Workshop organizers make last-minute changes to their schedule. Download this document again to get the latest changes, or use the ICML mobile application.

Schedule Highlights

July 12, 2020
WiML D&I Chairs Remarks: Sinead Williamson and Rachel Thomas

July 16, 2020
Graph Representation Learning and Beyond (GRL+): Velić, Bronstein, Deac, Hamilton, Hamrick, Hashemi, Jegelka, Leskovec, Liao, Monti, Sun, Swersky, Ying, Žitnik

XXAI: Extending Explainable AI Beyond Deep Models and Classifiers: Samek, Holzinger, Fong, Moon, Mueller

July 17, 2020
Self-supervision in Audio and Speech: Ravanelli, Serdyuk, Hjelm, Ramabhadran, Parcollet

5th ICML Workshop on Human Interpretability in Machine Learning (WHI): Weller, Xiang, Dharunthanhar, Kim, Wei, Varshney, Bhatt

Law & Machine Learning: Castets-Renard, Cussat-Blanc, Risser

Learning with Missing Values: Josse, Frellsen, Mattei, Varoquaux


Workshop on AI for Autonomous Driving (AIAD): Chao, McAllister, Gaidon, Li, Kreiss

Challenges in Deploying and Monitoring Machine Learning Systems: Tosi, Korda, Lawrence


Workshop on Continual Learning: Fayek, Chaudhry, Lopez-Paz, Belilovsky, Schwarz, Pickett, Aljundi, Ebrahimi, Dokania

Workshop on eXtreme Classification: Theory and Applications: Choromanska, Langford, Majzoubi, Prabhu

ICML 2020 Workshop on Computational Biology: Aghamirzaie, Anderson, Azizi, Diallo, Burdziak, Gallaher, Kundaje, Pe’er, Prabhakaran, Remita, Robertson-Tessi, Tansey, Vogt, Xie

Participatory Approaches to Machine Learning: Zhou, Madras, Raji, Mill, Kulynych, Zemel

Object-Oriented Learning: Perception, Representation, and Reasoning: Ahn, Kosirek, Hamrick, van Steenkiste, Bengio

Theoretical Foundations of Reinforcement Learning: Brunskill, Lykouris, Simchowitz, Sun, Wang

ML Interpretability for Scientific Discovery: Venugopalan, Brenner, Linderman, Kim

Uncertainty and Robustness in Deep Learning Workshop (UDL): Li, Lakshminarayanan, Hendrycks, Dietterich, Snoek

Beyond first order methods in machine learning systems: Berahas, Gholaminejad, Kynillidis, Mahoney, Roosta

July 18, 2020

4th Lifelong Learning Workshop: Sodhani, Chandar, Ravindran, Precup


Inductive Biases, Invariances and Generalization in Reinforcement Learning: Goyal, Ke, Wang, Weber, Viola, Schölkopf, Bauer

Negative Dependence and Submodularity: Theory and Applications in Machine Learning: Mariet, Derezinski, Gartrell

Federated Learning for User Privacy and Data Confidentiality: Baracaldo, Choudhury, Joshi, Raskar, Wang, Yu

Machine Learning for Global Health: Belgrave, Hyland, Onu, Furnham, Mwebaze, Lawrence

Bridge Between Perception and Reasoning: Graph Neural Networks & Beyond: Tang, Song, Leskovec, Liao, Li, Fidler, Zemel, Salakhutdinov

Economics of privacy and data labor: Vasiloglou, Cummings, Weyl, Koutris, Young, Jia, Dao, Waggoner


Real World Experiment Design and Active Learning: Bogunovic, Neiswanger, Yue

Workshop on Learning in Artificial Open Worlds: Szlam, Hofmann, Salakhutdinov, Kuno, Guss, Srinet, Houghton

1st Workshop on Language in Reinforcement Learning (LaReL): Nardelli, Luketina, Nardelli, Foerster, Zhong, Andreas, Grefenstette, Rocktäschel

Incentives in Machine Learning: Faltings, Liu, Parkes, Radanovic, Song

Machine Learning for Media Discovery: Schmidt, Nieto, Gouyon, Raimond, Kinnaid, Lanckriet

2nd ICML Workshop on Human in the Loop Learning (HILL): Zhang, Wang, Yu, Darrell
Graph Representation Learning and Beyond (GRL+)

Petar Veličković, Michael M. Bronstein, Andreea Deac, Will Hamilton, Jessica Hamrick, Milad Hashemi, Stefanie Jegelka, Jure Leskovec, Renjie Liao, Federico Monti, Yizhou Sun, Kevin Swersky, Zhilao Ying, Marinka Žitnik

Thu Jul 16, 23:40 PM

Recent years have seen a surge in research on graph representation learning, including techniques for deep graph embeddings, generalizations of CNNs to graph-structured data, and neural message-passing approaches. These advances in graph neural networks and related techniques have led to new state-of-the-art results in numerous domains: chemical synthesis, 3D-vision, recommender systems, question answering, continuous control, self-driving and social network analysis. Building on the successes of three related workshops from last year (at ICML, ICLR and NeurIPS), the primary goal for this workshop is to facilitate community building, and support expansion of graph representation learning into more interdisciplinary projects with the natural and social sciences. With hundreds of new researchers beginning projects in this area, we hope to bring them together to consolidate this fast-growing area into a healthy and vibrant subfield. Especially, we aim to strongly promote novel and exciting applications of graph representation learning across the sciences, reflected in our choices of invited speakers.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:40 PM</td>
<td>Opening Remarks Veličković, Deac</td>
</tr>
<tr>
<td>12:00 AM</td>
<td>Invited Talk: Xavier Bresson</td>
</tr>
<tr>
<td>12:30 AM</td>
<td>Invited Talk: Thomas Kip</td>
</tr>
<tr>
<td>01:00 AM</td>
<td>Q&amp;A / Discussions / Coffee 1</td>
</tr>
<tr>
<td>01:30 AM</td>
<td>Virtual Poster Session #1</td>
</tr>
<tr>
<td>02:30 AM</td>
<td>Novel Applications: Wiki-CS: A Wikipedia-Based Benchmark for Graph Neural Networks</td>
</tr>
<tr>
<td>02:40 AM</td>
<td>Novel Applications: Graph Neural Networks for Massive MIMO Detection</td>
</tr>
<tr>
<td>02:50 AM</td>
<td>Novel Applications: Embedding a random graph via GNN: Extended mean-field inference theory and RL applications to NP-Hard multi-robot/machine scheduling Kang</td>
</tr>
<tr>
<td>03:00 AM</td>
<td>Update: PyTorch Geometric Feny</td>
</tr>
<tr>
<td>03:10 AM</td>
<td>Update: Deep Graph Library Karypis</td>
</tr>
<tr>
<td>03:20 AM</td>
<td>Update: Open Graph Benchmark Leskovec</td>
</tr>
<tr>
<td>03:30 AM</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>04:30 AM</td>
<td>Original Research: Get Rid of Suspended Animation: Deep Diffusive Neural Network for Graph Representation Learning Zhang</td>
</tr>
<tr>
<td>04:45 AM</td>
<td>Original Research: Learning Graph Models for Template-Free Retrosynthesis Somnath</td>
</tr>
<tr>
<td>05:00 AM</td>
<td>Original Research: Frequent Subgraph Mining by Walking in Order Embedding Space Ying</td>
</tr>
<tr>
<td>05:15 AM</td>
<td>Invited Talk: Kyle Cranmer Cranmer</td>
</tr>
<tr>
<td>05:45 AM</td>
<td>Invited Talk: Danai Koutra Koutra</td>
</tr>
<tr>
<td>06:15 AM</td>
<td>Q&amp;A / Discussions / Coffee 2</td>
</tr>
<tr>
<td>06:45 AM</td>
<td>COVID-19 Applications: Navigating the Dynamics of Financial Embeddings over Time Gogoglou</td>
</tr>
<tr>
<td>07:05 AM</td>
<td>COVID-19 Applications: Gaining insight into SARS-CoV-2 infection and COVID-19 severity using self-supervised edge features and Graph Neural Networks Sehanobish</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>Invited Talk: Tina Eliassi-Rad Eliassi-Rad</td>
</tr>
<tr>
<td>07:45 AM</td>
<td>Invited Talk: Raquel Urtasun</td>
</tr>
<tr>
<td>Time</td>
<td>Session</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Q&amp;A / Discussions / Coffee</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Virtual Poster Session #2</td>
</tr>
<tr>
<td>09:45 AM</td>
<td>Closing Remarks</td>
</tr>
<tr>
<td>N/A</td>
<td>(#12 / Sess. 1) Deep Graph</td>
</tr>
<tr>
<td>N/A</td>
<td>(#29 / Sess. 2) Few-shot</td>
</tr>
<tr>
<td>N/A</td>
<td>(#37 / Sess. 1) Geometric</td>
</tr>
<tr>
<td>N/A</td>
<td>(#23 / Sess. 2) A Note on</td>
</tr>
<tr>
<td>N/A</td>
<td>(#80 / Sess. 2) Weisfeller</td>
</tr>
<tr>
<td>N/A</td>
<td>(#95 / Sess. 2) Graphein - a</td>
</tr>
<tr>
<td>N/A</td>
<td>(#31 / Sess. 2) Generalized</td>
</tr>
<tr>
<td>N/A</td>
<td>(#19 / Sess. 1) Neural Bipartite Matching</td>
</tr>
<tr>
<td>N/A</td>
<td>(#83 / Sess. 2) Connecting</td>
</tr>
<tr>
<td>N/A</td>
<td>(#84 / Sess. 2) UniKER: A</td>
</tr>
<tr>
<td>N/A</td>
<td>(#90 / Sess. 1) Pointer</td>
</tr>
<tr>
<td>N/A</td>
<td>(#97 / Sess. 1) Clustered</td>
</tr>
<tr>
<td>N/A</td>
<td>(#15 / Sess. 2) Learning</td>
</tr>
<tr>
<td>N/A</td>
<td>(#82 / Sess. 2) Scattering</td>
</tr>
<tr>
<td>N/A</td>
<td>(#8 / Sess. 2) Practical</td>
</tr>
<tr>
<td>N/A</td>
<td>(#58 / Sess. 1) Temporal</td>
</tr>
<tr>
<td>N/A</td>
<td>(#87 / Sess. 2) Bi-Level</td>
</tr>
<tr>
<td>N/A</td>
<td>(#46 / Sess. 2) Discrete</td>
</tr>
<tr>
<td>N/A</td>
<td>(#71 / Sess. 1) Population</td>
</tr>
<tr>
<td>N/A</td>
<td>(#64 / Sess. 1) Differentiable Graph Module (DGM) for Graph Convolutional Networks</td>
</tr>
<tr>
<td>N/A</td>
<td>(#62 / Sess. 2) Relate and</td>
</tr>
<tr>
<td>N/A</td>
<td>(#3 / Sess. 2) Spectral-designed Depthwise Separable Graph Neural Networks</td>
</tr>
<tr>
<td>N/A</td>
<td>(#92 / Sess. 1) From Graph</td>
</tr>
<tr>
<td>N/A</td>
<td>(#26 / Sess. 2) Message</td>
</tr>
<tr>
<td>N/A</td>
<td>(#74 / Sess. 2) Relation-Dependent Sampling for Multi-Relational Link Prediction</td>
</tr>
<tr>
<td>N/A</td>
<td>(#50 / Sess. 1) Graph</td>
</tr>
<tr>
<td>N/A</td>
<td>(#77 / Sess. 1) SIGN: Scalable Inception Graph Neural Networks</td>
</tr>
<tr>
<td>Session</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Geoopt: Riemannian Optimization in PyTorch</td>
</tr>
<tr>
<td>1</td>
<td>Set2Graph: Learning Graphs From Sets</td>
</tr>
<tr>
<td>1</td>
<td>TUDataset: A collection of benchmark datasets for learning with graphs</td>
</tr>
<tr>
<td>1</td>
<td>Scene Graph Reasoning for Visual Question Answering</td>
</tr>
<tr>
<td>2</td>
<td>Molecule Edit Graph Attention Network: Modeling Chemical Reactions as Sequences of Graph Edits</td>
</tr>
<tr>
<td>2</td>
<td>Continuous Graph Flow</td>
</tr>
<tr>
<td>2</td>
<td>Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction</td>
</tr>
<tr>
<td>1</td>
<td>Principal Neighbourhood Aggregation for Graph Nets</td>
</tr>
<tr>
<td>1</td>
<td>Improving Graph Neural Network Expressivity via Subgraph Isomorphism Counting</td>
</tr>
<tr>
<td>1</td>
<td>Graph Clustering with Graph Neural Networks</td>
</tr>
<tr>
<td>1</td>
<td>Hierarchically Attentive Graph Pooling with Subgraph Attention</td>
</tr>
<tr>
<td>2</td>
<td>A Graph VAE and Graph Transformer Approach to Generating Molecular Graphs</td>
</tr>
<tr>
<td>2</td>
<td>Learning Graph Structure with A Finite-State Automaton Layer</td>
</tr>
<tr>
<td>2</td>
<td>Multi-Graph Neural Operator for Parametric Partial Differential Equations</td>
</tr>
<tr>
<td>1</td>
<td>Graph neural induction of value iteration</td>
</tr>
<tr>
<td>1</td>
<td>Active Learning on Graphs via Meta Learning</td>
</tr>
<tr>
<td>2</td>
<td>Evaluating Logical Generalization in Graph Neural Networks</td>
</tr>
<tr>
<td>2</td>
<td>HINHN: Hypergraph Network with Hyperedge Neurons</td>
</tr>
<tr>
<td>2</td>
<td>Graph Neural Networks for the Prediction of Substrate-Specific Organic Reaction Conditions</td>
</tr>
<tr>
<td>1</td>
<td>Hierarchical Inter-Message Passing for Learning on Molecular Graphs</td>
</tr>
<tr>
<td>2</td>
<td>Degree-Quant: Quantization-Aware Training for Graph Neural Networks</td>
</tr>
<tr>
<td>1</td>
<td>Graphs, Entities, and Step Mixture</td>
</tr>
<tr>
<td>2</td>
<td>Graph Generation with Energy-Based Models</td>
</tr>
<tr>
<td>2</td>
<td>Uncovering the Folding Landscape of RNA Secondary Structure with Deep Graph Embeddings</td>
</tr>
<tr>
<td>1</td>
<td>Deep Lagrangian Propagation in Graph Neural Networks</td>
</tr>
<tr>
<td>2</td>
<td>Software Engineering Event Modeling using Relative Time in Temporal Knowledge Graphs</td>
</tr>
<tr>
<td>2</td>
<td>Are Hyperbolic Representations in Graphs Created Equal?</td>
</tr>
<tr>
<td>1</td>
<td>Uncertainty in Neural Relational Inference Trajectory Reconstruction</td>
</tr>
</tbody>
</table>
### Abstracts (71):

#### Abstract 6: Novel Applications: Wiki-CS: A Wikipedia-Based Benchmark for Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Mernyei 02:30 AM

We present Wiki-CS, a novel dataset derived from Wikipedia for benchmarking Graph Neural Networks. The dataset consists of nodes corresponding to Computer Science articles, with edges based on hyperlinks and 10 classes representing different branches of the field. We use the dataset to evaluate semi-supervised node classification and single-relation link prediction models. Our experiments show that these methods perform well on a new domain, with structural properties different from earlier benchmarks. We will make the dataset publicly available for future work.

#### Abstract 7: Novel Applications: Graph Neural Networks for Massive MIMO Detection in Graph Representation Learning and Beyond (GRL+), Scotti 02:40 AM

In this paper, we innovately use graph neural networks (GNNs) to learn a message-passing solution for the inference task of massive multiple-input multiple-output (MIMO) detection in wireless communication. We adopt a graphical model based on the Markov random field (MRF) where belief propagation (BP) yields poor results when it assumes a uniform prior over the transmitted symbols. Numerical simulations show that, under the uniform prior assumption, our GNN-based MIMO detection solution outperforms the minimum mean-squared error (MMSE) baseline detector, in contrast to BP. Furthermore, experiments demonstrate that the performance of the algorithm slightly improves by incorporating MMSE information into the prior.

#### Abstract 8: Novel Applications: Embedding a random graph via GNN: Extended mean-field inference theory and RL applications to NP-Hard multi-robot/machine scheduling in Graph Representation Learning and Beyond (GRL+), Kang 02:50 AM

We illustrate that developing a theory of how to embed a random graph using GNNs is the key to achieving the first near-optimal learning-based scheduling algorithm for an NP-hard multi-robot scheduling problem for tasks with time-varying rewards. We focus on a problem referred to as a Multi-Robot Reward Collection (MRRC) problem, of which immediate applications are ridesharing and pickup-and-delivery problems. We 1) observe that states in our robot scheduling problems can be represented as an extension of probabilistic graphical models (PGMs), which we refer to as random PGMs, and 2) develop a meanfield inference method for random PGMs. We then prove that a simple heuristic for applying deep graph encoder for random graph embedding is theoretically justified. We illustrate how a two-step hierarchical inference induces precise Q-function estimation. We empirically demonstrate that our method achieves near-optimality (plus transferability and scalability, machine scheduling (IPMS) applications in the appendix section). Arxiv preprint: https://arxiv.org/abs/1905.12204.
Abstract 14: Original Research: Learning Graph Models for Template-Free Retrosynthesis in Graph Representation Learning and Beyond (GRL+), Somnath 04:45 AM

Retrosynthesis prediction is a fundamental problem in organic synthesis, where the task is to identify precursor molecules that can be used to synthesize a target molecule. Despite advancements in neural retrosynthesis algorithms, they are unable to fully recapitulate the strategies employed by chemists and do not generalize well to infrequent reactions. In this paper, we propose a graph-based approach that capitalizes on the idea that graph topology of precursor molecules is largely unaltered during the reaction. The model first predicts the set of graph edits transforming the target into incomplete molecules called synthons. Next, the model learns to expand synthons into complete molecules by attaching relevant leaving groups. Since the model operates at the level of molecular fragments, it avoids full generation, greatly simplifying the underlying architecture and improving its ability to generalize. The model yields 11.7% absolute improvement over state-of-the-art approaches on the USPTO-50k dataset.

Abstract 15: Original Research: Frequent Subgraph Mining by Walking in Order Embedding Space in Graph Representation Learning and Beyond (GRL+), Ying 05:00 AM

Identifying frequent subgraphs or network motifs has been crucial in analyzing and predicting properties of real-world networks. However, finding large commonly-occurring motifs remains an open problem due to its NP-hard subroutine of subgraph counting, and the combinatorial growth of the number of possible subgraphs with their size. Here we present Subgraph Pattern Miner (SPMMiner), a novel neural approach to finding frequent subgraphs in a large target graph. SPMMiner integrates graph neural networks, order embedding space, and an efficient search strategy to identify network subgraph patterns that appear most frequently in a target graph dataset. SPMMiner first decomposes the target graph into many overlapping subgraphs and then encodes the subgraphs into order embeddings. SPMMiner then uses a monotonic walk in the order embedding space to identify frequent motifs. Compared to existing approaches and possible neural alternatives, SPMMiner is more accurate, faster, and more scalable. For 5- and 6-node motifs, we show that SPMMiner can identify almost all of the most frequent motifs while being 100x faster than exact enumeration methods. In addition, SPMMiner can also reliably identify frequent 10-node motifs, which is well beyond the size limit of exact enumeration approaches. And last, We show that SPMMiner can find large 10+ node motifs that appear 10-100x more frequently than those found by current widely-used approximate methods.

Abstract 19: COVID-19 Applications: Navigating the Dynamics of Financial Embeddings over Time in Graph Representation Learning and Beyond (GRL+), Gogolou 06:45 AM

Financial transactions constitute connections between entities and through these connections a large scale heterogeneous weighted graph is formulated. In this labyrinth of interactions that are continuously updated, there exists a variety of similarity-based patterns that can provide insights into the dynamics of the financial system. With the current work, we propose the application of Graph Representation Learning in a scalable dynamic setting as a means of capturing these patterns in a meaningful and robust way. We proceed to perform a rigorous qualitative analysis of the latent trajectories to extract real world insights from the proposed representations and their evolution over time that is to our knowledge the first of its kind in the financial sector. Shifts in the latent space are associated with known economic events and in particular the impact of the recent Covid-19 pandemic to consumer patterns. Capturing such patterns indicates the value added to financial modeling through the incorporation of latent graph representations.

Abstract 20: COVID-19 Applications: Integrating Logical Rules Into Neural Multi-Hop Reasoning for Drug Repurposing in Graph Representation Learning and Beyond (GRL+), Liu 06:55 AM

The graph structure of biomedical data differs from those in typical knowledge graph benchmarking tasks. A particular property of biomedical data is the presence of long-range dependencies, which can be captured by patterns described as logical rules. We propose a novel method that combines these rules with a neural multi-hop reasoning approach that uses reinforcement learning. We conduct an empirical study based on the real-world task of drug repurposing by formulating this task as a link prediction problem. We apply our method to the biomedical knowledge graph Hetionet and show that our approach outperforms several baseline methods.

Abstract 21: COVID-19 Applications: Gaining insight into SARS-CoV-2 infection and COVID-19 severity using self-supervised edge features and Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Sehanobish 07:05 AM

Graph Neural Networks (GNN) have been extensively used to extract meaningful representations from graph structured data and to perform predictive tasks such as node classification and link prediction. In recent years, there has been a lot of work incorporating edge features along with node features for prediction tasks. In this work, we present a framework for creating new edge features, via a combination of self-supervised and unsupervised learning which we then use along with node features for node classification tasks. We validate our work on two biological datasets comprising of single-cell RNA sequencing data of in vitro SARS-CoV-2 infection and human COVID-19 patients. We demonstrate that our method achieves better performance over baseline
Graph Attention Network (GAT) and Graph Convolutional Network (GCN) models. Furthermore, given the attention mechanism on edge and node features, we are able to interpret the cell types and genes that determine the course and severity of COVID-19, contributing to a growing list of potential disease biomarkers and therapeutic targets.

Abstract 27: (#12 / Sess. 1) Deep Graph Contrastive Representation Learning in Graph Representation Learning and Beyond (GRL+), Zhu N/A

Graph representation learning nowadays becomes fundamental in analyzing graph-structured data. Inspired by recent success of contrastive methods, in this paper, we propose a novel framework for unsupervised graph representation learning by leveraging a contrastive objective at the node level. Specifically, we generate two graph views by corruption and learn node representations by maximizing the agreement of node representations in these two views. To provide diverse node contexts for the contrastive objective, we propose a hybrid scheme for generating graph views on both structure and attribute levels. We perform empirical experiments on both transductive and inductive learning tasks using a variety of real-world datasets. Experimental experiments demonstrate that despite its simplicity, our proposed method consistently outperforms existing state-of-the-art methods by large margins. Notably, our method gains about 10% absolute improvements on protein function prediction. Our unsupervised method even surpasses its supervised counterparts on transductive tasks.

Abstract 28: (#29 / Sess. 2) Few-shot link prediction via graph neural networks for Covid-19 drug-repurposing (GRL+), Zheng N/A

Predicting interactions among heterogeneous graph structured data has numerous applications such as knowledge graph completion, recommendation systems and drug discovery. Often times, the links to be predicted belong to rare types such as the case in repurposing drugs for novel diseases. This motivates the task of few-shot link prediction. Typically, GCNs are ill-equipped in learning such rare link types since the contexts for the contrastive objective, we propose a hybrid scheme for generating graph views on both structure and attribute levels. We perform empirical experiments on both transductive and inductive learning tasks using a variety of real-world datasets. Experimental experiments demonstrate that despite its simplicity, our proposed method consistently outperforms existing state-of-the-art methods by large margins. Notably, our method gains about 10% absolute improvements on protein function prediction. Our unsupervised method even surpasses its supervised counterparts on transductive tasks.

Abstract 29: (#37 / Sess. 1) Geometric Matrix Completion: A Functional View in Graph Representation Learning and Beyond (GRL+), Sharma N/A

We propose a totally functional view of geometric matrix completion problem. Differently from existing work, we propose a novel regularization inspired from the functional map literature that is more interpretable and theoretically sound. On synthetic tasks with strong underlying geometric structure, our framework outperforms state of the art by a huge margin (two order of magnitude) demonstrating the potential of our approach. On real datasets, we achieve state-of-the-art results at a fraction of the computational effort of previous methods.
Graph neural networks (GNNs) have found application for learning in the space of algorithms. However, the algorithms chosen by existing research (sorting, Breadth-First search, shortest path finding, etc.) usually align perfectly with a standard GNN architecture. This report analyzes how deeper networks can be applied to a complex algorithm, such as finding maximum bipartite matching by reducing it to a flow problem and using Ford-Fulkerson to find the maximum flow. This is achieved via neural execution based only on features generated from a single GNN. The evaluation shows strongly generalising results with the network achieving optimal matching almost 100% of the time.

Graph convolution operator of the GCN model is originally motivated from a localized first-order approximation of spectral graph convolutions. This work stands on a different view; establishing a connection between graph convolution and graph-regularized PCA. Based on this connection, GCN architecture, shaped by stacking graph convolution layers, shares a close relationship with stacking graph-regularized PCA (GPCA). We empirically demonstrate that the unsupervised embeddings by GPCA paired with a logistic regression classifier achieves similar performance to GCN on semi-supervised node classification tasks. Further, we capitalize on the discovered relationship to design an effective baseline framework for GNN based protein stacking GPCA.

Graph Convolutional Networks (GCNs) have received increasing attention in recent machine learning. How to effectively leverage the rich structural information in complex graphs, such as knowledge graphs with heterogeneous types of entities and relations, is a primary open challenge in the field. Most GCN methods are either restricted to graphs with a homogeneous type of edges (e.g., citation links only), or focusing on representation learning for nodes only instead of jointly optimizing the embeddings of both nodes and edges for target-driven objectives. This paper addresses these limitations by proposing a novel framework, namely the GEneralized Multi-relational Graph Convolutional Networks (GEM-GCN), which combines the power of GCNs in graph-based belief propagation and the strengths of advanced knowledge-base embedding methods, and goes beyond. Our theoretical analysis shows that GEM-GCN offers an elegant unification of several well-known GCN methods as specific cases, with a new perspective of graph convolution. Experimental results on benchmark datasets show the advantageous performance of GEM-GCN over strong baseline methods in the tasks of knowledge graph alignment and entity classification.

Combining KGE and logical rules for better KG inference have gained increasing attention in recent years. Unfortunately, a majority of existing methods employ sampling strategies to randomly select only a small portion of ground rules or hidden triples, thus only partially leverage the power of logical rules in reasoning. In this paper, we propose a novel framework UniKER to address this challenge by restricting logical rules to be Horn rules, which can fully exploit the knowledge in logical rules and enable the mutual enhancement of logical rule-based reasoning and KGE in an extremely efficient way. Extensive experiments have demonstrated that our approach is superior to existing state-of-the-art algorithms in terms of both efficiency and effectiveness.
Graph neural networks (GNNs) are typically applied to static graphs that are assumed to be known upfront. This static input structure is often informed purely by insight of the machine learning practitioner, and might not be optimal for the actual task the GNN is solving. We introduce Pointer Graph Networks (PGNs) which augment sets or graphs with additional inferred edges for improved model expressivity. PGNs allow each node to dynamically point to another node, followed by message passing over these pointers. Despite its sparsity, this adaptable graph structure proves sufficiently expressive to simulate complex algorithms. The pointing mechanism is supervised to model long-term sequences of operations on classical data structures. PGNs can learn parallelisable variants of pointer-based data structures, and generalise out-of-distribution to 5x larger test inputs on dynamic graph connectivity tasks, outperforming unrestricted GNNs and Deep Sets.


Abstract 38: (#97 / Sess. 1) Clustered Dynamic Graph CNN for Biometric 3D Hand Shape Recognition in Graph Representation Learning and Beyond (GRL+), Svoboda N/A

The research in biometric recognition using hand shape has been somewhat stagnating in the last decade. Meanwhile, computer vision and machine learning have experienced a paradigm shift with the renaissance of deep learning, which has set the new state-of-the-art in many related fields. Inspired by successful applications of deep learning for other biometric modalities, we propose a novel approach to 3D hand shape recognition from RGB-D data based on geometric deep learning techniques. We show how to train our model on synthetic data and retain the performance on real samples during test time. To evaluate our method, we provide a new dataset NNHand RGB-D of short video sequences and show encouraging performance compared to diverse baselines on the new data, as well as current benchmark dataset HKPolyU. Moreover, the new dataset opens door to many new research directions in hand shape recognition.

[Teaser video](https://slideslive.com/38931485/clusted-dynamic-graph-cnn-for-biometric-3d-hand-shape-recognition) | [Zoom join link](https://zoom.us/j/94803775606)

Abstract 39: (#15 / Sess. 2) Learning Distributed Representations of Graphs with Geo2DR in Graph Representation Learning and Beyond (GRL+), Scherer N/A

We present Geo2DR (Geometric to Distributed Representations), a Python library for unsupervised learning on graph-structured data using discrete substructure patterns and neural language models. It contains efficient implementations of popular graph decomposition algorithms and neural language models in PyTorch which are combined to learn representations using the distributive hypothesis. Furthermore, Geo2DR comes with general data processing and loading methods which can bring substantial speed-up in the training of the neural language models. Through this we provide a unified set of tools and methods to quickly construct systems capable of learning distributed representations of graphs. This is useful for replication of existing methods, modification, or development of novel systems. This paper serves to present the Geo2DR library and perform a comprehensive comparative analysis of existing methods re-implemented using Geo2DR across several widely used graph classification benchmarks. Geo2DR displays a high reproducibility of results in published methods and interoperability with other libraries useful for distributive language modeling.


Abstract 40: (#82 / Sess. 2) Scattering GCN: Overcoming Oversmoothness in Graph Convolutional Networks in Graph Representation Learning and Beyond (GRL+), Wenkel N/A

Graph convolutional networks (GCNs) are widely used for semi-supervised node classification on graphs today. The graph structure is however only accounted for by considering the similarity of activations between adjacent nodes, in turn degrading the results. In this work, we augment GCN models by incorporating richer notions of regularity by leveraging cascades of band-pass filters, known as geometric scatterings. We introduce a new hybrid architecture for the task and demonstrate its potential on multiple graph datasets, where it outperforms leading GCN models.


Abstract 41: (#8 / Sess. 2) Practical Adversarial Attacks on Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Ding N/A

We study the black-box attacks on graph neural networks (GNNs) under a novel and realistic constraint: attackers have access to only a subset of nodes in the network, and they can only attack a small number of them. A node selection step is essential under this setup. We demonstrate that the structural inductive biases of GNN models can be an effective source for this type of attacks. Specifically, by exploiting the connection between the backward propagation of GNNs and random walks, we show that the common gradient-based white-box attacks can be generalized to the black-box setting via the connection between the gradient and an importance score similar to PageRank. In practice, we find attacks based on this importance score indeed increase the classification loss by a large margin, but they fail to significantly increase the mis-classification rate. Our further analyses suggest that there is a discrepancy between the loss and mis-classification rate, as the latter presents a diminishing-return pattern when the number of attacked nodes increases. Therefore, we propose a greedy procedure to correct the importance score that takes into account of the diminishing-return pattern. Experimental results show that the proposed procedure can significantly increase the mis-classification rate of common GNNs on real-world data without access to model parameters nor predictions.


Abstract 42: (#58 / Sess. 1) Temporal Graph Networks for Deep Learning on Dynamic Graphs in Graph Representation Learning and Beyond (GRL+), Rossi N/A
Graph Neural Networks (GNNs) have become increasingly popular due to their ability to learn complex systems of relations or interactions arising in a broad spectrum of problems ranging from biology and particle physics to social networks and recommendation systems. Despite the plethora of different models for deep learning on graphs, few approaches have been proposed thus far for dealing with graphs that present some sort of dynamic nature (e.g., evolving features or connectivity over time).

In this paper, we present Temporal Graph Networks (TGNs), a generic, efficient framework for deep learning on dynamic graphs. Thanks to a novel combination of memory modules and graph-based operators, TGNs are able to significantly outperform previous approaches, being at the same time more computationally efficient.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

We present Bi-Level Attention-Based Relational Graph Convolutional Networks (BR-GCN), unique neural network architectures that utilize masked self-attentional layers with relational graph convolutions, to effectively operate on highly multi-relational data. BR-GCN models use bi-level attention to learn node embeddings through (1) node-level attention, and (2) relation-level attention. BR-GCN's node-level self-attentional layers use intra-relational graph interactions to learn relation-specific node embeddings using a weighted aggregation of neighborhood features in a sparse subgraph region. BR-GCN's relation-level self-attentional layers use inter-relational graph interactions to learn the final node embeddings using a weighted aggregation of relation-specific node embeddings. BR-GCN's bi-level attention mechanism extends Transformer-based multiplicative attention from the natural language processing (NLP) domain, and Graph Attention Networks (GAT)-based attention, to large-scale heterogeneous graphs (HGs). On node classification, BR-GCN outperforms baselines from 0.02% to 7.40% and suggests to enrich HG embedding models. We also conduct ablation studies to evaluate the quality of BR-GCN's relation-level attention and discuss how its learning of graph structure may be transferred to enrich other Graph Neural Networks (GNNs). Through various experiments, we show that BR-GCN's attention mechanism is both scalable and more effective in learning compared to state-of-the-art GNNs.

Graph Neural Networks (GNNs) have become increasingly popular due to their ability to learn complex systems of relations or interactions arising in a broad spectrum of problems ranging from biology and particle physics to social networks and recommendation systems. Despite the plethora of different models for deep learning on graphs, few approaches have been proposed thus far for dealing with graphs that present some sort of dynamic nature (e.g., evolving features or connectivity over time).

In this paper, we present Temporal Graph Networks (TGNs), a generic, efficient framework for deep learning on dynamic graphs. Thanks to a novel combination of memory modules and graph-based operators, TGNs are able to significantly outperform previous approaches, being at the same time more computationally efficient.
our model provides a significant improvement over baselines both in transductive and inductive settings and achieves state-of-the-art results.

Abstract 47: (#62 / Sess. 2) Relate and Predict: Structure-Aware Prediction with Jointly Optimized Neural Dependency Graph in Graph Representation Learning and Beyond (GRL+), Sekhon N/A

Understanding relationships between feature variables is one important way humans use to make decisions. However, state-of-the-art deep learning studies either focus on task-agnostic statistical dependency learning or do not model explicit feature dependencies during prediction. We propose a deep neural network framework, dGAP, to learn neural dependency Graph and optimize structure-Aware target Prediction simultaneously. dGAP trains towards a structure self-supervision loss and a target prediction loss jointly. Our method leads to an interpretable model that can disentangle sparse feature relationships, informing the user how relevant dependencies impact the target task. We empirically evaluate dGAP on multiple simulated and real datasets. dGAP is not only accurate, but can also recover correct dependency structure.

Abstract 48: (#3 / Sess. 2) Spectral-designed Depthwise Separable Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Balcilar N/A

This paper aims at revisiting Convolutional Graph Neural Networks (ConvGNNs) by designing new graph convolutions in spectral domain with a custom frequency profile while applying them in the spatial domain. Within the proposed framework, we propose two ConvGNNs methods: one using a simple single-convolution kernel that operates as a low-pass filter, and one operating multiple convolution kernels called Depthwise Separable Graph Convolution Network (DSGCN). The latter is a generalization of the depthwise separable convolution framework for graph convolutional networks, which allows to decrease the total number of trainable parameters while keeping the capacity of the model unchanged. Our proposals are evaluated on both transductive and inductive graph learning problems, demonstrating that DSGCN outperforms the state-of-the-art methods.

Abstract 49: (#92 / Sess. 1) From Graph Low-Rank Global Attention to 2-FWL Approximation in Graph Representation Learning and Beyond (GRL+), Puny N/A

Graph Neural Networks are known to have an expressive power bounded by that of the vertex coloring algorithm \cite{Xu2019goalgown} of GNNs with the more powerful 2-folklore Weisfeiler-Lehman (FWL) algorithm. Furthermore, we provide a sample complexity bound for the module using kernel feature map interpretation of 2-FWL. Empirically, augmenting existing GNN layers with LRGA produces state of the art results on most datasets in a GNN standard benchmark.

Abstract 50: (#26 / Sess. 2) Message Passing Query Embedding in Graph Representation Learning and Beyond (GRL+), Daza N/A

Recent works on representation learning for Knowledge Graphs have moved beyond the problem of link prediction, to answering queries of an arbitrary structure. Existing methods are based on ad-hoc mechanisms that require training with a diverse set of query structures. We propose a more general architecture that employs a graph neural network to encode a graph representation of the query, where nodes correspond to entities and variables. The generality of our method allows it to encode a more diverse set of query types in comparison to previous work. Our method shows competitive performance against previous models for complex queries when trained for link prediction only. We show that the model learns entity embeddings that capture the notion of entity type without explicit supervision.

Abstract 51: (#74 / Sess. 2) Relation-Dependent Sampling for Multi-Relational Link Prediction in Graph Representation Learning and Beyond (GRL+), Thost, Feeney, Gupta N/A

Multi-relational graphs contain various types of relations that usually come with variable frequency and have different importance for the problem at hand. Existing graph sampling approaches ignore the multi-relational nature of such graphs. We propose an approach to modeling the importance of relation types for sampling and show that we can learn the right balance: relation-type probabilities that reflect both frequency and importance. We use relation-dependent sampling to develop a scalable graph neural network and apply it for multi-relational link prediction. Our experiments specifically consider drug-drug interaction (DDI) prediction, an important task during drug development. In that context, we further show the benefit of considering a special relation type, negative edges. Synergistic drug combinations (e.g., drugs that synergize by improved efficacy and reduced side effects) can be regarded as negative evidence for DDI (which often indicate adverse reactions). We add drug synergy data to provide extra expert knowledge which can be easily integrated into our model and yields improved performance.
Abstract 52: (#50 / Sess. 1) Graph Convolutional Gaussian Processes for Link Prediction in Graph Representation Learning and Beyond (GRL+), Opolka N/A

Link prediction aims to reveal missing edges in a graph. We address this task with a deep graph convolutional Gaussian process model. The Gaussian process is transformed using simplified graph convolutions to better leverage the topological information of the graph domain. To scale the Gaussian process model to larger graphs, we introduce a variational inducing point method that places pseudo-inputs on a graph-structured domain. The proposed model represents the first Gaussian process for link prediction that can make use of both node features and topological information. We evaluate our model on three graph data sets with up to thousands of nodes and report consistent improvements over existing Gaussian process models and state-of-the-art graph neural network approaches.

Abstract 53: (#77 / Sess. 1) SIGN: Scalable Inception Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Frasca N/A

The popularity of graph neural networks has sparked interest, both in academia and in industry, in developing methods that scale to very large graphs such as Facebook or Twitter social networks. In most of these approaches, the computational cost is alleviated by a sampling strategy retaining a subset of node neighbors or subgraphs at training time. In this paper we propose a new, efficient and scalable graph deep learning architecture, which sidesteps the need for graph sampling by using graph convolutional filters of different size that are amenable to efficient precomputation, allowing extremely fast training and inference. Our architecture allows using different local graph operators (e.g. motif-induced adjacency matrices or Personalized Page Rank diffusion matrix) to best suit the task at hand. We conduct extensive experimental evaluation on various open benchmarks and show that our approach is competitive with other state-of-the-art architectures, while requiring a fraction of training and inference time.

Abstract 54: (#93 / Sess. 1) Geoopt: Riemannian Optimization in PyTorch in Graph Representation Learning and Beyond (GRL+), Kochurov N/A

Geoopt is a research-oriented modular open-source package for Riemannian Optimization in PyTorch. The core of Geoopt is a standard Manifold interface that allows for the generic implementation of optimization algorithms. Geoopt supports basic Riemannian SGD as well as adaptive optimization algorithms. Geoopt also provides several algorithms and arithmetic methods for supported manifolds, which allow composing geometry-aware neural network layers that can be integrated with existing models.

Abstract 55: (#34 / Sess. 1) Set2Graph: Learning Graphs From Sets in Graph Representation Learning and Beyond (GRL+), Serviansky N/A

Many problems in machine learning (ML) can be cast as learning functions from sets to graphs, or more generally to hypergraphs; in short, Set2Graph functions. Examples include clustering, learning vertex and edge features on graphs, and learning features on triplets in a collection. A natural approach for building Set2Graph models is to characterize all linear equivariant set-to-hypergraph layers and stack them with non-linear activations. This posses two challenges: (i) the expressive power of these networks is not well understood; and (ii) these models would suffer from high, often intractable computational and memory complexity, as their dimension grows exponentially. This paper advocates a family of neural network models for learning Set2Graph transformations, both practical and of maximal expressive power (universal), that is, can approximate arbitrary continuous Set2Graph functions over compact sets. Testing these models on different machine learning tasks, mainly an application to particle physics, we find them favorable to existing baselines.

Abstract 56: (#79 / Sess. 1) TUDataset: A collection of benchmark datasets for learning with graphs in Graph Representation Learning and Beyond (GRL+), Kriege N/A

Recently, there has been an increasing interest in (supervised) learning with graph data, especially using graph neural networks. However, the development of meaningful benchmark datasets and standardized evaluation procedures is lagging, consequently hindering advancements in this area. To address this, we introduce the TUDataset for graph classification and regression. The collection consists of over 120 datasets of varying sizes from a wide range of applications. We provide Python-based data loaders, kernel and graph neural network baseline implementations, and evaluation tools. Here, we give an overview of the datasets, standardized evaluation procedures, and provide baseline experiments. All datasets are available at www.graphlearning.io. The experiments are fully reproducible from the code available at www.github.com/chrismr Ruseth dataset.

Abstract 57: (#53 / Sess. 1) Scene Graph Reasoning for Visual Question Answering in Graph Representation Learning and Beyond (GRL+), Koner N/A

Visual question answering is concerned with answering free-form questions about an image. Since it requires a deep linguistic understanding of the question and the ability to associate it with various objects that are present in the image, it is an ambitious task and requires
We propose a novel method that approaches the task by performing context-driven, sequential reasoning based on the objects and their semantic and spatial relationships present in the scene. As a first step, we derive a scene graph which describes the objects in the image, as well as their attributes and their mutual relationships. A reinforcement agent then learns to autonomously navigate over the extracted scene graph to generate paths, which are then the basis for deriving answers. We conduct a first experimental study on the challenging GQA dataset with manually curated scene graphs, where our method reaches the level of human performance.

One of the key challenges in automated chemical synthesis planning is to propose diverse and reliable reactions. A common approach is to generate reactions using reaction templates, which represent a reaction as a fixed graph transformation. This enables accurate and interpretable predictions but can suffer from limited diversity. On the other hand, template-free methods increase diversity but can be prone to making trivial mistakes. Inspired by the efficacy of reaction templates, we propose Molecule Edit Graph Attention Network (MEGAN), a template-free model that encodes reaction as a sequence of graph edits. Our model achieves state-of-the-art results on a standard retrosynthesis benchmark without any manual rule encoding.

In this paper, we propose Continuous Graph Flow, a generative continuous flow based method that aims to model complex distributions of graph-structured data. Our proposed model learns a joint probability density over a set of related random variables by formulating it as first order ordinary differential equation system with shared and reusable functions that operate over the graph structure. This leads to a reversible continuous message passing over time resulting in continuous transformations of probability distributions of the variables. We evaluate our model on a diverse set of generation tasks: graph generation, image puzzle generation, and layout generation from scene graphs. Experimental results show that CGF-based models outperform state-of-the-art graph generative models.

We introduce Bi-GNN for modeling biological link prediction tasks such as drug-drug interaction (DDI) and protein-protein interaction (PPI). Taking drug-drug interaction as an example, existing methods using machine learning either only utilize the link structure between drugs without using the graph representation of each drug molecule, or only leverage the individual drug compound structures without using graph structure for the higher-level DDI graph. The key idea of our method is to fundamentally view the data as a bi-level graph, where the highest level is a representing the interaction between biological entities (interaction graph), and each biological entity itself is further expanded to its intrinsic graph representation (representation graphs), where the graph is either a fixed drug compound or hierarchical like a protein with amino acid level graph, secondary structure, tertiary structure, etc. Our model not only allows the usage of information from both the high-level interaction graph and the low-level representation graphs, but also offers a baseline for future research opportunities to address the bi-level nature of the data.

Graph Neural Networks (GNNs) have been shown to be effective models for different predictive tasks on graph-structured data. Recent work on their expressive power has focused on isomorphism tasks and countable feature spaces. We extend this theoretical framework to include continuous features—which occur regularly in real-world input domains and within the hidden layers of GNNs—and we demonstrate the architecture’s potential for multi-level aggregation functions in this context. Accordingly, we propose Principal Neighbourhood Aggregation (PNA), a novel architecture combining multiple aggregators with degree-scalers (which generalize the sum aggregator). Finally, we compare the capacity of different models to capture and exploit the graph structure via a novel benchmark containing multiple tasks taken from classical graph theory, alongside existing benchmarks from real-world domains, all of which demonstrate the strength of our model.

Graph Neural Networks (GNNs) have been recently found to suffer from important limitations regarding their ability to capture the structure of the underlying graph. It has been shown that the expressive power of standard GNNs is bounded by the Weisfeiler-Lehman (WL) graph isomorphism test, from which they inherit proven limitations such as the inability to detect and count graph substructures. On the other hand, there is significant empirical evidence that substructures are often informative for downstream tasks, suggesting that it is desirable to
Graph Neural Networks (GNNs) have achieved state-of-the-art results on many graph analysis tasks such as node classification and link prediction. However, important unsupervised problems on graphs, such as graph clustering, have proved more resistant to advances in GNNs. In this paper, we study unsupervised training of GNN pooling in terms of their clustering capabilities. We draw a connection between graph clustering and graph pooling: intuitively, a good graph clustering is what one would expect from a GNN pooling layer. Counterintuitively, we show that this is not true for state-of-the-art pooling methods, such as MinCut pooling. To address these deficiencies, we introduce Deep Modularity Networks (DMoN), an unsupervised pooling method inspired by the modularity measure of clustering quality, and show how it tackles recovery of the challenging clustering structure of graphs. In order to clarify the regimes where existing methods fail, we carefully design a set of experiments on synthetic data which show that DMoN is able to jointly leverage the signal from the graph structure and node features. Similarly, on real-world data, we show that DMoN produces high quality clusters which correlate strongly with ground truth labels, achieving state-of-the-art results.

Graph Neural Networks get significant attention for graph representation and classification in machine learning community. Different types of neighborhood aggregation and pooling strategies have been proposed in the literature. In this work, we introduce a higher order hierarchical GNN algorithm (SubGattPool) by employing (i) an attention mechanism which learns the importance and aggregates neighboring subgraphs of a node instead of first-order neighbors, and (ii) a hierarchical pooling strategy which learns the importance of different hierarchies in a GNN. SubGattPool is able to achieve state-of-the-art graph classification performance on multiple real-world datasets.

We propose a combination of a variational autoencoder and a transformer based model which fully utilises graph convolutional and graph pooling layers to operate directly on graphs. The transformer model implements a novel node encoding layer, replacing the position encoding typically used in transformers, to create a transformer with no position information that operates on graphs, encoding adjacent node properties into the edge generation process. The proposed model builds on graph generative work operating on graphs with edge features, creating a model that offers improved scalability with the number of nodes in a graph. In addition, our model is capable of learning a disentangled, interpretable latent space that represents graph properties through a mapping between latent variables and graph properties. In experiments we chose a benchmark task of molecular generation, given the importance of both generated node and edge features. Using the QM9 dataset we demonstrate that our model performs strongly across the task of generating valid, unique and novel molecules. Finally, we demonstrate that the model is interpretable by generating molecules controlled by molecular properties, and we then analyse and visualise the learned latent representation.
Graph neural networks (GNNs) have gained popularity in simulating physical systems and solving partial differential equations (PDEs) since graphs offer a natural way of modeling particle interactions and discretizing the continuum models. However, the graphs constructed for approximating such tasks usually ignore long-range interactions due to unfavorable scaling of the computational complexity with respect to the number of nodes. The errors due to these approximations scale with the discretization of the system, thereby not allowing for generalization under mesh-refinement. Inspired by the classical multipole methods, we propose a novel multi-level graph neural network framework that captures interaction at all ranges with only linear complexity. Our multi-level formulation is equivalent to recursively adding inducing points to the kernel matrix, unifying GNNs with multi-resolution matrix factorization of the kernel. Experiments confirm our multi-graph network learns discretization-invariant solution operators to PDEs and can be evaluated in linear time.

Many reinforcement learning tasks can benefit from explicit planning based on an internal model of the environment. Previously, such planning components have been incorporated through a neural network that partially aligns with the computational graph of value iteration. Such network have so far been focused on restrictive environments (e.g. grid-worlds), and modelled the planning procedure only indirectly. We relax these constraints, proposing a graph neural network (GNN) that executes the value iteration (VI) algorithm, across arbitrary environment models, with direct supervision on the intermediate steps of VI. The results indicate that GNNs are able to model value iteration accurately, recovering favourable metrics and policies across a variety of out-of-distribution tests. This suggests that GNN executors with strong supervision are a viable component within deep reinforcement learning systems.

Active learning (AL) for semi-supervised node classification aims to reduce the number of labeled instances by selecting only the most informative nodes for labeling. The AL algorithms designed for other data types such as images and text do not perform well on graph-structured data. Although a few heuristics-based AL algorithms have been proposed for graphs, a principled approach is lacking. We propose MetAL, an AL algorithm that selects unlabeled items that directly improve the future performance of a graph neural network (GNN) model. We formulate the AL problem as a bilevel optimization problem. Based on recent work in meta-learning, we compute the meta-gradients to approximate the impact of unlabeled instances on the model uncertainty. We empirically demonstrate that MetAL outperforms existing AL algorithms.

Recent research has highlighted the role of relational inductive biases in building learning agents that can generalize and reason in a compositional manner. However, while relational learning algorithms such as graph neural networks (GNNs) show promise, we do not understand how effectively these approaches can adapt to new tasks. In this work, we study the task of logical generalization using GNNs by designing a benchmark suite grounded in first-order logic. Our benchmark suite, GraphLog, requires that learning algorithms perform rule induction in different synthetic logics, represented as knowledge graphs. GraphLog consists of relation prediction tasks on 57 distinct logical domains. We use GraphLog to evaluate GNNs in three different setups: single-task supervised learning, multi-task pretraining, and continual learning. Unlike previous benchmarks, our approach allows us to precisely control the logical relationship between the different tasks. We find that the ability for models to generalize and adapt is strongly determined by the diversity of the logical rules they encounter during training, and our results highlight new challenges for the design of GNN models.

Hypergraphs provide a natural representation for many real world datasets. We propose a novel framework, HNHN, for hypergraph representation learning. HNHN is a hypergraph convolution network with nonlinear activation functions applied to both hypernodes and hyperedges, combined with a normalization scheme that can flexibly adjust the importance of high-cardinality hyperedges and high-degree vertices depending on the dataset. We demonstrate improved performance of HNHN in both classification accuracy and speed on real world datasets when compared to state of the art methods.
classification tasks pertaining to the identification of experimental reagents and conditions. We find that models are able to identify specific graph features that affect reaction conditions and lead to accurate predictions. The results herein show great promise in advancing molecular machine learning.


Abstract 73: (#45 / Sess. 1) Hierarchical Inter-Message Passing for Learning on Molecular Graphs in Graph Representation Learning and Beyond (GRL+), *Fey* N/A

We present a hierarchical neural message passing architecture for learning on molecular graphs. Our model takes in two complementary graph representations: the raw molecular graph representation and its associated junction tree, where nodes represent meaningful clusters in the original graph, e.g., rings or bridged compounds. We then proceed to learn a molecule’s representation by passing messages inside each graph, and exchange messages between the two representations using a coarse-to-fine and fine-to-coarse information flow. Our method is able to overcome some of the restrictions known from classical GNNs, like detecting cycles, while still being very efficient to train. We validate its performance on the ZINC dataset and datasets stemming from the MoleculeNet benchmark collection.

[Teaser video](https://slideslive.com/38931406/hierarchical-intermessage-passing-for-learning-on-molecular-graphs)

[Zoom join link](https://zoom.us/j/97893366407)

Abstract 74: (#24 / Sess. 2) Degree-Quant: Quantization-Aware Training for Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), *Tailor* N/A

Graph neural networks have demonstrated strong performance modelling non-uniform structured data. However, there exists little research exploring methods to make them more efficient at inference time. In this work, we explore the viability of training quantized GNNs models, enabling the usage of low precision integer arithmetic for inference. We propose a method, Degree-Quant, to improve performance over existing quantization-aware training baselines commonly used on other architectures, such as CNNs. Our work demonstrates that it is possible to train models using 8-bit integer arithmetic at inference-time with similar accuracy to their full precision counterparts.


[Zoom join link](https://zoom.us/j/92128669371)

Abstract 75: (#89 / Sess. 1) Graphs, Entities, and Step Mixture in Graph Representation Learning and Beyond (GRL+), *Shin* N/A

Existing approaches for graph neural networks commonly suffer from the oversmoothing issue, regardless of how neighborhoods are aggregated. Most methods also focus on transductive scenarios for fixed graphs, leading to poor generalization for unseen graphs. To address these issues, we propose a new graph neural network that considers both edge-based neighborhood relationships and node-based entity features, i.e. Graph Entities with Step Mixture via random walk (GESM). GESM employs a mixture of various steps through random walk to alleviate the oversmoothing problem, attention to dynamically reflect interrelations depending on node information, and structure-based regularization to enhance embedding representation. With intensive experiments, we show that the proposed GESM achieves state-of-the-art or comparable performance on eight benchmark graph datasets.


[Zoom join link](https://zoom.us/j/96713836180)

Abstract 76: (#86 / Sess. 2) Graph Generation with Energy-Based Models in Graph Representation Learning and Beyond (GRL+), *Liu* N/A

We present a set of novel, energy-based models built on top of graph neural networks (GNN-EBMs) to estimate the unnormalized density of a distribution of graphs. GNN-EBMs can generate graphs implicitly via MCMC sampling. We compare the performance of GNN-EBMs trained using 3 different estimators: pseudolikelihood, conditional noise contrastive estimation, and persistent contrastive divergence (PCD). We find that all 3 estimators result in models that generalize well, while models trained with PCD generate samples that are competitive with state-of-the-art baselines. Finally, we discuss the potential of GNN-EBMs beyond generation for diverse tasks such as semi-supervised learning and outlier detection.


[Zoom join link](https://zoom.us/j/99890466920)

Abstract 77: (#43 / Sess. 2) Uncovering the Folding Landscape of RNA Secondary Structure with Deep Graph Embeddings in Graph Representation Learning and Beyond (GRL+), *Castro* N/A

Biomolecular graph analysis has recently gained much attention in the emerging field of geometric deep learning. Here we focus on organizing biomolecular graphs in ways that expose meaningful relations and variations between them. We propose a geometric scattering autoencoder (GSAE) network for learning such graph embeddings. Our embedding network first extracts rich graph features using the recently proposed geometric scattering transform. Then, it leverages a semi-supervised variational autoencoder to extract a low-dimensional embedding that retains the information in these features that enable prediction of molecular properties as well as characterize graphs. We show that GSAE organizes RNA graphs both by structure and energy, accurately reflecting bistable RNA structures. Also, the model is generative and can sample new folding trajectories.

[Teaser video](https://slideslive.com/38931487/uncovering-the-folding-landscape-of-rna-secondary-structure)

[Zoom join link](https://zoom.us/j/97816627053)

Abstract 78: (#51 / Sess. 1) Deep Lagrangian Propagation in Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), *Tiezzi* N/A

...
Graph Neural Networks (Scarselli et al., 2009) exploit an iterative diffusion procedure to compute the node states as the fixed point of the trainable state transition function. In this paper, we show how to cast this scheme as a constrained optimization problem, thus avoiding the unfolding procedure required for the computation of the fixed point. This is done by searching for saddle points of the Lagrangian function in the space of the weights, state variables and Lagrange multipliers. The proposed approach shows state-of-the-art performance in multiple standard benchmarks in graph domains.

[Zoom join link](https://zoom.us/j/97164886012)

Abstract 80: (#94 / Sess. 2) Are Hyperbolic Representations in Graphs Created Equal? In Graph Representation Learning and Beyond (GRL+), Kochurov N/A

Recently there was an increasing interest in applications of graph neural networks in non-Euclidean geometry; however, are non-Euclidean representations always useful for graph learning tasks? For different problems such as node classification and link prediction we compute hyperbolic embeddings and conclude that for tasks that require global prediction consistency it might be useful to use non-Euclidean embeddings, while for other tasks Euclidean models are superior. To do so we first fix an issue of the existing models associated with the optimization process at zero curvature. Current hyperbolic models deal with gradients at the origin in ad-hoc manner, which is inefficient and can lead to numerical instabilities. We solve the instabilities of kappa-Stereographic model at zero curvature cases and evaluate the approach of embedding graphs into the manifold in several graph representation learning tasks.

[Zoom join link](https://zoom.us/j/99884101698)

Abstract 83: (#2 / Sess. 1) When Spectral Domain Meets Spatial Domain in Graph Neural Networks in Graph Representation Learning and Beyond (GRL+), Balcilar N/A

Convolutional Graph Neural Networks (ConvGNNs) are designed either in the spectral domain or in the spatial domain. In this paper, we provide a theoretical framework to analyze these neural networks, by deriving some equivalence of the graph convolution processes, regardless if they are designed in the spatial or the spectral domain. We demonstrate the relevance of the proposed framework by providing a spectral analysis of the most popular ConvGNNs (ChebNet, CayleyNet, GCN and Graph Attention Networks), which allows to explain their performance and shows their limits.
ICML 2020 Workshop book

Abstract 84: (#9 / Sess. 1) Graph Neural Networks in TensorFlow and Keras with Spektral in Graph Representation Learning and Beyond (GRL+). Grattarola N/A

In this paper we present Spektral, an open-source Python library for building graph neural networks with TensorFlow and the Keras application programming interface. Spektral implements a large set of methods for deep learning on graphs, including message-passing and pooling operators, as well as utilities for processing graphs and loading popular benchmark datasets. The purpose of this library is to provide the essential building blocks for creating graph neural networks, focusing on the guiding principles of user-friendliness and quick prototyping on which Keras is based. Spektral is, therefore, suitable for absolute beginners and expert deep learning practitioners alike. In this work, we present an overview of Spektral's features and report the performance of the methods implemented by the library in scenarios of node classification, graph classification, and graph regression.

Abstract 85: (#28 / Sess. 1) Contrastive Graph Neural Network Explanation in Graph Representation Learning and Beyond (GRL+). Faber N/A

Graph Neural Networks achieve remarkable results on problems with structured data but come as black-box predictors. Transferring existing techniques, for example occlusion, to interpret models fails as even removing a single node or edge can lead to drastic changes in the graph. The resulting graphs can differ from all training examples, causing model confusion and wrong explanations. Thus, we argue that explicability must use graphs consistent with the distribution underlying the training data. We coin this property Distribution Compliant Explanation (DCE) and present a novel Contrastive GNN Explanation (CoGE) technique following this paradigm. An experimental study supports the efficacy of CoGE.

Abstract 86: (#18 / Sess. 1) Hierarchical Protein Function Prediction with Tail-GNNs in Graph Representation Learning and Beyond (GRL+). SpaleviAž N/A

Protein function prediction may be framed as predicting subgraphs (with certain closure properties) of a directed acyclic graph describing the hierarchy of protein functions. Graph neural networks (GNNs), with their built-in inductive bias for relational data, are hence naturally suited for this task. However, in contrast with most GNN applications, the graph is not related to the input, but to the label space. Accordingly, we propose Tail-GNNs, neural networks which naturally compose with the output space of any neural network for multi-task prediction, to provide relationally-reinforced labels. For protein function prediction, we combine a Tail-GNN with a dilated convolutional network which learns representations of the protein sequence, making significant improvement in F_1 score and demonstrating the ability of Tail-GNNs to learn useful representations of labels and exploit them in real-world problem solving.

Abstract 87: (#63 / Sess. 2) Stay Positive: Knowledge Graph Embedding Without Negative Sampling in Graph Representation Learning and Beyond (GRL+). Hajimoradiou N/A

Knowledge graphs (KGs) are typically incomplete and we often wish to infer new facts given the existing ones. This can be thought of as a binary classification problem; we aim to predict if new facts are true or false. Unfortunately, we generally only have positive examples (the known facts) but we also need negative ones to train a classifier. To resolve this, it is usual to generate negative examples using a negative sampling strategy. However, this can produce false negatives which may reduce performance, is computationally expensive, and does not produce calibrated classification probabilities. In this paper, we propose a training procedure that obviates the need for negative sampling by adding a spectral regularization term to the loss function. Our results for two relational embedding models (DistMult and SimplE) show the merit of our proposal both in terms of performance and speed.

Abstract 88: (#99 / Sess. 2) GraphNets with Spectral Message Passing in Graph Representation Learning and Beyond (GRL+). Stachenfeld N/A

Graph Neural Networks (GNNs) are the subject of intense focus by the machine learning community for problems involving relational reasoning. GNNs can be broadly divided into spatial and spectral approaches. Spatial approaches use a form of learned message-passing, in which interactions among vertices are computed locally, and information propagates over longer distances on the graph with greater numbers of message-passing steps. Spectral approaches use eigendecompositions of the graph Laplacian to produce a generalization of spatial convolutions to graph structured data which access information over short and long time scales simultaneously. Here we introduce a Spectral Graph Network, which applies message passing to both the spatial and spectral domains. Our model projects vertices of the spatial graph onto the Laplacian eigenvectors, which are each represented as vertices in a fully connected “spectral graph”, and then applies learned message passing to them. We apply this model to various benchmark tasks including a sparse graph-version of MNIST image classification, molecular classification (MoleculeNet), and molecular property prediction (QM9). The Spectral GN promotes efficient training, reaching high performance with fewer training iterations despite having more parameters. The model also provides robustness to edge dropout and outperforms baselines for the classification tasks.
Over the years, ML models have steadily grown in complexity, gaining predictivity often at the expense of interpretability. An active research area called explainable AI (or XAI) has emerged with the goal to produce models that are both predictive and understandable. XAI has reached important successes, such as robust heatmap-based explanations of DNN classifiers. From an application perspective, there is now a need to massively engage into new scenarios such as explaining unsupervised / reinforcement learning, as well as producing explanations that are optimally structured for the human. In particular, our planned workshop will cover the following topics:

- Explaining beyond DNN classifiers: random forests, unsupervised learning, reinforcement learning
- Explaining beyond heatmaps: structured explanations, Q/A and dialog systems, human-in-the-loop
- Explaining beyond explaining: Improving ML models and algorithms, verifying ML, getting insights

XAI has received an exponential interest in the research community, and awareness of the need to explain ML models have grown in similar proportions in industry and in the sciences. With the sizable XAI research community that has formed, there is now a key opportunity to achieve this push towards successful applications. Our hope is that our proposed XXAI workshop can accelerate this process, foster a more systematic use of XAI to produce improvement on models in applications, and finally, also serves to better identify in which way current XAI methods need to be improved and what kind of theory of XAI is needed.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:45 AM</td>
<td>Opening Remarks</td>
<td>Samek</td>
</tr>
<tr>
<td>12:00 AM</td>
<td>Invited Talk 1: Scott Lundberg - From local explanations to global understanding with trees</td>
<td>Samek</td>
</tr>
<tr>
<td>12:25 AM</td>
<td>Invited Talk 1 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>12:30 AM</td>
<td>Invited Talk 2: Bolei Zhou - Interpreting and Leveraging the Latent Semantics in Deep Generative Models</td>
<td>Samek</td>
</tr>
<tr>
<td>12:55 AM</td>
<td>Invited Talk 2 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>01:00 AM</td>
<td>Contributed Talk 1: Sun et al. - Understanding Image Captioning Models beyond Visualizing Attention</td>
<td>Samek</td>
</tr>
<tr>
<td>01:15 AM</td>
<td>Contributed Talk 2: Karimi et al. - Algorithmic recourse under imperfect causal knowledge: a probabilistic approach</td>
<td>Samek</td>
</tr>
<tr>
<td>01:30 AM</td>
<td>Poster Session 1</td>
<td>Samek</td>
</tr>
<tr>
<td>03:00 AM</td>
<td>Invited Talk 3: Grégoire Montavon - XAI Beyond Classifiers: Explaining Anomalies, Clustering, and More</td>
<td>Samek</td>
</tr>
<tr>
<td>03:25 AM</td>
<td>Invited Talk 3 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>03:30 AM</td>
<td>Invited Talk 4: Zeynep Akata - Modelling Conceptual Understanding Through Communication</td>
<td>Samek</td>
</tr>
<tr>
<td>03:55 AM</td>
<td>Invited Talk 4 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>04:00 AM</td>
<td>Invited Talk 5: Sepp Hochreiter - XAI and Strategy Extraction via Reward Redistribution</td>
<td>Samek</td>
</tr>
<tr>
<td>04:25 AM</td>
<td>Invited Talk 5 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>04:30 AM</td>
<td>Invited Talk 6: Ribana Roscher - Use of Explainable Machine Learning in the Sciences</td>
<td>Samek</td>
</tr>
<tr>
<td>04:55 AM</td>
<td>Invited Talk 6 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>05:00 AM</td>
<td>Invited Talk 7: Adrian Weller &amp; Umang Bhatt - Challenges in Deploying Explainable Machine Learning</td>
<td>Samek</td>
</tr>
<tr>
<td>05:25 AM</td>
<td>Invited Talk 7 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>05:30 AM</td>
<td>Invited Talk 8: Osbert Bastani - Interpretable, Robust, and Verifiable Reinforcement Learning</td>
<td>Samek</td>
</tr>
<tr>
<td>05:55 AM</td>
<td>Invited Talk 8 Q&amp;A</td>
<td>Samek</td>
</tr>
<tr>
<td>06:00 AM</td>
<td>Contributed Talk 3: Anders et al. - XAI for Analyzing and Unlearning Spurious Correlations in ImageNet</td>
<td>Samek</td>
</tr>
<tr>
<td>06:15 AM</td>
<td>Contributed Talk 4: Yau et al. - What did you think would happen? Explaining Agent Behaviour through Intended Outcomes</td>
<td>Samek</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Poster Session 2</td>
<td>Samek</td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Closing Remarks</td>
<td>Samek</td>
</tr>
</tbody>
</table>
Abstract 1: Invited Talk 1: Scott Lundberg - From local explanations to global understanding with trees in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers. Samek 12:00 AM

Tree-based machine learning models are popular nonlinear predictive models, yet comparatively little attention has been paid to explaining their predictions. In this talk I will explain how to improve their interpretability through the combination of many local game-theoretic explanations. I'll show how combining many high-quality local explanations allows us to represent global structure while retaining local faithfulness to the original model. This will enable us to identify high-magnitude but low-frequency nonlinear mortality risk factors in the US population, to highlight distinct population subgroups with shared risk characteristics, and to identify nonlinear interaction effects among risk factors for chronic kidney disease, and to monitor a machine learning model deployed in a hospital by identifying which features are degrading the model’s performance over time.


Recent progress in deep generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) has enabled synthesizing photo-realistic images, such as faces and scenes. However, it remains much less explored on what has been learned inside the deep representations learned from synthesizing images. In this talk, I will present some of our recent progress in interpreting the semantics in the latent space of the GANs, as well as reversing real images back into the latent space. Identifying these semantics not only allows us to better understand the internal mechanism in generative models, but also facilitates versatile real image editing applications.

Abstract 3: Contributed Talk 1: Sun et al. - Understanding Image Captioning Models beyond Visualizing Attention in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers. Samek 01:00 AM

This paper explains predictions of image captioning attention models beyond visualizing the attention itself. In this paper, we develop variants of layer-wise relevance backpropagation (LRP) tailored to image captioning models with attention mechanisms. We show that the explanations, firstly, correlate to object locations with higher precision than attention, secondly, identify object words that are unsupported by image content, and thirdly, provide guidance to improve the model. Results are reported using two different image captioning attention models trained with Flickr30K and MSCOCO2017 datasets. Experimental analyses show the strength of explanation methods for understanding image captioning attention models.


Recent work has discussed the limitations of counterfactual explanations to recommend actions for algorithmic recourse and argued for the need of taking causal relationships between features into consideration. Unfortunately, in practice, the true underlying structural causal model is generally unknown. In this work, we first show that it is impossible to guarantee recourse without access to the true structural equations. To address this limitation, we propose two probabilistic approaches to select optimal actions that achieve recourse with high probability given limited causal knowledge (e.g., only the causal graph). The first captures uncertainty over structural equations under additive Gaussian noise and uses Bayesian model averaging to estimate the counterfactual distribution. The second removes any assumptions on the structural equations by instead computing the average effect of recourse actions on individuals similar to the person who seeks recourse, leading to a novel subpopulation-based interventional notion of recourse. We derive a gradient-based procedure for selecting optimal recourse actions and empirically show that the proposed approaches lead to more reliable recommendations under imperfect causal knowledge than non-probabilistic baselines.

Abstract 5: Poster Session 1 in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers. Samek 01:30 AM

- Sun et al. “Understanding Image Captioning Models beyond Visualizing Attention”
- Sun et al. “Explain and Improve: Cross-Domain-Few-Shot-Learning Using Explanations”
- Manupriya et al. “SEA-NN: Submodular Ensembled Attribution for Neural Networks”

Abstract 6: Contributed Talk 3: Grégoire Montavon - XAI Beyond Classifiers: Explaining Anomalies, Clustering, and More in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers. Samek 03:00 AM

Unsupervised models such as clustering or anomaly detection are routinely used for data discovery and summarization. To gain maximum insight from the data, we also need to explain which input features (e.g. pixels) support the cluster assignments and the anomaly detections. So far, XAI has mainly focused on supervised models. In this talk, a novel systematic approach to explain various unsupervised models is presented. The approach is based on finding, without retraining, neural network equivalents of these models. Their predictions can then be readily explained using common XAI procedures developed for neural networks.


Zoom Room 2

- Macdonald et al. "Explaining Neural Network Decisions Is Hard"
- Molnar et al. "Pitfalls to Avoid when Interpreting Machine Learning Models"

Zoom Room 1

- Sun et al. "Understanding Image Captioning Models beyond Visualizing Attention"
- Sun et al. "Explain and Improve: Cross-Domain-Few-Shot-Learning Using Explanations"
- Manupriya et al. "SEA-NN: Submodular Ensembled Attribution for Neural Networks"
An agent who interacts with a wide population of other agents needs to be aware that there may be variations in their understanding of the world. Furthermore, the machinery which they use to perceive may be inherently different, as is the case between humans and machines. In this work, we present both an image reference game between a speaker and a population of listeners where reasoning about the concepts other agents can comprehend is necessary and a model formulation with this capability. We focus on reasoning about the conceptual understanding of others, as well as adapting to novel gameplay partners and dealing with differences in perceptual machinery. Our experiments on three benchmark image/attribute datasets suggest that our learner indeed encodes information directly pertaining to the understanding of other agents, and that leveraging this information is crucial for maximizing gameplay performance.


Assigning credit for a received reward to previously performed actions is one of the central tasks in reinforcement learning. Credit assignment often uses world models, either in a forward or in a backward view. In a forward view, the future return is estimated by replacing the environment through a model or by rolling out sequences until episode end. A backward view either learns a backward model or performs a backward analysis of a forward model that predicts or models the return of an episode. Our method RUDDER performs a backward analysis to construct a reward redistribution to credit those actions that caused a reward. Its extension Align-RUDDER learns a reward redistribution from few demonstrations. An optimal reward redistribution has zero expected future reward and, therefore, immediately credits actions for all they will cause. XAI aims at credit assignment, too, when asking what caused a model to produce a particular output given an input. Even further, XAI wants to know how and why a policy solved a task, why an agent is better than humans, why a decision was made. Humans best comprehend a strategy of an agent if all its actions are immediately evaluated and do not have hidden consequences in the future. Reward redistributions learned by RUDDER and Align-RUDDER help to understand task-solving strategies of both humans and machines.


For some time now, machine learning methods have been indispensable in many application areas. Especially with the recent development of neural networks, these methods are increasingly used in the sciences to obtain scientific outcomes from observational or simulated data. Besides a high accuracy, a desired goal is to learn explainable models. In order to reach this goal and obtain explanation, knowledge from the respective domain is necessary, which can be integrated into the model or applied post-hoc. This talk focuses on explainable machine learning approaches which are used to tackle common challenges in the sciences such as the provision of reliable and scientific consistent results. It will show that recent advances in machine learning to enhance transparency, interpretability, and explainability are helpful in overcoming these challenges.

Abstract 17: Invited Talk 7: Adrian Weller & Umang Bhatt - Challenges in Deploying Explainable Machine Learning in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers, Sameek 05:00 AM

Explainable machine learning offers the potential to provide stakeholders with insights into model behavior, yet there is little understanding of how organizations use these methods in practice. In this talk, we discuss recent research exploring how organizations view and use explainability. We find that the majority of deployments are not for end-users but rather for machine learning engineers, who use explainability to debug the model. There is thus a gap between explainability in practice and the goal of external transparency since explanations are primarily serving internal stakeholders. Providing useful external explanations requires careful consideration of the needs of stakeholders, including end-users, regulators, and domain experts. Despite this need, little work has been done to facilitate inter-stakeholder conversation around explainable machine learning. To help address this gap, we report findings from a closed-door, day-long workshop between academics, industry experts, legal scholars, and policymakers to develop a shared language around explainability and to understand the current shortcomings of and potential solutions for deploying explainable machine learning in the service of external transparency goals.

Abstract 19: Invited Talk 8: Osbert Bastani - Interpretable, Robust, and Verifiable Reinforcement Learning in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers, Sameek 05:30 AM

Structured control policies such as decision trees, finite-state machines, and programs have a number of advantages over more traditional models: they are easier for humans to understand and debug, they generalize more robustly to novel environments, and they are easier to formally verify. However, learning these kinds of models has proven to be challenging. I will describe recent progress learning structured policies, along with evidence demonstrating their benefits.

Abstract 21: Contributed Talk 3: Anders et al. - XAI for Analyzing and Unlearning Spurious Correlations in ImageNet in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers, Sameek 06:00 AM

Contemporary learning models for computer vision are typically trained on very large data sets with millions of samples. There may, however, be biases, artifacts, or errors in the data that have gone unnoticed and are exploitable by the model, which in turn becomes a biased “Clever-Hans” predictor. In this paper, we contribute by providing a comprehensive analysis framework based on a scalable statistical analysis of attributions from explanation methods for large data corpora, here ImageNet. Based on Spectral Relevance Analysis we propose the following technical contributions and resulting findings: (a) a scalable quantification of artificial classes where the ML models under study exhibit Clever-Hans behavior, (b) an approach denoted as Class-Artifact Compensation (CArC) that allows to fine-tune an existing model to effectively eliminate its focus on artifacts and biases yielding significantly reduced Clever-Hans behavior.


We present a novel form of explanation for Reinforcement Learning (RL), based around the notion of intended outcome. This describes what
outcome an agent is trying to achieve by its actions. Given this definition, we provide a simple proof that general methods for post-hoc explanations of this nature are impossible in traditional reinforcement learning. Rather, the information needed for the explanations must be collected in conjunction with training the agent. We provide approaches designed to do this for several variants of Q-function approximation and prove consistency between the explanations and the Q-values learned. We demonstrate our method on multiple reinforcement learning problems.

Abstract 23: Poster Session 2 in XXAI: Extending Explainable AI Beyond Deep Models and Classifiers, Samek 06:30 AM

Zoom Room 1

- Bhatt et al. "Machine Learning Explainability for External Stakeholders"
- Karimi et al. "Algorithmic recourse under imperfect causal knowledge: a probabilistic approach"
- Wang et al. "Towards Probabilistic Sufficient Explanations"

Zoom Room 2

- Agarwal et al. "Neural Additive Models: Interpretable Machine Learning with Neural Nets"
- Alaniz et al. "Learning Decision Trees Recurrently through Communication"
- Anders et al. "XAI for Analyzing and Unlearning Spurious Correlations in ImageNet"

Zoom Room 3

- Dasgupta et al. "Explainable k-Means Clustering: Theory and Practice"
- Dhurandhar et al. "Leveraging Simple Model Predictions for Enhancing its Performance"
- Brophy and Lowd: "TREX: Tree-Ensemble Representer-Point Explanations"

Zoom Room 4

- Lin et al. "Contrastive Explanations for Reinforcement Learning via Embedded Self Predictions"
- Yau et al. "What Did You Think Would Happen? Explaining Agent Behaviour through Intended Outcomes"
- Danesh et al. "Understanding Finite-State Representations of Recurrent Policy Networks"

Zoom Room 5

- Quint et al. "Contrastive Attribution with Feature Visualization"
- Zhao "Fast Real-time Counterfactual Explanations"
- Chrysos et al. "Unsupervised Controllable Generation with Self-Training"

[link](https://us02web.zoom.us/j/81618586384?pwd=c2wzaldJaVhhdpjKyUeFFic1Ndaza0dz09)
Even though supervised learning using large annotated corpora is still the dominant approach in machine learning, self-supervised learning is gaining considerable popularity. Applying self-supervised learning to audio and speech sequences, however, remains particularly challenging. Speech signals, in fact, are not only high-dimensional, long, and variable-length sequences, but also entail a complex hierarchical structure that is difficult to infer without supervision (e.g., phonemes, syllables, words). Moreover, speech is characterized by an important variability due to different speaker identities, accents, recording conditions and noises that highly increase the level of complexity.

We believe that self-supervised learning will play a crucial role in the future of artificial intelligence, and we think that great research effort is needed to efficiently take advantage of it in audio and speech applications. With our initiative, we wish to foster more progress in the field, and we hope to encourage a discussion amongst experts and practitioners from both academia and industry that might bring different points of view on this topic. Furthermore, we plan to extend the debate to multiple disciplines, encouraging discussions on how insights from other fields (e.g., computer vision and robotics) can be applied to speech, and how findings on speech can be used on other sequence processing tasks. The workshop will be conceived to promote communication and exchange of ideas between machine learning and speech communities. Throughout a series of invited talks, contributed presentations, poster sessions, as well as a panel discussion we want to foster a fruitful scientific discussion that cannot be done with that level of detail during the main ICML conference.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:05 AM</td>
<td>Opening Remarks</td>
<td>Ravanelli</td>
</tr>
<tr>
<td>12:15 AM</td>
<td>Invited Talk: Representation learning on sequential data with latent priors</td>
<td>Chorowski</td>
</tr>
<tr>
<td>12:40 AM</td>
<td>Invited Talk: Contrastive Predictive Coding for audio representation learning</td>
<td>van den Oord</td>
</tr>
<tr>
<td>01:05 AM</td>
<td>Q&amp;A Invited Talks</td>
<td></td>
</tr>
<tr>
<td>01:30 AM</td>
<td>Adversarial representation learning for private speech generation</td>
<td>Ericsson</td>
</tr>
<tr>
<td>01:45 AM</td>
<td>Investigating self-supervised pre-training for end-to-end speech translation</td>
<td>Nguyen</td>
</tr>
</tbody>
</table>
| 02:00 AM | Analysis of Predictive Coding Models for phonemic representation learning in small datasets | Cruz Bland
| 02:15 AM | COALA: Co-Aligned autoencoders for learning semantically enriched audio representations   | Favory            |
| 02:30 AM | Understanding self-attention of self-supervised audio transformers                        | Yang              |
| 02:45 AM | Q&A Contributed Talks                                                                     |                   |
| 04:00 AM | Invited Talk: Denoising and real-vs-corrupted classification as two fundamental paradigms in self-supervised learning | Hyvarinen         |
| 04:25 AM | Invited Talk: Unsupervised pre-training of bidirectional speech encoders via masked reconstruction | Livescu          |
| 04:50 AM | Q&A Invited Talks                                                                         |                   |
| 05:15 AM | Learning Speech representations from raw audio by joint audiovisual self-supervision      | Shukla            |
| 05:30 AM | OtoWorld: Towards learning to separate by learning to move                                 | Ranadive          |
| 05:45 AM | Language Agnostic speech embeddings for emotion classification                            | Nandan            |
| 06:00 AM | End-to-end ASR: from supervised to semi-supervised learning with modern architectures     | Kahn              |
| 06:15 AM | Using self-supervised learning of birdsong for downstream industrial audio classification | Ryan              |
| 06:30 AM | Q&A Contributed Talks                                                                     |                   |
| 06:55 AM | Invited Talk: self-supervised video models from sound and speech, Lorenzo Torresani       | Torresani         |
| 07:20 AM | Invited Talk: Sights and sounds in 3D spaces                                               | Grauman           |
As more and more data is collected in various settings across organizations, companies, and countries, there has been an increase in the demand of user privacy. Developing privacy preserving methods for data analytics is thus an important area of research. In this work we present a model based on generative adversarial networks (GANs) that learns to obfuscate specific sensitive attributes in speech data. We train a model that learns to hide sensitive information in the data, while preserving the meaning in the utterance. The model is trained in two steps: first to filter sensitive information in the spectrogram domain, and then to generate new and private information independent of the filtered one. The model is based on a U-Net CNN that takes mel-spectrograms as input. A MelGAN is used to invert the spectrograms back to raw audio waveforms. We show that it is possible to hide sensitive information such as gender by generating new data, trained adversarially to maintain utility and realism.

Link to the video: https://slideslive.com/38930731/adversarial-representation-learning-for-private-speech-generation

Abstract 6: Investigating Self-supervised Pre-training for End-to-end Speech Translation in Self-supervision in Audio and Speech, Nguyen 01:45 AM

Self-supervised learning from raw speech has been proven beneficial to improve automatic speech recognition (ASR). We investigate here its impact on end-to-end automatic speech translation (AST) performance. We use a contrastive predictive coding (CPC) model pre-trained from unlabeled speech as a feature extractor for a downstream AST task. We show that self-supervised pre-training is particularly efficient in low resource settings and that fine-tuning CPC models on the AST training data further improves performance. Even in higher resource settings, ensembling AST models trained with filter-bank and CPC representations leads to near state-of-the-art models without using any ASR pre-training. This might be particularly beneficial when one needs to develop a system that translates from speech in a language with poorly standardized orthography or even from speech in an unwritten language.

Link to the video: https://slideslive.com/38930733/investigating-selfsupervised-pretraining-for-endtoend-speech-

Abstract 7: Analysis of Predictive Coding Models for Phonemic Representation Learning in Small Datasets in Self-supervision in Audio and Speech, Cruz Blandón 02:00 AM

Neural network models using predictive coding are interesting from the viewpoint of computational modelling of human language acquisition, where the objective is to understand how linguistic units could be learned from speech without any labels. Even though several promising predictive coding-based learning algorithms have been proposed in the literature, it is currently unclear how well they generalise to different languages and training dataset sizes. In addition, despite that such models have shown to be effective phonemic feature learners, it is unclear whether minimisation of the predictive loss functions of these models also leads to optimal phoneme-like representations. The present study investigates the behaviour of two predictive coding models, Autoregressive Predictive Coding and Contrastive Predictive Coding, in a phoneme discrimination task (ABX task) for two languages with different dataset sizes. Our experiments show a strong correlation between the autoregressive loss and the phoneme discrimination scores with the two datasets. However, to our surprise, the CPC model shows rapid convergence already after one pass over the training data, and, on
average, its representations outperform those of APC on both languages.

Link to the video:
https://slideslive.com/38930734/analysis-of-predictive-coding-models-for-phonemic-representation

Abstract 8: COALA: Co-Aligned Autoencoders for Learning Semantically Enriched Audio Representations in Self-supervision in Audio and Speech, Favory 02:15 AM

Audio representation learning based on deep neural networks (DNNs) emerged as an alternative approach to hand-crafted features. For achieving high performance, DNNs often need a large amount of annotated data which can be difficult and costly to obtain. In this paper, we propose a method for learning audio representations, aligning the learned latent representations of audio and associated tags. Aligning is done by maximizing the agreement of the latent representations of audio and tags, using a contrastive loss. The result is an audio embedding model which reflects acoustic and semantic characteristics of sounds. We evaluate the quality of our embedding model, measuring its performance as a feature extractor on three different tasks (namely, sound event recognition, and music genre and musical instrument classification), and investigate what type of characteristics the model captures. Our results show that our method is in par with the state-of-the-art in the considered tasks and the embeddings produced with our method are well correlated with some acoustic descriptors.

Link to the video:
https://slideslive.com/38930732/coala-coaligned-autoencoders-for-learning-semanticaily-enriched-audio-representations

Abstract 9: Understanding Self-Attention of Self-Supervised Audio Transformers in Self-supervision in Audio and Speech, Yang 02:30 AM

Self-supervised Audio Transformers (SAT) enable great success in many downstream speech applications like ASR, but how they work has not been widely explored yet. In this work, we present multiple strategies for the analysis of attention mechanisms in SAT. We categorize attentions into explainable categories, where we discover each category possesses its own unique functionality. We provide a visualization tool for understanding multi-head self-attention, importance ranking strategies for identifying critical attention, and attention refinement techniques to improve model performance.

Link to the video:
https://slideslive.com/38930730/understanding-selfattention-of-selfsupervised-audio-transformers

Abstract 11: Invited Talk: Denoising and real-vs-corrupted classification as two fundamental paradigms in self-supervised learning in Self-supervision in Audio and Speech, Hyvarinen 04:00 AM

The basic idea in self-supervised learning (SSL) is to turn an unsupervised learning task into a supervised task, and use well-known supervised methods to solve it. Even though the data initially has no labels or targets to enable supervised learning, we artificially define a “pretext” supervised task, with some labels or targets of our choosing. Here, I focus on two widely-used and fundamental paradigms for SSL. First, adding Gaussian noise to the data and then learning to denoise it, is a special case of the more general SSL principle of corrupting the data and learning to repair it. Second, classification can be used for SSL by first corrupting the data and then learning to discriminate between the original data and the corrupted version; in the extreme case, this means learning to discriminate between the data and pure noise. While these are very intuitive principles, a sophisticated theoretical analysis is possible in both cases. In particular, deep connections to energy-based modeling and independent component analysis can be shown.

Link to the video:

Abstract 12: Invited Talk: Unsupervised pre-training of bidirectional speech encoders via masked reconstruction in Self-supervision in Audio and Speech, Livescu 04:25 AM

We propose an approach for pre-training speech representations via a masked reconstruction loss. Our pre-trained encoder networks are bidirectional and can therefore be used directly in typical bidirectional speech recognition models. The pre-trained networks can then be fine-tuned on a smaller amount of labelled data for speech recognition. In addition, we address the problem of domain differences between the pre-training and fine-tuning data, by adding an explicit adaptation layer during fine-tuning. Experiments with this approach on the LibriSpeech and Wall Street Journal corpora show promising results. The gain from pre-training is additive to that from supervised data augmentation.

Link to the video:

Abstract 14: Learning Speech Representations from Raw Audio by Joint Audiovisual Self-Supervision in Self-supervision in Audio and Speech, Shukla 05:15 AM

The intuitive interaction between the audio and visual modalities is valuable for cross-modal self-supervised learning. This concept has been demonstrated for generic audiovisual tasks like video action recognition and acoustic scene classification. However, self-supervision remains under-explored for audiovisual speech. We propose a method to learn self-supervised speech representations from the raw audio waveform. We train a raw audio encoder by combining audio-only self-supervision (by predicting informative audio attributes) with visual self-supervision (by generating talking faces from audio). The visual pretext task drives the audio representations to capture information related to lip movements. This enriches the audio encoder with visual information and the encoder can be used for evaluation without the visual modality. Our method attains competitive performance with respect to existing self-supervised audio features on established isolated word classification benchmarks, and significantly outperforms other methods at learning from fewer labels. Notably, our method also outperforms fully supervised training, thus providing a strong initialization for speech related tasks. Our results demonstrate the potential of multimodal self-supervision in audiovisual speech for learning good audio representations.

Link to the video:

Abstract 15: OtoWorld: Towards Learning to Separate by Learning to Move in Self-supervision in Audio and Speech, Ranadive 05:30 AM

We present OtoWorld, an interactive environment in which agents must learn to listen in order to solve navigational tasks. The purpose of OtoWorld is to facilitate reinforcement learning research in computer audition, where agents must learn to listen to the world around them to navigate. OtoWorld is built on three open source libraries: OpenAI Gym
for environment and agent interaction, PyRoomAcoustics for ray-tracing and acoustics simulation, and nussl for training deep computer audition models. OtoWorld is the audio analogue of GridWorld, a simple navigation game. OtoWorld can be easily extended to more complex environments and games. To solve one episode of OtoWorld, an agent must move towards each sounding source in the auditory scene and “turn it off”. The agent receives no other input than the current sound of the room. The sources are placed randomly within the room and can vary in number. The agent receives a reward for turning off a source. We present preliminary results on the ability of agents to win at OtoWorld. OtoWorld is open-source and available.

Link to the video: https://slideslive.com/38930738/otoworld-toward-learning-to-separate-by-learning-to-move

Abstract 16: Language Agnostic Speech Embeddings for Emotion Classification in Self-supervision in Audio and Speech, Nandan 05:45 AM

In this paper, we propose a technique for learning speech representations or embeddings in a self supervised manner, and show their performance on emotion classification task. We also investigate the usefulness of these embeddings for languages different from the pretraining corpus. We employ a convolutional encoder model and contrastive loss function on augmented Log Mel spectrograms to learn meaningful representations from an unlabelled speech corpus. Emotion classification experiments are carried out on SAVEE corpus, German EmoDB, and CaFE corpus. We find that: (1) These pretrained embeddings perform better than MFCCs, openSMILE features and PASE+ encodings for emotion classification task. (2) These embeddings improve accuracies in emotion classification task on languages different from that used in pretraining thus confirming language agnostic behaviour.

Link to the video: https://slideslive.com/38930739/language-agnostic-speech-embeddings-for-emotion-classification

Abstract 17: End-to-End ASR: from Supervised to Semi-Supervised Learning with Modern Architectures in Self-supervision in Audio and Speech, Kahn 06:00 AM

We study pseudo-labeling for the semi-supervised training of ResNet, Time-Depth Separable ConvNets, and Transformers for speech recognition, with either CTC or Seq2Seq loss functions. We perform experiments on the standard Librispeech dataset, and leverage additional unlabeled data from Librivox through pseudo-labeling. We show that while Transformer-based acoustic models have superior performance with the supervised dataset alone, semi-supervision improves all models across architectures and loss functions and bridges much of the performance gaps between them. In doing so, we reach a new state-of-the-art for end-to-end acoustic models decoded with an external language model in the standard supervised learning setting, and a new absolute state-of-the-art with semi-supervised training. Finally, we study the effect of leveraging different amounts of unlabeled audio, propose several ways of evaluating the characteristics of unlabeled audio which improve acoustic modeling, and show that acoustic models trained with more audio rely less on external language models.

Link to the video: https://slideslive.com/38930740/endtoend-asr-from-supervised-to-semisupervised-learning-with-modern-architectures

Abstract 18: Using Self-Supervised Learning of Birdsong for Downstream Industrial Audio Classification in Self-supervision in Audio and Speech, Ryan 06:15 AM

In manufacturing settings, workers rely on their sense of hearing, and their knowledge of what sounds correct to help them identify machine quality problems based on the sound pitch, rhythm, timbre and other characteristics of machine operation. Using Machine Learning to classify these sounds has broad applications to automate the manual quality recognition work currently being done, including automating machine operator training, automating quality control detection, and diagnostics across manufacturing and mechanical service industries. We previously established that models taking input pitch information from music can dramatically improve classification model performance on industrial machine audio leveraging the CREPE pretrained pitch model. In this work we explore the use of self-supervised learning on pitch-intensive birdsong rather than the CREPE model. To reduce our reliance on a pretrained pitch model and reduce the quantity of labeled industrial audio required, we implement self-supervised representation learning using plentiful, license-free unlabeled, pitch intensive wild birdsong recordings, with audio data augmentation to perform classification on industrial audio. We show that: 1. We can preprocess the unlabeled birdsong data sample with unsupervised methods to eliminate low signal sample and mask low frequency noise leaving just desirable chirp-rich sample. 2. We can identify effective representations and approaches for learning birdsong pitch content by comparing select self-supervised pretext task training of temporal sequence prediction and sequence generation. 3. We can identify effective augmentation methods for learning pitch through comparison of the impact of a variety of audio data augmentation methods on self-supervised learning. And 4. Downstream fine-tuned models deliver improved performance classifying industrial motor audio. We demonstrate that motorized sound classification models using self-supervised learning with a dataset of pitch intensive birdsong, combined with select data augmentation, achieves better results than using the pre-trained CREPE pitch model.

Link to the video: https://slideslive.com/38930741/selfsupervised-video-models-from-sound-and-speech

Abstract 20: Invited Talk: Self-Supervised Video Models from Sound and Speech, Lorenzo Torresani in Self-supervision in Audio and Speech, Torresani 06:55 AM

Existing manually-annotated datasets for video understanding differ substantially in their label spaces. Coupled with the limited sizes of these collections, this causes fully-supervised video models to transfer poorly across datasets and tasks.

Link to the video: https://slideslive.com/38930742/selfsupervised-video-models-from-sound-and-speech

Abstract 21: Invited Talk: Sights and sounds in 3D spaces in Self-supervision in Audio and Speech, Grauman 07:20 AM


Abstract 22: Invited Talk: Self-supervised learning of speech representations with wav2vec in Self-supervision in Audio and Speech, Beskow 07:45 AM
We show for the first time that learning powerful representations from speech audio alone followed by fine-tuning on transcribed speech can outperform the best semi-supervised methods while being conceptually simpler. wav2vec 2.0 masks the speech input in the latent space and solves a contrastive task defined over a quantization of the latent representations which are jointly learned. We set a new state of the art on both the 100 hour subset of LibriSpeech as well as on TIMIT phoneme recognition. When lowering the amount of labeled data to one hour, our model outperforms the previous state of the art on the 100 hour subset while using 100 times less labeled data. Using just ten minutes of labeled data and pre-training on 539 hours of unlabeled data still achieves 5.7/10.1 WER on the noisy/clean test sets of LibriSpeech. This demonstrates the feasibility of speech recognition with limited amounts of labeled data. Fine-tuning on all of LibriSpeech achieves 1.9/3.5 WER using a simple baseline model architecture.

Link to the video:

Abstract 24: Self-supervised Pitch Detection by Inverse Audio Synthesis in Self-supervision in Audio and Speech, Engel 08:40 AM

Audio scene understanding, parsing sound into a hierarchy of meaningful parts, is an open problem in representation learning. Sound is a particularly challenging domain due to its high dimensionality, sequential dependencies and hierarchical structure. Differentiable Digital Signal Processing (DDSP) greatly simplifies the forward problem of generating audio by introducing differentiable synthesizer and effects modules that combine strong signal priors with end-to-end learning. Here, we focus on the inverse problem, inferring synthesis parameters to approximate an audio scene. We demonstrate that DDSP modules can enable a new approach to self-supervision, generating synthetic audio with differentiable synthesizers and training feature extractor networks to infer the synthesis parameters. By building a hierarchy from sinusoidal to harmonic representations, we show that it is possible to use such an inverse modeling approach to disentangle pitch from timbre, an important task in audio scene understanding.

Link to the video:


Supervised approaches to single-channel speech separation rely on synthetic mixtures, so that the individual sources can be used as targets. Good performance depends upon how well the synthetic mixture data match real mixtures. However, matching synthetic data to the acoustic properties and distribution of sounds in a target domain can be challenging. Instead, we propose an unsupervised method that requires only single-channel acoustic mixtures, without ground-truth source signals. In this method, existing mixtures are mixed together to form a mixture of mixtures, which the model separates into latent sources. We propose a novel loss that allows the latent sources to be remixed to approximate the original mixtures. Experiments show that this method can achieve competitive performance on speech separation compared to supervised methods. In a semi-supervised learning setting, our method enables domain adaptation by incorporating unsupervised mixtures from a matched domain. In particular, we demonstrate that significant improvement to reverberant speech separation performance can be achieved by incorporating reverberant mixtures.

Link to the video:

Abstract 26: Bootstrapping Unsupervised Deep Music Separation from Primitive Auditory Grouping Principles in Self-supervision in Audio and Speech, Seetharaman 09:10 AM

Separating an audio scene, such as a cocktail party with multiple overlapping voices, into meaningful components (e.g., individual voices) is a core task in computer audition, analogous to image segmentation in computer vision. Deep networks are the state-of-the-art approach. They are typically trained on synthetic audio mixtures made from isolated sound source recordings so that ground truth for the separation is known. However, the vast majority of available audio is not isolated. The human brain performs an initial segmentation of the audio scene using primitive cues that are broadly applicable to many kinds of sound sources. We present a method to train a deep source separation model in an unsupervised way by bootstrapping using multiple primitive cues. We apply our method to train a network on a large set of unlabeled music recordings to separate vocals from accompaniment without the need for ground truth isolated sources or artificial training mixtures. A companion notebook with audio examples and code for experiments is available: https://github.com/pseeth/bootstrapping-computer-audition.

Link to the video:

5th ICML Workshop on Human Interpretability in Machine Learning (WHI)

Adrian Weller, Alice Xiang, Amit Dhurandhar, Been Kim, Dennis Wei, Kush Varshney, Umang Bhatt

Fri Jul 17, 01:00 AM

This workshop will bring together artificial intelligence (AI) researchers who study the interpretability of AI systems, develop interpretable machine learning algorithms, and develop methodology to interpret black-box machine learning models (e.g., post-hoc interpretations). This is a very exciting time to study interpretable machine learning, as the advances in large-scale optimization and Bayesian inference that have enabled the rise of black-box machine learning are now also starting to be exploited to develop principled approaches to large-scale interpretable machine learning. Interpretability also forms a key bridge between machine learning and other AI research directions such as machine reasoning and planning. Participants in the workshop will exchange ideas on these and allied topics.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:30 AM</td>
<td>Contributed Talk 1 and 2</td>
</tr>
<tr>
<td>02:00 AM</td>
<td>Invited Talk: Sandra Wachter</td>
</tr>
<tr>
<td>02:30 AM</td>
<td>Spotlights</td>
</tr>
<tr>
<td>03:30 AM</td>
<td>Break</td>
</tr>
<tr>
<td>05:00 AM</td>
<td>Invited Talk: Finale Doshi-Velez</td>
</tr>
</tbody>
</table>
Law & Machine Learning

Céline Castets-Renard, Sylvain Cussat-Blanc, Laurent Risser

Fri Jul 17, 01:30 AM

Description: the workshop proposal in Law and Machine Learning aims to contribute to the research on social and legal risks of the deployment of AI systems using machine learning based decisions. Today, algorithms have been infiltrating and governing every aspect of our lives as individuals and as a society. Specifically, Algorithmic Decision Systems (ADS) are involved in many social decisions. For instance, such systems are increasingly used to support decision-making in fields, such as child welfare, criminal justice, school assignment, teacher evaluation, fire risk assessment, homelessness prioritization, healthcare, Medicaid benefit, immigration decision systems or risk assessment, and predictive policing, among other things. Law enforcement agencies are increasingly using facial recognition, algorithmic predictive policing systems to forecast criminal activity and allocate police resources. However, these predictive systems challenge fundamental rights and guarantees of the criminal procedure. For several years, numerous studies have revealed, social risks of ML, especially the risks of opacity, bias, manipulation of information.

While it is only the starting point of the deployment of such systems, more interdisciplinary research is needed. Our purpose is to contribute to this new field which brings together legal researchers, mathematicians and computer scientists, by bridging the gap between the performance of algorithmic systems and legal standards. For instance, notions like privacy or fairness are formulated in law, as well as mathematical definitions in computer science. However, the meaning and the impact of such requirements are not necessarily identical. Besides, legal norms to regulate AI systems appear in certain national laws but have to be relevant and compatible with technical requirements. Furthermore, these standards must be checked by legal experts and regulators, which presupposes that AI systems are sufficiently meaningful and transparent. These issues emerge in different topics, such as privacy in data analysis and fairness in algorithmic decision-making. The topic will cover the research that denounces the risks and, above all, multidisciplinary research that proposes solutions, especially legal and technical solutions.

Schedule

02:00 AM Workshop presentation by Castets-Renard, Cussat-Blanc, Risser

02:15 AM Professor Frederik Zuiderveen Borgesius (Amsterdam University & Radboud University): Legal Protection in Europe against Discrimination by Machine Learning systems

03:00 AM Talks 1: 5 talks of 15 minutes each

04:15 AM Live Q&A Castets-Renard, Cussat-Blanc, Risser

05:00 AM Lunch time

05:30 AM Live poster session

06:30 AM Professor Olivier Sylvain (Fordham Law School): Recovering Tech's Humanity

07:15 AM Live Q&A Castets-Renard, Cussat-Blanc, Risser

07:45 AM Talks 2: 5 talks of 15 minutes each

09:00 AM Live Q&A Castets-Renard, Cussat-Blanc, Risser

09:45 AM Live roundtable about the future of the workshop Law & Machine Learning Castets-Renard, Cussat-Blanc, Risser

Abstracts (3):

Abstract 3: Talks 1: 5 talks of 15 minutes each in Law & Machine Learning, 03:00 AM

- 10:00-10:15: Algorithmic Recourse: from Counterfactual Explanations to Interventions (Amir-Hossein Karimi, Isabel Valera, Bernhard Schölkopf)
- 10:30-10:45: Online publication of court records: circumventing the privacy-transparency trade-off (Tristan Allard, Louis Beziaud, Sébastien Gambs)
- 10:45-11:00: The Gap between Deep Learning and Law: Predicting Employment Notice (Jason T. Lam, Rohan Bhambhoria, David Liang, Xiaodan Zhu, Samuel Dahan)

Abstract 6: Live poster session in Law & Machine Learning, 05:30 AM

A virtual room was created to talk about each poster. Please go to the workshop website for the links
12:30-13:00: Live posters session 1
* Impact of Legal Requirements on Explainability in Machine Learning (Adrien Bibal, Michael Lognoul, Alexandre de Streel, Benoît FrÉnay)
* Content Moderation as Legal Compliance: Annotating Hate Speech Using Judicial Legal Frameworks for Natural Language Processing Tasks (Thales Bertaglia, Giovanni De Gregorio, Catalina Goanta, Jerry Spanaklis)
* Predicting Court Decisions for Alimony: Avoiding Extra-legal Factors in Decision made by Judges and Not Understandable AI Models (Fabrice Muhlenbach, Isabelle Sayn, Long Nguyen-Phuoc)
* Detecting and Explaining Unfairness in Consumer Contracts with Memory Networks (Federico Ruggeri, Francesca Lagioia, Marco Lippi, Paolo Torroni)
* A Causal Linear Model to Quantify Edge Unfairness for Unfair Edge Prioritization and Discrimination Removal (Pavan Ravishankar, Pranshu Malviya, Balaraman Ravindran)
* The interrelation between Data and AI Ethics in the context of Impact Assessments (Emre Kazim, Adrian Soares Koshiyama) - Poster - zoom link

13:00-13:30: Live posters session 2
* Accuracy Bounding: A Regulatory Path Forward for the Algorithmic Society (Aileen Nielsen)
* The AI Accident Network: Artificial Intelligence Liability Meets Network Theory (Anat Lior)
* Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices (Manish Raghavan, Solon Barocas, Jon Kleinberg, Karen Levy)
* Conceptualizing Facial Recognition Technology in its Technical and Legal Dimensions (Natalia Menendez)
* Facial Recognition: A cross-national Survey on Public Acceptance, Privacy, and Discrimination (Damian Borth, Lea Steinacker)

Abstract 9: Talks 2: 5 talks of 15 minutes each in Law & Machine Learning, 07:45 AM
* 14:45-15:00: Regulating Accuracy-Efficiency Trade-Offs in Distributed Machine Learning Systems (A. Feder Cooper, Karen Levy, Christopher De Sa)
* 15:00-15:15: Punishing Race, Poverty and Trauma: the Data behind predictive algorithms in the American justice system (Claire Boine, Jeffrey D. Krupa)
* 15:15-15:30 : Legal Risks of Adversarial Machine Learning Research (Ram Shankar Siva Kumar, Jonathon Penney, Bruce Schneier, Kendra Albert)
* 15:30-15:45: What Is a Proxy and Why Is It a Problem? (Margarita Boyarskaya, Solon Barocas, Hanna Wallach)
* 15:45-16:00: Formalizing Data Deletion in the Context of the Right to be Forgotten (Sanjam Garg, Shafi Goldwasser, Prashant Nalini Vasudevan)

Learning with Missing Values

Julie Josse, Jes Frellsen, Pierre-Alexandre Mattei, Gael Varoquaux

Fri Jul 17, 01:45 AM

Analysis of large amounts of data offers new opportunities to understand many processes better. Yet, data accumulation often implies relaxing acquisition procedures or compounding diverse sources, leading to many observations with missing features. From questionnaires to collaborative filtering, from electronic health records to single-cell analysis, missingness is everywhere at play and is rather the norm than the exception. Even â€œcleanedâ€ versions of incomplete data setsâ€ with all the unfortunate biases this cleaning process may have created.

Despite this ubiquity, tackling missing values is often overlooked. Handling missing values poses many challenges, and there is a vast literature in the statistical community, with many implementations available. Yet, there are still many open issues and the need to design new methods or to introduce new point of views: for missing values in a supervised-learning setting, in deep learning architectures, to adapt available methods for high dimensional observed data with different type of missing values, deal with feature mismatch and distribution mismatch. Missing data is one of the eight pillars of causal wisdom for Judea Pearl who brought graphical model reasoning to tackle some missing not at random values.

To the best of our knowledge, this is the first workshop at the major machine learning conferences focusing primarily on missing value problems in recent years. The goal of our workshop is to give more momentum and exposition to research on missing values, both theoretical and methodological, and emphasize the connections with other areas of machine learning (e.g. causal inference, generative modelling, uncertainty quantification, transfer learning, distributional shift, etc.). We will also attach importance to discussing the reproducibility problems that can be caused by missing data, the danger of forgetting the missing values issues and the importance of providing sound implementations.

We welcome both academic and industrial practitioners/researchers. In particular, since missing data is a critical issue in many applications, we would like to federate industrial/applied know-how and various academic approaches.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:45 AM</td>
<td>Opening Session</td>
</tr>
<tr>
<td>02:00 AM</td>
<td>Poster session 1</td>
</tr>
<tr>
<td>04:30 AM</td>
<td>Invited Talk: Learning despite the unknown - missing data imputation in healthcare</td>
</tr>
<tr>
<td>05:10 AM</td>
<td>Invited Talk: Imputing Missing Data with the Gaussian Copula</td>
</tr>
<tr>
<td>05:50 AM</td>
<td>Discussion and Q&amp;A by Gael Varoquaux, Julie Josse and Pierre Alexandre Mattei</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Invited Talk: Efficient Missing-value Acquisition with Variational Autoencoders</td>
</tr>
</tbody>
</table>

Generated Tue Sep 29, 2020
07:10 AM Invited Talk: What Interpretable Machine Learning Can Tell Us About Missing Values

Caruana

07:50 AM Discussion and Q&A by Gael Varoquaux and Jes Frellsen

08:30 AM Poster session 2

09:10 AM Invited Talk: Graphical Models based Solutions for Missing Data Problems. Mohan

09:50 AM Invited Talk: Sequentially additive nonignorable missing data modelling using auxiliary marginal information Sadinle

10:30 AM Discussion and Q&A by Ilya Shpitser - Identifiability of the full law in graphical missing data models

N/A Informal gathering with drinks to celebrate

Abstracts (8):

Abstract 2: Poster session 1 in Learning with Missing Values, 02:00 AM

*Please do not share or post Zoom links*

**A Random Matrix Analysis of Learning with Ï¿-Dropout**

*Mohamed El Amine Seddik, Romain Couillet, Mohamed Tamaazousti*

[[Paper]](https://openreview.net/forum?id=upKKnBFv28Y)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-upKKnBFv28Y.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/8730845669?pwd=VG9iYmVqZEVzVzZwN0Z2YHRvkWtXUT09)

**Visna---Visualising Multivariate Missing Values**

*Antony Unwin, Alexander Pilhoefer*

[[Paper]](https://openreview.net/forum?id=czpT0YW6-C)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-czpT0YW6-C.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/86329756247?pwd=SkYyVERS5M0xqiQ1FhMOJueHJRUUZmQT09)

**Multi-output prediction of global vegetation distribution with incomplete data**

*Rita Beigaite, Jesse Read, Indre Zliobaite*

[[Paper]](https://openreview.net/forum?id=6cppShZBHg)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-6cppShZBHg.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/89492168092?pwd=VIN6NkVzVHJNY2VzNEi5azZn

**Path Imputation Strategies for Signature Models**

*Michael Moor, Max Horn, Christian Bock, Karsten Borgwardt, Bastian Riek*

[[Paper]](https://openreview.net/forum?id=P0DL7M6T57o)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-P0DL7M6T57o.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/81561993102?pwd=UGtjdkEvREwwUXBJMGHmHN0

**Lung Segmentation from Chest X-rays using Variational Data Imputation**

*Raghavendra Selvan, Erik Dam, Nicki Skatte Betlefens, Sofus Rischel, Kaining Sheng, Mads Nielsen, Akshay Pai*

[[Paper]](https://openreview.net/forum?id=dizQM28tq2W)
[[Join Zoom]](https://us02web.zoom.us/j/85694734356?pwd=U3ZmT0QyRXRCWddoRG5IT

**Clustering Data with nonignorable Missingness using Semi-Parametric Mixture Models**

*Marie Du Roy de Chaumaray, Matthieu Marbac*

[[Paper]](https://openreview.net/forum?id=aGzQ2qwxTrN)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-aGzQ2qwxTrN.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/85960176392?pwd=QVl2ZTJyVnJhcXMUTvRYjh

**Estimating conditional density of missing values using deep Gaussian mixture model**

*Marcin PrzewiÄ™Å‘likowski, Marek ÅŚmieja, Å‘ukasz Struski*

[[Paper]](https://openreview.net/forum?id=VR6mXmaHacL)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-VR6mXmaHacL.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/88659912949?pwd=VG5yUT12YQ4WG1RURJTNK

**Missing the Point: Non-Convergence in Iterative Imputation Algorithms**

*Hanne I. Oberman, Stef van Buuren, Gerko Vink*

[[Paper]](https://openreview.net/forum?id=fHSVg6mVqpw)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-fHSVg6mVqpw.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/89225286945?pwd=dHpMdVhxV0Q3ZGZxajVsR0R0

**The Dynamic Latent Block Model for Sparse and Evolving Count Matrices**

*Giulia Marchello, Marco Corneli, Charles Bouveyron*

[[Paper]](https://openreview.net/forum?id=P-pDhjAKOu)
[[Poster]](https://artemiss-workshop.github.io/files/posters/Artimiss-poster-P-pDhjAKOu.pdf)
[[Join Zoom]](https://us02web.zoom.us/j/81561993102?pwd=UGtjdkEvREwwUXBJMGHmHN0

**Informal gathering with drinks to celebrate**
Zoom[[https://us02web.zoom.us/j/82500475133?pwd=REImdnJKWHNkZl0d1BZ3f8pWFNnZzo9]

**Predicting Feature Imputability in the Absence of Ground Truth**

*Niamh McCombe, Xuemei Ding, Girijesh Prasad, David P Finn, Stephen Todd, Paula L McClean, Kongfatt Wong-Lin*

**How to miss data? Reinforcement learning for environments with high observation cost**

*Mehmet Koseoglu, Ayca Ozcelikkale*

[[Poster][https://artemiss-workshop.github.io/files/posters/Artimiss-poster-_wVkmK7wBhW.pdf]

**Missing rating imputation based on product reviews via deep latent variable models**

*Dingge Liang, Marco Cornel, Pierre Latouche, Charles Bouveyron*

**How to deal with missing data in supervised deep learning?**

*Niels Bruun Ipsen, Pierre-Alexandre Mattei, Jes Frellsen*


**The impact of incomplete data on quantile regression for longitudinal data**

*Anneleen Verhasselt, Alvaro JosÃ© FiÃ§rez, Ingrid Van Keilegom, Geert Molenberghs*

**Working with Deep Generative Models and Tabular Data Imputation**

*Ramiro Camino, Christian Hammerschmidt, Radu State*


**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*Jaspar Pahl, Ines Rieger, Dominik Seuss*

**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*[Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau]*

**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*

[[Poster][https://artemiss-workshop.github.io/files/posters/Artimiss-poster-R4w3PTkCD4.pdf]

**How to miss data? Reinforcement learning for environments with high observation cost**

*Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau*

**How to deal with missing data in supervised deep learning?**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

[[Paper][https://openreview.net/forum?id=jEN6sD09HWe]

**Predicting Feature Imputability in the Absence of Ground Truth**

*Niamh McCombe, Xuemei Ding, Girijesh Prasad, David P Finn, Stephen Todd, Paula L McClean, Kongfatt Wong-Lin*

**How to miss data? Reinforcement learning for environments with high observation cost**

*Mehmet Koseoglu, Ayca Ozcelikkale*


**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*

[[Paper][https://openreview.net/forum?id=HUw-EqC-WMj]

**The impact of incomplete data on quantile regression for longitudinal data**

*Anneleen Verhasselt, Alvaro JosÃ© FiÃ§rez, Ingrid Van Keilegom, Geert Molenberghs*

**Working with Deep Generative Models and Tabular Data Imputation**

*Ramiro Camino, Christian Hammerschmidt, Radu State*


**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*Jaspar Pahl, Ines Rieger, Dominik Seuss*

**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*[Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau]*

**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*


**How to miss data? Reinforcement learning for environments with high observation cost**

*Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau*

**How to deal with missing data in supervised deep learning?**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

[[Paper][https://openreview.net/forum?id=jEXxzPUMYVZ]

**Predicting Feature Imputability in the Absence of Ground Truth**

*Niamh McCombe, Xuemei Ding, Girijesh Prasad, David P Finn, Stephen Todd, Paula L McClean, Kongfatt Wong-Lin*

**How to miss data? Reinforcement learning for environments with high observation cost**

*Mehmet Koseoglu, Ayca Ozcelikkale*


**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*

[[Paper][https://openreview.net/forum?id=jEXxzPUMYVZ]

**The impact of incomplete data on quantile regression for longitudinal data**

*Anneleen Verhasselt, Alvaro JosÃ© FiÃ§rez, Ingrid Van Keilegom, Geert Molenberghs*

**Working with Deep Generative Models and Tabular Data Imputation**

*Ramiro Camino, Christian Hammerschmidt, Radu State*


**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*Jaspar Pahl, Ines Rieger, Dominik Seuss*

**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*[Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau]*

**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*

[[Paper][https://openreview.net/forum?id=HUw-EqC-WMj]

**The impact of incomplete data on quantile regression for longitudinal data**

*Anneleen Verhasselt, Alvaro JosÃ© FiÃ§rez, Ingrid Van Keilegom, Geert Molenberghs*

**Working with Deep Generative Models and Tabular Data Imputation**

*Ramiro Camino, Christian Hammerschmidt, Radu State*


**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*Jaspar Pahl, Ines Rieger, Dominik Seuss*

**Multi-label Learning with Missing Values using Combined Facial Action Unit Datasets**

*[Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau]*

**Inferring Causal Dependencies between Chaotic Dynamical Systems from Sporadic Time Series**

*Chao Ma, Sebastian Tschatschek, Richard E. Turner, JosÃ© Miguel HernÃ¡ndez-Lobato, Cheng Zhang*

**VAEM: a Deep Generative Model for Heterogeneous Mixed Type Data**

*Mehmet Koseoglu, Ayca Ozcelikkale*
The algorithm models mixed data as a Gaussian copula. This model can fit arbitrary marginals for continuous variables and can handle ordinal variables with many levels, including Boolean variables as a special case.

We develop an efficient approximate EM algorithm to estimate copula parameters from incomplete mixed data. The resulting model reveals the statistical associations among variables. Experimental results on several synthetic and real datasets show the superiority of the proposed algorithm to state-of-the-art imputation algorithms for mixed data.

Abstract 6: Invited Talk: Efficient Missing-value Acquisition with Variational Autoencoders in Learning with Missing Values, Hernandez-Lobato 06:30 AM

Abstract: In many real-world problems we have to make predictions from feature vectors with missing values. However, we may also be able to observe some of the missing values in the feature vector at a cost. Given the currently observed values, how can we decide which missing values to observe next so that prediction accuracy increases as fast as possible as a function of the observation cost? This problem appears in many different application areas, including medical diagnosis, surveys, recommender systems, insurance, etc. In this talk, I will describe how to solve the problem using an information theoretic approach and novel variational autoencoder models that can effectively deal with missing data.

Abstract 7: Invited Talk: What Interpretable Machine Learning Can Tell Us About Missing Values in Learning with Missing Values, Caruana 07:10 AM

Missing values are everywhere, and lâ€™ve been dealing with them one way or another for many years. Recently lâ€™ve been doing research in interpretable machine learning. To my surprise, interpretable machine learning has completely changed how I work with missing values. Interpretable learning provides new methods for detecting, understanding, and modeling missing values. In the presentation lâ€™m show a few surprises where interpretability makes it clear the impact missing values have been having on our machine learning models all along, but which are only visible now thanks to interpretable methods.

Abstract 9: Poster session 2 in Learning with Missing Values, 08:30 AM

*Please do not share or post Zoom links*

**Optimal recovery of missing values for non-negative matrix factorization: A probabilistic error bound**

*Rebecca Chen, Lav R. Varshney*

**Handling Missing Data in Decision Trees: A Probabilistic Approach**

*Eric Landgrebe, yuxuan zhao, Madeleine Udell*

**Imputation of Missing Behavioral Measures in Connectome-based Predictive Modelling**

*Qinghao Liang, Dustin Scheinost*

**Does imputation matter? Benchmark for real-life classification problems.**

*Katarzyna WoÅ›nica, Przemyslaw Biecek*

**VAEs in the Presence of Missing Data**

*Mark Collier, Alfredo Nazabal, Chris Williams*

**Variance estimation after Kernel Ridge Regression Imputation**

*Hengfang Wang, Jae Kwang Kim*

**Online Mixed Missing Value Imputation Using Gaussian Copula**

*Eric Landgrebe, yuxuan zhao, Madeleine Udell*

**Imputation of Missing Behavioral Measures in Connectome-based Predictive Modelling**

*Qinghao Liang, Dustin Scheinost*
**Processing of incomplete images by (graph) convolutional neural networks**

*Tomasz Danel, Marek Âłâmieja, Łukasz Struski, Przemysław Spurek, Łukasz Maziarka*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Information Theoretic Approaches for Testing Missingness in Predictive Models**

*Shreyas A Bhave, Rajesh Ranganath, Adler Perotte*

**Conditioning on "and nothing else": Simple Models of Missing Data between Naive Bayes and Logistic Regression**

*David Poole, Ali Mohammad Mehr, Wan Shing Martin Wang*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Processing of incomplete images by (graph) convolutional neural networks**

*Tomasz Danel, Marek Âłâmieja, Łukasz Struski, Przemysław Spurek, Łukasz Maziarka*

We study a class of missingness mechanisms, referred to as sequentially additive nonignorable, for modelling multivariate data with item nonresponse. These mechanisms explicitly allow the probability of nonresponse for each variable to depend on the value of that variable, as well as representing nonignorable missingness mechanisms. These missing data models are identified by making use of auxiliary information or by local distributions, such as marginal probabilities for multivariate categorical variables or moments for numeric variables. We prove identification results and illustrate the use of these mechanisms in an application.

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

We study a class of missingness mechanisms, referred to as sequentially additive nonignorable, for modelling multivariate data with item nonresponse. These mechanisms explicitly allow the probability of nonresponse for each variable to depend on the value of that variable, as well as representing nonignorable missingness mechanisms. These missing data models are identified by making use of auxiliary information or by local distributions, such as marginal probabilities for multivariate categorical variables or moments for numeric variables. We prove identification results and illustrate the use of these mechanisms in an application.

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

We study a class of missingness mechanisms, referred to as sequentially additive nonignorable, for modelling multivariate data with item nonresponse. These mechanisms explicitly allow the probability of nonresponse for each variable to depend on the value of that variable, as well as representing nonignorable missingness mechanisms. These missing data models are identified by making use of auxiliary information or by local distributions, such as marginal probabilities for multivariate categorical variables or moments for numeric variables. We prove identification results and illustrate the use of these mechanisms in an application.

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

We study a class of missingness mechanisms, referred to as sequentially additive nonignorable, for modelling multivariate data with item nonresponse. These mechanisms explicitly allow the probability of nonresponse for each variable to depend on the value of that variable, as well as representing nonignorable missingness mechanisms. These missing data models are identified by making use of auxiliary information or by local distributions, such as marginal probabilities for multivariate categorical variables or moments for numeric variables. We prove identification results and illustrate the use of these mechanisms in an application.

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

**Multi-Time Attention Networks for Irregularly Sampled Time Series**

*Satya Narayan Shukla, Benjamin Marlin*

We study a class of missingness mechanisms, referred to as sequentially additive nonignorable, for modelling multivariate data with item nonresponse. These mechanisms explicitly allow the probability of nonresponse for each variable to depend on the value of that variable, as well as representing nonignorable missingness mechanisms. These missing data models are identified by making use of auxiliary information or by local distributions, such as marginal probabilities for multivariate categorical variables or moments for numeric variables. We prove identification results and illustrate the use of these mechanisms in an application.
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:30 AM</td>
<td>Live Q&amp;A Invited Speaker Panel</td>
<td></td>
</tr>
<tr>
<td>08:30 AM</td>
<td>AI in Diagnosis &amp; Therapy II</td>
<td></td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Epidemiology &amp; Policy I</td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Medical Data, Synthesis, &amp; Privacy II</td>
<td></td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Health Systems &amp; Delivery II</td>
<td></td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Epidemiology &amp; Policy II</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>Monitoring Mental Health:</td>
<td>Kumar</td>
</tr>
<tr>
<td></td>
<td>Identifying Depressive and Suicidal Sentiments on Online Forums using Deep Learning</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>Robert Morris -- Invited Talk</td>
<td>Morris</td>
</tr>
<tr>
<td>N/A</td>
<td>Enabling autonomous clinical decision support systems in space through AI-enhanced wearables</td>
<td>Antoniadou</td>
</tr>
<tr>
<td>N/A</td>
<td>Trishan Panch &amp; Mohammad Jouni -- Invited Talk</td>
<td>Panch</td>
</tr>
<tr>
<td>N/A</td>
<td>Noa Dagan -- Invited Talk</td>
<td>Dagan</td>
</tr>
<tr>
<td>N/A</td>
<td>Rumi Chunara -- Invited Talk</td>
<td>Chunara</td>
</tr>
<tr>
<td>N/A</td>
<td>Barbara Engelhardt -- Invited Talk</td>
<td>Engelhardt</td>
</tr>
<tr>
<td>N/A</td>
<td>Who Should We Test for COVID-19? A Triage Model Built from National Symptom Surveys</td>
<td>Shoer</td>
</tr>
<tr>
<td>N/A</td>
<td>MPC-guided Imitation Learning of Neural Network Policies for the Artificial Pancreas</td>
<td>Chen</td>
</tr>
<tr>
<td>N/A</td>
<td>Predicting Length of Stay in the Intensive Care Unit with Temporal Pointwise Convolutional Networks</td>
<td>Rocheteau</td>
</tr>
<tr>
<td>N/A</td>
<td>Deep Claim: Payer Response Prediction from Claims Data with Deep Learning</td>
<td>Kim</td>
</tr>
</tbody>
</table>

Abstracts (8):

**Abstract 1:** Medical Data, Synthesis, & Privacy I in Healthcare Systems, Population Health, and the Role of Health-tech, 03:30 AM

Moderator: Creighton Heaukulani

Zoom join link:
https://us04web.zoom.us/j/79416976497?pwd=djZNREZxUm9QN2licVVJTVYtMHpSUT09

Synthetic Tabular Data Generation with Oblivious Variational Autoencoders : Alleviating the Paucity of Personal Tabular Data for Open Research

Synthesis of Time Series Physiological Data from Wearables using Deep Networks

Enabling autonomous clinical decision support systems in space through AI-enhanced wearables

**Abstract 2:** Health Systems & Delivery I in Healthcare Systems, Population Health, and the Role of Health-tech, 05:00 AM

Moderator: Creighton Heaukulani

Zoom join link:
https://us04web.zoom.us/j/7738399130?pwd=cXYyeE9JRXZMTy9acGVZaHVnNU5NVz09

Interpretable Outcome Prediction with Sparse Bayesian Neural Networks in Intensive Care

Predicting Length of Stay in the Intensive Care Unit with Temporal Pointwise Convolutional Networks

Accurate prediction of population health costs

Rapid deployment of a nationwide, one-minute, online symptoms survey during the outbreak and spread of COVID-19 - framework and applications

**Abstract 3:** AI in Diagnosis & Therapy I in Healthcare Systems, Population Health, and the Role of Health-tech, 05:00 AM

Moderator: Konstantina Palla

Zoom join link:
https://us04web.zoom.us/j/74464246195?pwd=dWo3dzArYk9LcjdzZi9VbGJ4MkFDdz09

Deep Semi-Supervised Embedded Clustering (DSEC) for Stratification of Heart Failure Patients

Prediction of the onset of cardiovascular diseases from electronic health records using multi-task gated recurrent units

Enhancing diagnosis of tuberculosis in children with novel Mycobacterium tuberculosis antigens

**Abstract 5:** AI in Diagnosis & Therapy II in Healthcare Systems, Population Health, and the Role of Health-tech, 08:30 AM

Moderator: Konstantina Palla

Zoom join link:
https://us04web.zoom.us/j/76986651479?pwd=enBTaUxyR3BIV0wzcndER1h5QWJDd0

MPC-guided Imitation Learning of Neural Network Policies for the Artificial Pancreas
Monitoring Mental Health: Identifying Depressive and Suicidal Sentiments on Online Forums using Deep Learning

Using Capsule Neural Network to predict Tuberculosis in lens-free microscopic images

Identifying patterns in cystic fibrosis physiotherapy using unsupervised clustering

Abstract 6: Epidemiology & Policy I in Healthcare Systems, Population Health, and the Role of Health-tech, 08:30 AM
Moderator: Niranjani Prasad
Zoom join link: https://us04web.zoom.us/j/78050144268?pwd=QnZhQ080cklHOWjVUVNuVG1DQ1VwZz09

An Unsupervised Machine Learning Approach to Assess the ZIP Code Level Impact of COVID-19 in NYC

In the Danger Zone: U-Net Driven Quantile Regression can Predict High-risk SARS-CoV-2 Regions via Pollutant Particulate Matter and Satellite Imagery

Who Should We Test for COVID-19? A Triage Model Built from National Symptom Surveys

Abstract 7: Medical Data, Synthesis, & Privacy II in Healthcare Systems, Population Health, and the Role of Health-tech, 10:00 AM
Moderator: Niranjani Prasad
Zoom join link: https://us04web.zoom.us/j/74714284876?pwd=ZnNiUFVkJ2RFcmdWdYeWZ2OU9CQT09

Patient-Reported Outcomes: A Privacy-Centric and Federated Approach to Machine Learning

Benchmarking Differentially Private Residual Networks for Medical Imagery

Towards User Friendly Medication Mapping Using Entity-Boosted Two-Tower Neural Network

Abstract 8: Health Systems & Delivery II in Healthcare Systems, Population Health, and the Role of Health-tech, 11:30 AM
Moderator: Katherine Heller
Zoom join link: https://us04web.zoom.us/j/79454426445?pwd=b2F2NJicnBGTVU6K1sd2VQeGZmQT09

Deep Claim: Payer Response Prediction from Claims Data with Deep Learning

AI-based Monitoring and Response System for Hospital Preparedness towards COVID-19 in Southeast Asia

A Model is Not Enough: A case of AI-Enabled Healthcare Delivery


Abstract 9: Epidemiology & Policy II in Healthcare Systems, Population Health, and the Role of Health-tech, 01:00 PM
Moderator: Katherine Heller
Zoom join link: https://us04web.zoom.us/j/72221040261?pwd=OUNuanJyTXZWRGQ4Wi8yc0w0bDErQT09

Forecasting Influenza Prevalence with Deep Transformer Models

Interpretable Covid-19 Forecasting

An Epidemiological Modelling Approach for Covid19 via Data Assimilation

Workshop on AI for Autonomous Driving (AIAD)

Wei-Lun (Harry) Chao, Rowan McAllister, Adrien Gaidon, Li Erran Li, Sven Kreiss

Fri Jul 17, 05:00 AM

Self-driving cars and advanced safety features present one of today’s greatest challenges and opportunities for Artificial Intelligence (AI). Despite billions of dollars of investments and encouraging progress under certain operational constraints, there are no driverless cars on public roads today without human safety drivers. Autonomous Driving research spans a wide spectrum, from modular architectures -- composed of hardcoded or independently learned sub-systems -- to end-to-end deep networks with a single model from sensors to controls. In any system, Machine Learning is a key component. However, there are formidable learning challenges due to safety constraints, the need for large-scale manual labeling, and the complex high dimensional structure of driving data, whether inputs (from cameras, HD maps, inertial measurement units, wheel encoders, LiDAR, radar, etc.) or predictions (e.g., world state representations, behavior models, trajectory forecasts, plans, controls). The goal of this workshop is to explore the frontier of learning approaches for safe, robust, and efficient Autonomous Driving (AD) at scale. The workshop will span both theoretical frameworks and practical issues especially in the area of deep learning.

Website: https://sites.google.com/view/aiad2020

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:00 AM</td>
<td>Open Remark 1</td>
</tr>
<tr>
<td></td>
<td>Chao, McAllister, Li, Kreiss, Gaidon</td>
</tr>
<tr>
<td>05:10 AM</td>
<td>Invited Talk: Deep Direct Visual SLAM (Daniel Cremers)</td>
</tr>
<tr>
<td></td>
<td>Cremers</td>
</tr>
<tr>
<td>05:40 AM</td>
<td>Q&amp;A: Daniel Cremers</td>
</tr>
<tr>
<td></td>
<td>Cremers</td>
</tr>
<tr>
<td>05:50 AM</td>
<td>Invited Talk: Raster-based Motion Prediction for Safe Self-Driving (Nemanja Djuric)</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>06:20 AM</td>
<td>Q&amp;A: Nemanja Djuric</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Coffee break</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Q&amp;A: Ingmar Posner</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>Invited Talk: Motion Prediction for Vulnerable Road Users (Dariu Gavrila)</td>
</tr>
<tr>
<td>07:50 AM</td>
<td>Q&amp;A: Dariu Gavrila</td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Panel Discussion 1</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Paper presentation opening</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Deep Representation Learning and Clustering of Traffic Scenarios</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Towards Map-Based Validation of Semantic Segmentation Masks</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Multi-agent Graph Reinforcement Learning for Connected Automated Driving</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Autonomous Driving with Reinforcement Learning and Rule-based Policies</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Can Autonomous Vehicles Identify, Recover From, and Adapt to Distribution Shifts?</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: INSTA-YOLO: Real-Time Instance Segmentation based on YOLO</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Imitation Learning Approach for AI Driving Olympics Trained on Real-world and Simulation Data Simultaneously</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper Q&amp;A session 1</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Trajectograms: Which Semi-Supervised Trajectory Prediction Model to Use?</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Interpretable End-to-end Autonomous Driving with Reinforcement Learning</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Learning Invariant Representations for Reinforcement Learning without Reconstruction</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Depth Meets CNN: A Fusion Based Approach for Semantic Road Segmentation</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Paper spotlight: Learning Multiplicative Interactions with Bayesian Neural Networks for Visual-Inertial Odometry</td>
</tr>
<tr>
<td>10:25 AM</td>
<td>Open Remark 2</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Invited Talk: Neural Motion Planning for Self-Driving (Raquel Urtasun)</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Q&amp;A: Raquel Urtasun</td>
</tr>
<tr>
<td>11:10 AM</td>
<td>Invited Talk: Autonomous Driving: The Way Forward (Vladlen Koltun)</td>
</tr>
<tr>
<td>11:40 AM</td>
<td>Coffee break</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Invited Talk: What we learned from Argoverse Competitions (James Hays)</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Q&amp;A: James Hays</td>
</tr>
</tbody>
</table>
Invited Talk: Feedback in Imitation Learning: Confusion on Causality and Covariate Shift (Arun Venkatraman & Sanjiban Choudhury)

Q&A: Arun Venkatraman & Sanjiban Choudhury

Invited Talk: INTERPRET: INTERACTION-dataset-based PREdicTion Challenge (Wei Zhan)

Q&A: Wei Zhan

Panel Discussion 2

Closing remark (best paper award: sponsored by NVIDIA)

Paper Q&A session 2

Abstract: The reconstruction of our 3D world from moving cameras is among the central challenges in computer vision. I will present recent developments in camera-based reconstruction of the world. In particular, I will discuss direct methods for visual SLAM (simultaneous localization and mapping). These recover camera motion and 3D structure directly from brightness consistency thereby providing better performance in terms of precision and robustness compared to classical keypoint-based techniques. Moreover, I will demonstrate how we can leverage the predictive power of deep networks in order to significantly boost the performance of direct SLAM methods. The resulting methods allow us to track a single camera with a precision that is on par with state-of-the-art approaches. We show that they are capable of operating in a wide range of environments including the 'Best Paper of the Year 2003' (Int. Pattern Recognition Society), the 'Olympus Award 2004' (German Soc. for Pattern Recognition) and the '2005 UCLA Chancellor's Award for Postdoctoral Research'. For pioneering research he received five grants from the European Research Council, including a Starting Grant, a Consolidator Grant and an Advanced Grant. In 2018 he organized the largest ever European Conference on Computer Vision in Munich. He is member of the Bavarian Academy of Sciences and Humanities. In December 2010 he was listed among "Germany's top 40 researchers below 40" (Capital). On March 1st 2016, Prof. Cremers received the Gottfried Wilhelm Leibniz Award, the biggest award in German academia. He is co-founder of several companies, most recently the high-tech startup Artsense.

Abstract 3: Q&A: Daniel Cremers in Workshop on AI for Autonomous Driving (AIAD), Cremers 05:40 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYvU1RDbzB1dz09

Abstract 4: Invited Talk: Raster-based Motion Prediction for Safe Self-Driving (Nemanja Djuric) in Workshop on AI for Autonomous Driving (AIAD), Djuric 05:50 AM

Video: https://slideslive.com/38930749/rasterbased-motion-prediction-for-safe-selfdriving

Abstract: Motion prediction is a critical component of self-driving technology, tasked with inferring future behavior of traffic actors as well as modeling behavior uncertainty. In the talk, we focus on this important problem, and discuss raster-based methods that have shown state-of-the-art performance. These approaches take top-down images of the surrounding area as their input, providing near-complete contextual information necessary to accurately predict traffic motion. We present a number of recently proposed models, and show how to develop methods that obey map and other physical constraints of the environment.

Bio: Nemanja Djuric is a Staff Engineer and Tech Lead Manager at Uber ATG, for the past 5 years working on motion prediction, object detection, and other technologies supporting self-driving vehicles. Prior to ATG he worked as a research scientist at Yahoo Labs, which he joined after obtaining his PhD at Temple University.

Abstract 5: Q&A: Nemanja Djuric in Workshop on AI for Autonomous Driving (AIAD), Djuric 06:20 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYvU1RDbzB1dz09


Video: https://youtu.be/VhCNOuNmgpA

Abstract: In providing long-range information and significant robustness to environmental conditions radar complements perfectly some of the more commonly used sensing modalities in autonomous driving. However, radar data is also notoriously difficult to work with. Significant, context-dependent sensing artefacts and noise characteristics make interpretation and use of this data a real challenge. In this talk I will describe some of the work done in the Applied AI Lab at Oxford in leveraging learning to enable radar-based perception and navigation.
particular, I will talk about how we use system-level self-supervision - the use of adjacent sensing or subsystems to derive a learning signal during training - in order to make radar data palatable during deployment. I will introduce work that explicitly accounts for the particular noise-characteristics of a radar in order to map from raw radar scans to occupancy grids; I will describe an approach to interpretable ego-motion estimation learning an inherent distraction suppression; and I will give an overview of how we can construct a fully fledged radar-based navigation system.

Bio: Ingmar leads the Applied Artificial Intelligence Lab at Oxford University and is a founding director of the Oxford Robotics Institute. His goal is to enable robots to robustly and effectively operate in complex, real-world environments. His research is guided by a vision to create machines which constantly improve through experience. In doing so Ingmar's work explores a number of intellectual challenges at the heart of robot learning, such as unsupervised scene interpretation, action inference and machine introspection. All the while Ingmarâ€™s research remains grounded in real-world robotics applications such as manipulation, autonomous driving, logistics and space exploration. In 2014 Ingmar co-founded Oxbotica, a multi-award winning provider of mobile autonomy software solutions.

Abstract 8: Q&A: Ingmar Posner in Workshop on AI for Autonomous Driving (AIAD), Posner 07:10 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGHQRXppZ1pKUMxURHYyUl1RDdzB2UT09

Abstract 9: Invited Talk: Motion Prediction for Vulnerable Road Users (Dariu Gavrila) in Workshop on AI for Autonomous Driving (AIAD), Gavrila 07:20 AM

Video: https://youtu.be/qWKfhrDItU

Abstract: Sensors are meanwhile very good at measuring 3D in the context of environment perception for self-driving vehicles. Scene labeling and object detection have also made big strides, mainly due to advances in deep learning. Time has now come to focus on the next frontier: modeling and anticipating the motion of road users. The potential benefits are large, such as earlier and more effective system reactions in dangerous traffic situations. To reap these benefits, however, it is necessary to use sophisticated predictive motion models based on intent-relevant (context) cues. In this talk, I give an overview of predictive motion models and intent-relevant cues with respect to the vulnerable road users (i.e. pedestrians, cyclists). In particular, I discuss the pros and cons of having these models handcrafted by an expert compared to learning them from data. I present results from a recent case study on cyclist path prediction involving a Dynamic Bayesian Network and a Recurrent Neural Network.

Bio: Dariu M. Gavrila received the PhD degree in computer science from the University of Maryland at College Park, USA, in 1996. During 1997 - 2016, he was with Daimler R&D in Ulm, Germany, where he became a Distinguished Scientist. During 2003 - 2018, he was also professor at the University of Amsterdam, chairing the area of Intelligent Perception Systems (part time). Since 2016 he is head of the Intelligent Vehicles group at TU Delft, full time (www.intelligent-vehicles.org). Over the past 20 years, Prof. Gavrila has focused on visual systems for detecting humans and their activity, with application to intelligent vehicles, smart surveillance and social robotics. He led the multi-year pedestrian detection research effort at Daimler R&D, which was commercialized in the Mercedes-Benz S-, E-, and C-Class models (2013-2014). He now performs research on self-driving cars in complex urban environment and focusses on the anticipation of pedestrian and cyclist behavior. Prof. Gavrila is frequently cited in the scientific literature (Google Scholar: 13.000+ times) and he received the I/O 2007 Award from the Netherlands Organisation for Scientific Research (NWO) and the IEEE Intelligent Transportation Systems Application Award 2014 (as part of a Daimler team). He served as Area and Program Co-Chair at several conferences (IV, ICCV, ECCV, AVSS). His research was covered in various print and broadcast media, such as Wired Magazine, Der Spiegel, BBC Radio, 3Sat Nano and NOS.

Abstract 10: Q&A: Dariu Gavrila in Workshop on AI for Autonomous Driving (AIAD), Gavrila 07:50 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGHQRXppZ1pKUMxURHYyUl1RDdzB2UT09

Abstract 11: Panel Discussion 1 in Workshop on AI for Autonomous Driving (AIAD), Cremers, Djunic, Posner, Gavrila 08:00 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGHQRXppZ1pKUMxURHYyUl1RDdzB2UT09

Abstract 12: Paper presentation opening in Workshop on AI for Autonomous Driving (AIAD), McAllister, Li, Gaidon, Kreiss, Chao 08:30 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGHQRXppZ1pKUMxURHYyUl1RDdzB2UT09

Abstract 13: Paper spotlight: Deep Representation Learning and Clustering of Traffic Scenarios in Workshop on AI for Autonomous Driving (AIAD), Harmening, GA, innemann, BiloA 08:35 AM

Video: https://slideslive.com/38931754

Individual Zoom meeting: https://us02web.zoom.us/j/82456487107?pwd=T2p1ZVBVRDhYyXSN2FzdFJRzVnQ0Z0

Abstract 14: Paper spotlight: Towards Map-Based Validation of Semantic Segmentation Masks in Workshop on AI for Autonomous Driving (AIAD), von Rueden 08:35 AM

Video: https://slideslive.com/38931748

Individual Zoom meeting: https://us02web.zoom.us/j/89373300557?pwd=REZRMXZrXJJKQ0NTOU91TzFzVTNvZ0

Abstract 15: Paper spotlight: Multi-agent Graph Reinforcement Learning for Connected Automated Driving in Workshop on AI for Autonomous Driving (AIAD), SHI 08:35 AM

Video: https://slideslive.com/38931756/multiagent-graph-reinforcement-learning-for-connected-automated-driving

Individual Zoom meeting: https://us02web.zoom.us/j/83693184668?pwd=SEZWTGqIT11oaFZYORzVQXUVjNG12

Abstract 16: Paper spotlight: Probabilistic Object Detection: Strengths, Weaknesses, Opportunities in Workshop on AI for Autonomous Driving (AIAD), Bhatt 08:35 AM

Video: https://slideslive.com/38931756/
Abstract 17: Paper spotlight: Autonomous Driving with Reinforcement Learning and Rule-based Policies in Workshop on AI for Autonomous Driving (AIAD), Likmeta 08:35 AM

Video: https://slideslive.com/38931747
Individual Zoom meeting: https://us02web.zoom.us/j/85016432368?pwd=aU9qVFJqVmZ0WW5LWmNjSxJllzjHQT09

Abstract 18: Paper spotlight: Can Autonomous Vehicles Identify, Recover From, and Adapt to Distribution Shifts? in Workshop on AI for Autonomous Driving (AIAD), Filos, Tigas 08:35 AM

Video: https://slideslive.com/38931743
Individual Zoom meeting: https://us02web.zoom.us/j/82135269788?pwd=RzNvRkIJSU5h5zZKZnNUNUFzMXJGQT09

Abstract 19: Paper spotlight: INSTA-YOLO: Real-Time Instance Segmentation based on YOLO in Workshop on AI for Autonomous Driving (AIAD), Mohamed Abd El Rahman 08:35 AM

Video: https://slideslive.com/38931744
Individual Zoom meeting: https://us02web.zoom.us/j/84540391510?pwd=SudCeVNpcE5iMTFCN0NmMGtUS3JGQT09

Abstract 20: Paper spotlight: Imitation Learning Approach for AI Driving Olympics Trained on Real-world and Simulation Data Simultaneously in Workshop on AI for Autonomous Driving (AIAD), Sazanovich 08:35 AM

Video: https://slideslive.com/38931755
Individual Zoom meeting: https://us02web.zoom.us/j/89707469981?pwd=R3pmcGJYajJVZ1VRbURsOFOYdYyCQ

Abstract 21: Paper Q&A session 1 in Workshop on AI for Autonomous Driving (AIAD), 08:35 AM

Video: https://slideslive.com/38931752
Individual Zoom meeting: https://us02web.zoom.us/j/89198067302?pwd=WG14dGMrRUNUVRWTUNXMEZVoFI

Abstract 22: Paper spotlight: SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving in Workshop on AI for Autonomous Driving (AIAD), AKSOY 08:35 AM

Video: https://slideslive.com/38931751
Individual Zoom meeting: https://us02web.zoom.us/j/82135269788?pwd=by9QWmYxUzFaQXArRExQdkt1azZKQT09

Abstract 23: Paper spotlight: Trajectograms: Which Semi-Supervised Trajectory Prediction Model to Use? in Workshop on AI for Autonomous Driving (AIAD), Lamm, Drori, Jaiprakash 08:35 AM

Video: https://slideslive.com/38931746
Individual Zoom meeting: https://us02web.zoom.us/j/89198067302?pwd=WG14dGMrRUNUVRWTUNXMEZVoFI

Abstract 24: Paper spotlight: Interpretable End-to-end Autonomous Driving with Reinforcement Learning in Workshop on AI for Autonomous Driving (AIAD), Chen 08:35 AM

Video: https://slideslive.com/38931746
Individual Zoom meeting: https://us02web.zoom.us/j/89198067302?pwd=WG14dGMrRUNUVRWTUNXMEZVoFI
Abstract 25: Paper spotlight: Learning Invariant Representations for Reinforcement Learning without Reconstruction in Workshop on AI for Autonomous Driving (AIAD), Zhang 08:35 AM

Video: https://slideslive.com/38931749

Abstract 26: Paper spotlight: Depth Meets CNN: A Fusion Based Approach for Semantic Road Segmentation in Workshop on AI for Autonomous Driving (AIAD), Atrishi, Singh, Gupta, Marwaha 08:35 AM

Video: https://slideslive.com/38931753

Abstract 27: Paper spotlight: Learning Multiplicative Interactions with Bayesian Neural Networks for Visual-Inertial Odometry in Workshop on AI for Autonomous Driving (AIAD), Shinde, Lee 08:35 AM

Video: https://slideslive.com/38931750

Abstract 28: Open Remark 2 in Workshop on AI for Autonomous Driving (AIAD), Chao, Kreiss, McAllister, Li, Gaidon 10:25 AM

Zoom webinar: https://us02web.zoom.us/j/89706077961?pwd=cm0yY3UwZ3g4eXRKNDY0aDZyaXJXRG90dz09

Abstract 29: Invited Talk: Neural Motion Planning for Self-Driving (Raquel Urtasun) in Workshop on AI for Autonomous Driving (AIAD), Urtasun 10:30 AM

Video: https://slideslive.com/38930753/neural-motion-planning-for-selfdriving

Bio: Raquel Urtasun is the Chief Scientist for Uber ATG and the Head of Uber ATG Toronto. She is also a Professor at the University of Toronto, a Canada Research Chair in Machine Learning and Computer Vision and a co-founder of the Vector Institute for AI. She received her Ph.D. from the Ecole Polytechnique Federal de Lausanne (EPFL) in 2006 and did her postdoc at MIT and UC Berkeley. She is a recipient of an NSERC EWR Steacie Award, an NVIDIA Pioneers of AI Award, a Ministry of Education and Innovation Early Researcher Award, three Google Faculty Research Awards, an Amazon Faculty Research Award, a Connaught New Researcher Award, a Fallona Family Research Award and two Best Paper Runner up Prize awarded CVPR in 2013 and 2017. She was also named Chatelaine 2018 Woman of the year, and 2018 Toronto’s top influencers by Adweek magazine.

Abstract 30: Q&A: Raquel Urtasun in Workshop on AI for Autonomous Driving (AIAD), Urtasun 11:00 AM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXxz1pKUmUxURHYvU1RDbzBWUT09

Abstract 31: Invited Talk: Autonomous Driving: The Way Forward (Vladlen Koltun) in Workshop on AI for Autonomous Driving (AIAD), Koltun 11:10 AM

The video link is here: https://youtu.be/XmtTqjmW3g
If you have questions to Vladlen Koltun, please contact him by http://vladlen.info/contact/

Bio: Vladlen Koltun is the Chief Scientist for Intelligent Systems at Intel. He directs the Intelligent Systems Lab, which conducts high-impact basic research in computer vision, machine learning, robotics, and related areas. He has mentored more than 50 PhD students, postdocs, research scientists, and PhD student interns, many of whom are now successful research leaders. Web: http://vladlen.info

Abstract 32: Invited Talk: What we learned from Argoverse Competitions (James Hays) in Workshop on AI for Autonomous Driving (AIAD), Hays 12:00 PM

Video: https://slideslive.com/38930752/what-we-learned-from-argoverse-competitions

Abstract: This talk will have two parts. First, I’ll discuss what we’ve learned from Argoverse competitions in 2020 and 2019. We’ll analyze the strategies used by the top scoring teams in 3D tracking and Motion forecasting, and examine situations where there is still room for improvement.

In the second part, I’ll discuss the “inflation” of 2D instance segmentations into 3D cuboids suitable for training 3D object detectors. With the help of an HD map, 2D instance masks can be converted into surprisingly accurate 3D training data for LiDAR-based detectors. We show that we can mine 3D cuboids from unlabeled self-driving logs and train a 3D detector that outperforms a human-supervised baseline.

Bio: James Hays is an associate professor of computing at Georgia Institute of Technology since fall 2015. Since 2017, He also works with Argo AI to create self-driving cars. Previously, he was the Manning assistant professor of computer science at Brown University. He received his Ph.D. from Carnegie Mellon University and was a postdoc at Massachusetts Institute of Technology. His research interests span computer vision, robotics, and machine learning. His research often involves exploiting non-traditional data sources (e.g. internet imagery, crowdsourced annotations, thermal imagery, human sketches, autonomous vehicle sensor data) to explore new research problems (e.g. global geolocalization, sketch to real, hand-object contact prediction).

Abstract 33: Q&A: James Hays in Workshop on AI for Autonomous Driving (AIAD), Hays 12:30 PM

Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXxz1pKUmUxURHYvU1RDbzBWUT09

Abstract 34: Invited Talk: Feedback in Imitation Learning: Confusion on Causality and Covariate Shift (Arun Venkatraman & Sanjiban Choudhury) in Workshop on AI for Autonomous Driving (AIAD), Choudhury, Venkatraman 12:40 PM

Video: https://us02web.zoom.us/j/89706077961?pwd=cm0yY3UwZ3g4eXRKNDY0aDZyaXJXRG90dz09

Abstract: This talk will have two parts. First, I’ll discuss what we’ve learned from Argoverse competitions in 2020 and 2019. We’ll analyze the strategies used by the top scoring teams in 3D tracking and Motion forecasting, and examine situations where there is still room for improvement.

In the second part, I’ll discuss the “inflation” of 2D instance segmentations into 3D cuboids suitable for training 3D object detectors. With the help of an HD map, 2D instance masks can be converted into surprisingly accurate 3D training data for LiDAR-based detectors. We show that we can mine 3D cuboids from unlabeled self-driving logs and train a 3D detector that outperforms a human-supervised baseline.

Bio: James Hays is an associate professor of computing at Georgia Institute of Technology since fall 2015. Since 2017, He also works with Argo AI to create self-driving cars. Previously, he was the Manning assistant professor of computer science at Brown University. He received his Ph.D. from Carnegie Mellon University and was a postdoc at Massachusetts Institute of Technology. His research interests span computer vision, robotics, and machine learning. His research often involves exploiting non-traditional data sources (e.g. internet imagery, crowdsourced annotations, thermal imagery, human sketches, autonomous vehicle sensor data) to explore new research problems (e.g. global geolocalization, sketch to real, hand-object contact prediction).

Abstract 35: Q&A: Arun Venkatraman & Sanjiban Choudhury in Workshop on AI for Autonomous Driving (AIAD), 12:40 PM
and a 2013 Siebel's Scholar. He worked with Sidd Srinivasa. He is the recipient of best paper awards at Postdoctoral fellow at the University of Washington, CSE where he learn from prior experience to speed up online planning. He was a advised by Sebastian Scherer. His thesis showed how robots can full-scale helicopters, self-driving cars and mobile manipulators. He has a Much of his research has been deployed on real-world robotic systems - algorithms at the intersection of machine learning and motion planning. with the best to solve self-driving at scale. He focuses on theory and Sanjiban Choudhury is a research engineer at Aurora, where he works to develop the Aurora Driver.

Sanjiban Choudhury is a research engineer at Aurora, the company delivering self-driving technology safely, quickly, and broadly. Arun graduated with a BS with Honors from the California Institute of Technology and completed his PhD, Training Strategies for Time Series: Learning for Prediction, Filtering, and Reinforcement Learning, at the Robotics Institute at Carnegie Mellon University co-advised by Dr. Drew Bagnell and Dr. Martial Hebert. During his time at CMU and NREC, Arun worked on a variety of robotics applications and received a best paper award at Robotics Science and Systems 2015 for work on autonomy assisted teleoperation via a brain-computer interface. At Aurora, Arun leads the Motion Planning Machine Learning team, bringing together the best in machine learning with the best practices in robotics development to develop the Aurora Driver.

Bio: Wei Zhan is a Postdoctoral Scholar at UC Berkeley working with Professor Masayoshi Tomizuka. He received his Ph.D. from UC Berkeley in 2019. His research focus is interactive prediction and planning for autonomous driving, and his research interests span robotics, control, computer vision and machine learning. He has been coordinating the research activities in Autonomous Driving Group in Mechanical Systems Control Lab for years, from perception and prediction to decision and control on real autonomous vehicles. One of his publications on probabilistic prediction received the Best Student Paper Award in IEEE Intelligent Vehicle Symposium 2018 (IV’18). He is the lead author of the INTERACTION dataset, which provides highly interactive driving behavior in various complex scenarios from different countries. He is a key organizer of the prediction challenge based on the INTERACTION dataset as a NeurIPS’20 Competition. He also organized several workshops on Behavior Prediction and Decision (IV’19), Prediction Dataset and Benchmark (iROS’19), and Socially Compatible Behavior Generation (IV’20).

Abstract 38: Q&A: Wei Zhan in Workshop on AI for Autonomous Driving (AIAD), Zhan 01:35 PM
Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYyU1RDbzBTVmQT09

Abstract 39: Panel Discussion 2 in Workshop on AI for Autonomous Driving (AIAD), Hays, Bagnell, Urtasun 01:40 PM
Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYyU1RDbzBTVmQT09

Abstract 40: Closing remark (best paper award: sponsored by NVIDIA) in Workshop on AI for Autonomous Driving (AIAD), Chao, McAllister, Li, Gaidon, Kreiss 02:10 PM
Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYyU1RDbzBTVmQT09

Abstract 41: Paper Q&A session 2 in Workshop on AI for Autonomous Driving (AIAD), 02:15 PM
Zoom webinar: https://us02web.zoom.us/j/83855151644?pwd=TGhQRXppZ1pKUmxURHYyU1RDbzBTVmQT09

Abstract 37: Invited Talk: INTERPRET: INTERACTION-dataset-based PREdiciOn Challenge (Wei Zhan) in Workshop on AI for Autonomous Driving (AIAD), Zhan 01:20 PM

Video: https://.slideslive.com/38930879/interpret-interactiondataset-based-prediction-challenge

Abstract: It is a consensus in both academia and industry that behavior prediction is one of the most challenging problems blocking the realization of fully autonomous vehicles. It is a key asset for the behavior-related research community to have motion datasets with highly interactive driving behavior and critical situations in complex scenarios with different driving cultures. Prediction benchmarks with comprehensive evaluations are also crucial. This talk presents the INTERACTION dataset, which provides the highly accurate trajectories of various road users with densely interactive and critical behavior from different countries. Corresponding HD maps with full semantics of lane connections and traffic rules are also included in the dataset. The prediction challenge based on the INTERACTION dataset, INTERPRET as a NeurIPS’20 Competition, is also presented in this talk. The challenge offers multiple tracks to test the capabilities of the prediction model on data approximation, generalizability, as well as fatality in open-loop and closed-loop. The results on the leaderboard in the preliminary stage of the challenge are also briefly mentioned.

Zoom links:
Multi-agent Graph Reinforcement Learning for Connected Automated Driving
https://us02web.zoom.us/j/83693184668?pwd=SEZWTGqjT1oaFZORzVQXUvNQ1ZWd0UGJVSAA%
Challenges in Deploying and Monitoring Machine Learning Systems

**Alessandra Tosi, Nathan Korda, Neil Lawrence**

Fri Jul 17, 05:00 AM

Until recently, Machine Learning has been mostly applied in industry by consulting academics, data scientists within larger companies, and a number of dedicated Machine Learning research labs within a few of the world’s most innovative tech companies. Over the last few years we have seen the dramatic rise of companies dedicated to providing Machine Learning software-as-a-service tools, with the aim of democratizing access to the benefits of Machine Learning. All these efforts have revealed major hurdles to ensuring the continual delivery of good performance from deployed Machine Learning systems. These hurdles range from challenges in MLOps, to fundamental problems with deploying certain algorithms, to solving the legal issues surrounding the ethical involvement in letting algorithms make decisions for your business.

This workshop will invite papers related to the challenges in deploying and monitoring ML systems. It will encourage submission on: subjects related to MLOps for deployed ML systems (such as testing ML systems, deploying ML systems, monitoring ML systems, debugging ML models, deploying ML at scale); subjects related to the ethics around deploying ML systems (such as ensuring fairness, trust and transparency of ML systems, providing privacy and security on ML Systems); useful tools and programming languages for deploying ML systems; specific challenges relating to deploying reinforcement learning in ML systems and finally data challenges for deployed ML systems.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:00 AM</td>
<td>Opening remarks</td>
</tr>
<tr>
<td>Tosi, Korda</td>
<td></td>
</tr>
<tr>
<td>05:10 AM</td>
<td>Deploying Machine Learning Models in a Developing Country</td>
</tr>
<tr>
<td>Mwebaze</td>
<td></td>
</tr>
<tr>
<td>05:55 AM</td>
<td>System-wide Monitoring Architectures with Explanations</td>
</tr>
<tr>
<td>Gilpin</td>
<td></td>
</tr>
<tr>
<td>06:20 AM</td>
<td>First Break</td>
</tr>
<tr>
<td>06:50 AM</td>
<td>Bridging the gap between research and production in machine learning</td>
</tr>
<tr>
<td>Nguyen</td>
<td></td>
</tr>
<tr>
<td>07:20 AM</td>
<td>Monitoring and explainability of models in production</td>
</tr>
<tr>
<td>Klaise</td>
<td></td>
</tr>
<tr>
<td>07:30 AM</td>
<td>Gradient-Based Monitoring of Learning Machines</td>
</tr>
<tr>
<td>Liu</td>
<td></td>
</tr>
<tr>
<td>07:40 AM</td>
<td>Not Your Grandfather's Test Set: Reducing Labeling Effort for Testing</td>
</tr>
<tr>
<td>Taskazan</td>
<td></td>
</tr>
<tr>
<td>07:50 AM</td>
<td>Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models</td>
</tr>
<tr>
<td>Anthony</td>
<td></td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Serverless inferencing on Kubernetes</td>
</tr>
<tr>
<td>Cox</td>
<td></td>
</tr>
<tr>
<td>08:10 AM</td>
<td>Do You Sign Your Model?</td>
</tr>
<tr>
<td>Aramoon</td>
<td></td>
</tr>
<tr>
<td>08:20 AM</td>
<td>PareCO: Pareto-aware Channel Optimization for Slimmable Neural Networks</td>
</tr>
<tr>
<td>Chin</td>
<td></td>
</tr>
</tbody>
</table>
For reconciling system-wide errors, I developed a comprehensive architecture that I call "Anomaly Detection through Explanations" (ADE). The ADE architecture contributes an explanation synthesizer that produces an argument tree, which in turn can be traced and queried to determine the support of a decision, and to construct counterfactual explanations. I have applied this methodology to detect incorrect labels in semi-autonomous vehicle data, and to reconcile inconsistencies in simulated anomalous driving scenarios.

In conclusion, I discuss the difficulties in evaluating these types of monitoring systems. I argue that meaningful evaluation tasks should be dynamic: designing collaborative tasks (between a human and machine) that require explanations for success.

Abstract 5: Bridging the gap between research and production in machine learning in Challenges in Deploying and Monitoring Machine Learning Systems, Nguyen 06:50 AM

Machine learning has found increasing use in the real world, and yet a framework for productionizing machine learning algorithms is lacking. This talk discusses how companies can bridge the gap between research and production in machine learning. It starts with the key differences between the research and production environments: data, goals, compute requirements, and evaluation metrics. It also breaks down the different phases of a machine learning production cycle, the infrastructure currently available for the process, and the industry best practices.

Live presentation

Abstract 6: Monitoring and explainability of models in production in Challenges in Deploying and Monitoring Machine Learning Systems, Klaise 07:20 AM

The machine learning lifecycle extends beyond the deployment stage. Monitoring deployed models is crucial for continued provision of high quality machine learning enabled services. Key areas include model performance and data monitoring, detecting outliers and data drift using statistical techniques, and providing explanations of historic predictions. We discuss the challenges to successful implementation of solutions in each of these areas with some recent examples of production ready solutions using open source tools.

Abstract 7: Gradient-Based Monitoring of Learning Machines in Challenges in Deploying and Monitoring Machine Learning Systems, Liu 07:30 AM

The widespread use of machine learning algorithms calls for automatic change detection algorithms to monitor their behaviour over time. As a machine learning algorithm learns from a continuous, possibly evolving, stream of data, it is desirable and often critical to supplement it with a companion change detection algorithm to facilitate its monitoring and control. We present a generic score-based change detection method that can detect a change in any number of (hidden) components of a machine learning model trained via empirical risk minimization. This proposed statistical hypothesis test can be readily implemented for such models designed within a differentiable programming framework. We establish the consistency of the hypothesis test and show how to calibrate it based on our theoretical results. We illustrate the versatility of the approach on synthetic and real data.

Abstract 8: Not Your Grandfather’s Test Set: Reducing Labeling Effort for Testing in Challenges in Deploying and Monitoring Machine Learning Systems, Taskazan 07:40 AM

The\textsuperscript{2}
Building and maintaining high-quality test sets remains a laborious and expensive task. As a result, test sets in the real world are often not properly kept up to date and drift from the production traffic they are supposed to represent. The frequency and severity of this drift raise serious concerns over the value of manually labelled test sets in the QA process.

This paper proposes a simple but effective technique that drastically reduces the effort needed to construct and maintain a high-quality test set (reducing labelling effort by 80-100% across a range of practical scenarios). This result encourages a fundamental rethinking of the testing process by both practitioners, who can use these techniques immediately to improve their testing and researchers who can help address many of the open questions raised by this new approach.


Deep learning (DL) can achieve impressive results across a wide variety of tasks, but this often comes at the cost of training models for extensive periods on specialized hardware accelerators. This energy-intensive workload has seen immense growth in recent years. Machine learning (ML) may become a significant contributor to climate change if this exponential trend continues. If practitioners are aware of their energy and carbon footprint, then they may actively take steps to reduce it whenever possible. In this work, we present carbonTracker, a tool for tracking and predicting the energy and carbon footprint of training DL models. We propose that energy and carbon footprint of model development and training is reported alongside performance metrics using tools like carbonTracker. We hope this will promote responsible computing in ML and encourage research into energy-efficient deep neural networks.

Abstract 10: Serverless Inference on Kubernetes in Challenges in Deploying and Monitoring Machine Learning Systems. Cox 08:00 AM

Organisations are increasingly putting machine learning models into production at scale. The increasing popularity of serverless scale-to-zero paradigms presents an opportunity for deploying machine learning models to help mitigate infrastructure costs when many models may not be in continuous use. We will discuss the KFServing project which builds on the KNative serverless paradigm to provide a serverless machine learning inference solution that allows a consistent and simple interface for data scientists to deploy their models. We will show how it solves the challenges of autoscaling GPU based inference and discuss some of the lessons learnt from using it in production.

Abstract 11: Do You Sign Your Model? in Challenges in Deploying and Monitoring Machine Learning Systems. Aramoon 08:10 AM

Engineering a top-notch deep neural network (DNN) is an expensive procedure which involves collecting data, hiring human resources with expertise in machine learning, and providing high computational resources. For that reason, DNNs are considered as valuable Intellectual Properties (IPs) of the model vendors. To ensure a reliable commercialization of these products, it is crucial to develop techniques to protect model vendors against IP infringements. One of such techniques that recently has shown great promise is digital watermarking. In this paper, we present GradSigns, a novel watermarking framework for DNNs. GradSigns embeds owner's signature into gradient of cross-entropy cost function with respect to inputs to the model. Our approach has negligible impact on the performance of the protected model, and can verify ownership of remotely deployed models through prediction APIs. We evaluate GradSigns on DNNs trained for different image classification tasks using CIFAR-10, SVHN and YTF datasets, and experimentally show that unlike existing methods, GradSigns is robust against counter-watermark attacks, and can embed large amount of information into DNNs.

Abstract 12: PareCO: Pareto-Aware Channel Optimization for Slimmable Neural Networks in Challenges in Deploying and Monitoring Machine Learning Systems. Chin 08:20 AM

Slimmable neural networks have been proposed recently for resource-constrained settings such as mobile devices as they provide a flexible trade-off front between prediction error and computational cost (such as the number of floating-point operations or FLOPs) with the same storage cost as a single model. However, current slimmable neural networks use a single width-multiplier for all the layers to arrive at sub-networks with different performance profiles, which neglects that different layers affect the network's prediction accuracy differently and have different FLOP requirements. We formulate the problem of optimizing slimmable networks from a multi-objective optimization lens, which leads to a novel algorithm for optimizing both the shared weights and the width-multipliers for the sub-networks. While slimmable neural networks introduce the possibility of only maintaining a single model instead of many, our results make it more realistic to do so by improving their performance.


The development and deployment of machine learning systems can be executed easily with modern tools, but the process is typically rushed and means-to-an-end. The lack of diligence can lead to technical debt, scope creep and misaligned objectives, model misuse and failures, and expensive consequences. Engineering systems, on the other hand, follow well-defined processes and testing standards to streamline development for high-quality, reliable results. The extreme is spacecraft systems, where mission critical measures and robustness are ingrained in the development process. Drawing on experience in both spacecraft engineering and AI/ML (from research through product), we propose a proven systems engineering approach for machine learning development and deployment. Our "Technology Readiness Levels for ML" (TRL4ML) framework defines a principled process to ensure robust systems while being streamlined for ML research and product, including key distinctions from traditional software engineering. Even more, TRL4ML defines a common language for people across the organization to work collaboratively on ML technologies.


Q&A live session for the contributed talks that have been played in the previous session. Each poster presenter is in a separate Zoom Meeting.

- "Monitoring and explainability of models in production", Klaise, [Join Zoom](https://us02web.zoom.us/j/86768187826?pwd=SzkyYjd1SmpVa1iQNVRUNUI1TmNjZGQ9)
- "Gradient-Based Monitoring of Learning Machines", Liu, [Join Zoom](https://washington.zoom.us/j/93560179317)
A major challenge in deploying machine learning algorithms for decision-making problems is the lack of guarantee for the performance of their resulting policies, especially those generated during the initial exploratory phase of these algorithms. Online decision-making algorithms, such as those in bandits and reinforcement learning (RL), learn a policy while interacting with the real system. Although these algorithms will eventually learn a good or an optimal policy, there is no guarantee for the performance of their intermediate policies, especially at the very beginning, when they perform a large amount of exploration. Thus, in order to increase their applicability, it is important to control their exploration and to make it more conservative.

To address this issue, we define a notion of safety that we refer to as safety w.r.t. a baseline. In this definition, a policy considered to be safe if it performs at least as well as a baseline, which is usually the current strategy of the company. We formulate this notion of safety in bandits and RL and show how it can be integrated into these algorithms as a constraint that must be satisfied uniformly in time. We derive contextual linear bandits and RL algorithms that minimize their regret, while ensure that at any given time, their expected sum of rewards remains above a fixed percentage of the expected sum of rewards of the baseline policy. This fixed percentage depends on the amount of risk that the manager of the system is willing to take. We prove regret bounds for our algorithms that control the cost of satisfying the constraint (conservative exploration) can be controlled. Finally, we report experimental results to validate our theoretical analysis. We conclude the talk by discussing a few other constrained bandit formulations.

MLRetrospectives: A Venue for Self-Reflection in ML Research

Jessica Forde, Jesse Dodge, Mayoope Jaiswal, Ryan Lowe, Rosanne Liu, Joelle Pineau, Yoshua Bengio

 Fri Jul 17, 05:50 AM

The ICML Workshop on Retrospectives in Machine Learning will build upon the success of the 2019 NeurIPS Retrospectives workshop to further encourage the publication of retrospectives. A retrospective of a paper or a set of papers, by its author, takes the form of an informal paper. It provides a venue for authors to reflect on their previous publications, to talk about how their thoughts have changed following publication, to identify shortcomings in their analysis or results, and to discuss resulting extensions. The overarching goal of MLRetrospectives is to improve the science, openness, and accessibility of the machine learning field, by widening what is publishable and helping to identify opportunities for improvement. Retrospectives also give researchers and practitioners unable to attend conferences access to the author’s updated understanding of their work, which would otherwise only be accessible to their immediate circle. The machine learning community would benefit from retrospectives on much of the research which shapes
our field, and this workshop will present an opportunity for a few retrospectives to be presented.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:20 AM</td>
<td>Welcome</td>
<td>Lowe, Forde</td>
</tr>
<tr>
<td>05:30 AM</td>
<td>Invited Talk: Bernhard Scholkopf</td>
<td>Scholkopf</td>
</tr>
<tr>
<td>05:50 AM</td>
<td>Q&amp;A: Bernhard Scholkopf</td>
<td>Scholkopf, Jaiswal</td>
</tr>
<tr>
<td>06:00 AM</td>
<td>Invited Talk: Pascale Fung</td>
<td>FUNG</td>
</tr>
<tr>
<td>06:20 AM</td>
<td>Q&amp;A: Pascale Fung</td>
<td>FUNG, Liu</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Invited Talk: Deborah Raji</td>
<td>Raji</td>
</tr>
<tr>
<td>06:50 AM</td>
<td>Q&amp;A: Deborah Raji</td>
<td>Raji, Forde</td>
</tr>
<tr>
<td>07:00 AM</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Contributed Talk: Jonathan Frankie</td>
<td>Frankie</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>Q&amp;A: Jonathan Frankie</td>
<td>Frankie, Jaiswal</td>
</tr>
<tr>
<td>07:30 AM</td>
<td>Invited Talk: Dani Yogatama</td>
<td>Yogatama</td>
</tr>
<tr>
<td>07:50 AM</td>
<td>Q&amp;A: Dani Yogatama</td>
<td>Yogatama, Dodge, Forde</td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Panel Session</td>
<td>Forde</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited Talks: Anima Anandakumar</td>
<td>Anandkumar</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Q&amp;A: Anima Anandakumar</td>
<td>Anandkumar, Forde</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Lunch Break</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited Talk: Margaret Mitchell</td>
<td>Dodge</td>
</tr>
<tr>
<td>11:20 AM</td>
<td>Q&amp;A: Margaret Mitchell</td>
<td>Dodge</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Invited Talk: Chris Maddison</td>
<td>Maddison</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Q&amp;A: Chris Maddison</td>
<td>Maddison, Forde, Dodge</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Contributed Talk: Jason Hartford</td>
<td>Hartford</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Q&amp;A: Jason Hartford</td>
<td>Hartford, Dodge</td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Invited Talk: Peter Henderson</td>
<td>Henderson</td>
</tr>
<tr>
<td>01:20 PM</td>
<td>Q&amp;A: Peter Henderson</td>
<td>Henderson, Jaiswal, Lowe</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Brainstorming &amp; Closing</td>
<td>Jaiswal, Lowe, Dodge, Forde, Liu</td>
</tr>
</tbody>
</table>

Workshop on Continual Learning

Haytham Fayek, Arslan Chaudhry, David Lopez-Paz, Eugene Belilovsky, Jonathan Schwarz, Marc Pickett, Rahaf Aljundi, Sayna
11:30 AM  Invited Talk: Claudia Clopath "Continual learning though consolidation â€“ a neuroscience angle"

12:00 PM  Live Q&A: Claudia Clopath

12:05 PM  Spotlight Talk: Deep Reinforcement Learning amidst Lifelong Non-Stationarity

12:20 PM  Invited Talk: Jeff Clune

12:50 PM  Live Q&A: Jeff Clune

12:55 PM  Spotlight Talk: Supermasks in Superposition

01:10 PM  Live Invited Talk: Alexi Efros "Imagining a Post-Dataset Era"

01:40 PM  Live Q&A: Alexi Efros

01:45 PM  Best paper: Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics

02:00 PM  Closing remarks

02:05 PM  Poster Session 2

N/A  Continual Deep Learning by Functional Regularisation of Memorable Past

N/A  Online Inducing Points Selection for Gaussian Processes

N/A  Task-Agnostic Continual Learning via Stochastic Synapses

N/A  Routing Networks with Co-training for Continual Learning

N/A  UNCLEAR: A Straightforward Method for Continual Reinforcement Learning

N/A  Understanding Regularisation Methods for Continual Learning

N/A  SOLA: Continual Learning with Second-Order Loss Approximation

N/A  Continual Reinforcement Learning with Multi-Timescale Replay

N/A  Variational Auto-Regressive Gaussian Processes for Continual Learning

N/A  Beyond Catastrophic Forgetting in Continual Learning: An Attempt with SVM

N/A  Disentanglement of Color and Shape Representations for Continual Learning

N/A  On Class Orderings for Incremental Learning

N/A  Continual Learning from the Perspective of Compression

N/A  Continual Learning in Human Activity Recognition: An empirical analysis of Regularization

N/A  Supermasks in Superposition

N/A  Combining Variational Continual Learning with FiLM Layers

N/A  Task Agnostic Continual Learning via Meta Learning

N/A  Variational Beam Search for Continual Learning

N/A  Online Hyperparameter Tuning for Multi-Task Learning

N/A  Evaluating Agents Without Rewards

N/A  Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics

N/A  A General Framework for Continual Learning of Compositional Structures

N/A  Attention Option-Critic

N/A  Visually Grounded Continual Learning of Compositional Phrases

N/A  Understanding the Role of Training Regime in Continual Learning
In recent years we have seen an explosion of approaches that aim at transferring information between different learning tasks, in particular meta-learning and continual or lifelong learning. In my talk, I discuss ways to study these formally, using tools from learning theory that abstract away the specific details of implementation. In particular, I will discuss which assumptions one has to make on the tasks to be learned in order to guarantee a successful transfer of information.

Abstract 3: Live Q&A: Christoph H. Lampert in Workshop on Continual Learning, 06:35 AM

Ask your questions here: https://app.sli.do/event/izt9dbaz/live/questions

Abstract 4: Invited Talk: Razvan Pascanu "Continual Learning from an Optimization/Learning-dynamics perspective" in Workshop on Continual Learning, Pascanu 06:55 AM

Continual learning is usually described through a list of desiderata, however some of the "wants" on this list are in contradiction with each other, hence a solution to continual learning implies finding suitable trade-offs between the different objectives. Such trade-offs can be given by grounding ourselves into a particular domain or set of tasks. Alternatively, I believe, one can also rely on framing or looking at continual learning through different perspectives to gain this grounding. In this talk I’m looking at optimization and learning dynamics. From this perspective, continual learning can be seen as looking for a more suitable credit assignment mechanism for learning, one that does not rely on tug-of-war dynamics that result from gradient based optimization techniques. I exemplify in what sense this grounds us, and present a few recent projects I’ve been involved in that could be thought of as looking at continual learning from this perspective.

Abstract 5: Live Q&A: Razvan Pascanu in Workshop on Continual Learning, 07:25 AM

Ask your questions here: https://app.sli.do/event/cye40uex/live/questions

Abstract 6: Invited Talk: Bing Liu "Learning on the Job in the Open World" in Workshop on Continual Learning, Liu 07:45 AM

In existing machine learning (ML) applications, once a model is built it is deployed to perform its intended task. During the application, the model is fixed due to the closed-world assumption of the classic ML paradigm â€“ everything seen in testing/application must have been seen in training. However, many real-life environments - such as those for chatbots and self-driving cars - are full of unknown, which are called the open environments/worlds. We humans can deal with such environments comfortably - detecting unknowns and learning them continuously in the interaction with other humans and the environment to adapt to the new environment and to become more and more knowledgeable. In fact, we humans never stop learning. After formal education, we continue to learn on the job or while working. AI systems should have the same on-the-job learning capability. It is impossible for them to rely solely on manually labeled data and offline training to deal with the dynamic open world. This talk discusses this problem and presents some initial work in the context of natural language processing.

Abstract 9: Live Q&A: Bing Liu in Workshop on Continual Learning, 08:15 AM

Ask your questions here: https://app.sli.do/event/5g97klgd/live/questions

Abstract 11: Panel Discussion in Workshop on Continual Learning, 09:00 AM

Ask your questions here: https://app.sli.do/event/3d5xogjl/live/questions

Abstract 13: Invited Talk: Claudia Clopath "Continual learning though consolidation â€“ a neuroscience angle" in Workshop on Continual Learning, Clopath 11:30 AM

I will review the different mechanisms the brain might use to mitigate catastrophic forgetting in the brain and present a couple of brain-inspired agents in a reinforcement learning set up.

[Update] Claudia kindly asked us to keep this talk accessible for a limited time only. Therefore, this talk will no longer be available for you to watch.

Abstract 14: Live Q&A: Claudia Clopath in Workshop on Continual Learning, 12:00 PM

Ask your questions here: https://app.sli.do/event/elucy8a2/live/questions

Abstract 16: Invited Talk: Jeff Clune in Workshop on Continual Learning, Clune 12:20 PM

A dominant trend in machine learning is that hand-designed pipelines are replaced by higher-performing learned pipelines once sufficient compute and data are available. I argue that trend will apply to machine learning itself, and thus that the fastest path to truly powerful AI is to create AI-generating algorithms (AI-GAs) that on their own "learn" to solve the hardest AI problems. This paradigm is an all-in bet on meta-learning. After introducing these ideas, the talk focuses on one example of this paradigm: Learning to Continually Learn. I describe a Neuromodulated Meta-Learning algorithm (ANML), which uses meta-learning to try to solve catastrophic forgetting, producing state-of-the-art results.

Abstract 17: Live Q&A: Jeff Clune in Workshop on Continual Learning, 12:50 PM

Ask your questions here: https://app.sli.do/event/oivbvz6e/live/questions

Abstract 19: Live Invited Talk: Alexi Efros "Imagining a Post-Dataset Era" in Workshop on Continual Learning, Efros 01:10 PM

Large-scale datasets have been key to the progress in fields like computer vision during the 21st century. Yet, the over-reliance on datasets has brought new challenges, such as various dataset biases,
Abstract 20: Live Q&A: Alexi Efros in Workshop on Continual Learning, 01:40 PM

Ask your questions here: https://app.sli.do/event/pxks1d8c/live/questions

Abstract 24: Continual Deep Learning by Functional Regularisation of Memorable Past in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/84040119219?pwd=M1lqY091c2g1dBNaThVL1d1TzRZQzA9

Abstract 25: Online Inducing Points Selection for Gaussian Processes in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85369361995?pwd=YjY5RTFaN1FsM0FiM3BmRW9uRlkzY2ZzNQ==

Abstract 26: Task-Agnostic Continual Learning via Stochastic Synapses in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85112351300?pwd=TTdtK0k4UGhUS3N3WnxUTcrayWnpQPQ==

Abstract 27: Routing Networks with Co-training for Continual Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/83415438372?pwd=R3FDajNiR1U0FSTFBKY1RrTTWFWyZzNQ==

Abstract 28: UNCLEAR: A Straightforward Method for Continual Reinforcement Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/88604978097?pwd=eLpVz3VWTSpVcDczY2hvncUg3Z2ZzNQ==

Abstract 29: Understanding Regularisation Methods for Continual Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/86970194383?pwd=amhZRWcyNXVscDBWN25WRkZaNXNpZzA9

Abstract 30: SOLA: Continual Learning with Second-Order Approximation in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85871137797?pwd=aTRqZG5sWGZsWGZDNw5VVkxxbGdJSkNHQzA9

Abstract 31: Continual Reinforcement Learning with Multi-Timescale Replay in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85486198097?pwd=ZnVnQlRObjNsWE1sYmd4VNlVcTRHbQ==

Abstract 32: Variational Auto-Regressive Gaussian Processes for Continual Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/86777903933?pwd=NjU1QlFmdkE5Q3poRzRzVnRXmQzT09

Abstract 33: Beyond Catastrophic Forgetting in Continual Learning: An Attempt with SVM in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85785774573?pwd=cXhmVnZxR2g1TH9cVAzcWNIWTgwUT09

Abstract 34: Disentanglement of Color and Shape Representations for Continual Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/81201880957?pwd=RzR2TDZVNmhXSEoyUzU0cRlB1VUQz

Abstract 35: On Class Orderings for Incremental Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/89543456102?pwd=czQreW91lRc3bTIVE56YxklenZjd09

Abstract 36: Continual Learning from the Perspective of Compression in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/83517402371?pwd=TlRxM3EzcSarbHBjTWhxoa9sTTFZzNQ==

Abstract 37: Continual Learning in Human Activity Recognition: An Empirical analysis of Regularization in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/82763897108?pwd=WTItTdvWmNzVDR4UlyNjUQys5dYzA9

Abstract 38: Supermasks in Superposition in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/86564942704?pwd=ODc3THZRVXVoMBZGlStsYpaTRMDzA9

Abstract 39: Combining Variational Continual Learning with FiLM Layers in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/83405630312?pwd=cWM0ODkrRihwUzzFFU0BOakJOZfX1J

Abstract 40: Task Agnostic Continual Learning via Meta Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85298431987?pwd=WHVlU0JocGtyUFOxcWFELm03YTMyZzA9

Abstract 41: Variational Beam Search for Continual Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/86677693560?pwd=NU1aRFxNWXJySFTaVTh3hyWFBoLDc5

Abstract 42: Online Hyperparameter Tuning for Multi-Task Learning in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/85656031477?pwd=dXUybm9aM2JkejBXTFYvSNGaZHfadzA9

Abstract 43: Evaluating Agents Without Rewards in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/84217423656?pwd=bk1CaFpsaXQ0UERVaTeSyWSZxM2152

Abstract 44: Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/84436400626?pwd=WTd5WDdvakxYYz1SMWlhWNltcDRzNQ==

Abstract 45: A General Framework for Continual Learning of Compositional Structures in Workshop on Continual Learning, N/A

https://us02web.zoom.us/j/86173688177?pwd=MEVDZ1daZHpN2ROYVBUeExCAIEz

Abstract 46: Attention Option-Critic in Workshop on Continual Learning, N/A
Extreme classification is a rapidly growing research area focusing on multi-class and multi-label problems, where the label space is extremely large. It brings many diverse approaches under the same umbrella including natural language processing (NLP), computer vision, information retrieval, recommendation systems, computational advertising, and embedding methods. Extreme classifiers have been deployed in many real-world applications in the industry ranging from language modelling to document tagging in NLP, face recognition to learning universal feature representations in computer vision, etc.

Moreover, extreme classification finds application in recommendation, tagging, and ranking systems since these problems can be reformulated as multi-label learning tasks where each item to be ranked or recommended is treated as a separate label. Such reformulations have led to significant gains over traditional collaborative filtering and content-based recommendation techniques.

The proposed workshop aims to offer a timely collection of information to benefit the researchers and practitioners working in the aforementioned research fields of core supervised learning, theory of extreme classification, as well as application domains. These issues are well-covered by the Topics of Interest in ICML 2020. The workshop aims to bring together researchers interested in these areas to encourage discussion, facilitate interaction and collaboration and improve upon the state-of-the-art in extreme classification. The workshop will provide plethora of opportunities for research discussions, including poster sessions, invited talks, contributed talks, and a panel. During the panel the speakers will discuss challenges & opportunities in the field of extreme classification, in particular: 1) how to deal with the long tail labels problem?, 2) how to effectively combine deep learning approaches with extreme multi-label classification techniques?, 3) how to develop the theoretical foundations for this area? We expect a healthy participation from both industry and academia.
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:50 AM</td>
<td>Spotlight Talk 4 - Generalizing across (in)visible spectrum</td>
</tr>
<tr>
<td>10:55 AM</td>
<td>Spotlight Talk 5 - Extreme Regression for Ranking &amp; Recommendation</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Spotlight Talk 6 - Heteroskedastic and Imbalanced Deep Learning with Adaptive Regularization</td>
</tr>
<tr>
<td>11:05 AM</td>
<td>Break 2</td>
</tr>
<tr>
<td>11:05 AM</td>
<td>Poster Session</td>
</tr>
<tr>
<td>12:05 PM</td>
<td>Speaker Session</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Invited Talk 5 - Multi-Output Prediction: Theory and Practice</td>
</tr>
<tr>
<td>12:45 PM</td>
<td>Invited Talk 5 Q&amp;A - Inderjit Dhillon</td>
</tr>
<tr>
<td>12:50 PM</td>
<td>Invited Talk 6 - Efficient contextual bandits via reduction to extreme multiclass classification</td>
</tr>
<tr>
<td>01:20 PM</td>
<td>Invited Talk 6 Q&amp;A - Chicheng Zhang</td>
</tr>
<tr>
<td>01:25 PM</td>
<td>Poster Session</td>
</tr>
<tr>
<td>01:25 PM</td>
<td>Break 3</td>
</tr>
<tr>
<td>02:10 PM</td>
<td>Speaker Introduction</td>
</tr>
<tr>
<td>02:15 PM</td>
<td>Invited Talk 7 - Generalizing to Novel Tasks in the Low-Data Regime</td>
</tr>
<tr>
<td>02:45 PM</td>
<td>Invited Talk 7 Q&amp;A - Jure Leskovec</td>
</tr>
<tr>
<td>02:50 PM</td>
<td>Discussion Panel</td>
</tr>
</tbody>
</table>

Abstracts (9):


In this talk we propose the DeepXML framework for deep extreme multi-label learning and apply it to short-text document classification. We demonstrate that DeepXML can: (a) be used to analyze seemingly disparate deep extreme classifiers; (b) can lead to improvements in leading algorithms such as XML-CNN & MACH when they are recast in the proposed framework; and (c) can lead to a novel algorithm called Astec which can be up to 12% more accurate and up to 40x faster to train than the state-of-the-art for short text document classification. Finally, we show that when flighted on Bing, Astec can be used for personalized search, ads and recommendation for billions of users. Astec can handle billions of events per day, can process more than a hundred thousand events per second and leads to a significant improvement in key metrics as compared to state-of-the-art methods in production.

Abstract 5: Invited Talk 2 - Historical perspective on extreme classification in language modeling - Tomas Mikolov in Workshop on eXtreme Classification: Theory and Applications, Mikolov 06:50 AM

In this talk, I will present several simple ideas that were proposed a long time ago to deal with extremely large output spaces in the language modeling. These include various types of hierarchical softmax, and other approaches that decompose the labels into smaller spaces such as sub-word language modeling.

Abstract 9: Invited Talk 3 - Extreme Classification with Logarithmic-depth Streaming Multi-label Decision Trees - Maryam Majzoubi in Workshop on eXtreme Classification: Theory and Applications, Majzoubi 08:00 AM

We consider multi-label classification where the goal is to annotate each data point with the most relevant subset of labels from an extremely large label set. Efficient annotation can be achieved with balanced tree predictors, i.e. trees with logarithmic-depth in the label complexity, whose leaves correspond to labels. Designing prediction mechanisms like such trees for real data applications is non-trivial as it needs to accommodate sending examples to multiple leaves while at the same time sustain high prediction accuracy. In this paper we develop the LdSM algorithm for the construction and training of multi-label decision trees, where in every node of the tree we optimize a novel objective function that favors balanced splits, maintains high class purity of children nodes, and allows sending examples to multiple directions but with a penalty that prevents tree over-growth. Each node of the tree is trained once the previous node is completed leading to a streaming approach for training. We analyze the proposed objective theoretically and show that minimizing it leads to pure and balanced data splits. Furthermore, we show a boosting theorem that captures its connection to the multi-label classification error. Experimental results on benchmark data sets demonstrate that our approach achieves high prediction accuracy and low prediction time and position LdSM as a competitive tool among existing state-of-the-art approaches.


This talk is about a new learned dynamic memory controller for organizing prior experiences in a way that is empirically useful for a number of downstream tasks. The controller supports logarithmic time operations and can thus be integrated into existing statistical learning algorithms as an augmented memory unit without substantially increasing training and inference computation. It also supports optional reward reinforcement, which brings a steady improvement empirically. The controller operates as a reduction to online classification, allowing it to benefit from advances in representation or architecture. This is joint
work with Wen Sun, Hal Daume, John Langford, and Paul Mineiro (published at ICML-2019).

Abstract 22: Poster Session in Workshop on eXtreme Classification: Theory and Applications, 11:05 AM

***Use the below Link and Password to joining the Zoom poster sessions. Please do not share the below links/password on any chat or public forum or with any person unregistered for ICML.***

**Password**: xc2020p

**Zoom Links**: 

Poster 1 - Unbiased Estimates of Decomposable Losses for Extreme Classification With Missing Labels

https://zoom.us/j/98614464694?pwd=bHorMlRZUmNkNXUvTnNjT3hKbnZkdz09

Poster 2 - Online probabilistic label trees

https://zoom.us/j/98050411730?pwd=VGxoakZsaWNLZFl3VtbzBKU1RRZz09

Poster 3 - Visualizing Classification Structure in Large-Scale Classifiers

https://zoom.us/j/92665878817?pwd=OE9MTG1cdlIDMFpSa0Vtb3FOUjB3dz09

Poster 4 - Generalizing across (in)visible spectrum

https://zoom.us/j/97890519071?pwd=MVliMWpaSW1lazVaXlqQjA2Zmp2Zz09

Poster 5 - Extreme Regression for Ranking & Recommendation

https://zoom.us/j/96685271629?pwd=MGdEN2k1YmtnMjZhUmNUZ241S2FFZz09

Poster 6 - Heteroskedastic and Imbalanced Deep Learning with Adaptive Regularization

https://zoom.us/j/92929137916?pwd=NmZiZ28veUF2SW93VDdrVEJrNTN1BZdz09


Many challenging problems in modern applications amount to finding relevant results from an enormous output space of potential candidates, for example, finding the best matching product from a large catalog or suggesting related search phrases on a search engine. The size of the output space for these problems can be in the millions to billions. Moreover, observational or training data is often limited for many of the so-called long-tail of items in the output space. Given the inherent paucity of training data for most of the items in the output space, developing machine learned models that perform well for spaces of this size is challenging. Fortunately, items in the output space are often correlated thereby presenting an opportunity to alleviate the data sparsity issue. In this talk, I will first discuss the challenges in modern multi-output prediction, including missing values, features associated with outputs, absence of negative examples, and the need to scale up to enormous data sets. Bilinear methods, such as Inductive Matrix Completion ((IMC)), enable us to handle missing values and output features in practice, while coming with theoretical guarantees. Nonlinear methods such as nonlinear IMC and DSSM (Deep Semantic Similarity Model) enable more powerful models that are used in practice in real-life applications. However, inference in these models scales linearly with the size of the output space. In order to scale up, I will present the Prediction for Enormous and Correlated Output Spaces (PECOS) framework, that performs prediction in three phases: (i) in the first phase, the output space is organized using a semantic indexing scheme, (ii) in the second phase, the indexing is used to narrow down the output space by orders of magnitude using a machine learned matching scheme, and (iii) in the third phase, the matched items are ranked by a final ranking scheme. The versatility and modularity of PECOS allows for easy plug-and-play of various choices for the indexing, matching, and ranking phases, and it is possible to ensemble various models, each arising from a particular choice for the three phases.


We create a computationally tractable algorithm for contextual bandit learning with one-dimensional continuous actions with unknown structure on the loss functions. In a nutshell, our algorithm, Continuous Action Tree with Smoothing (CATS), reduces continuous-action contextual bandit learning to cost-sensitive extreme multiclass classification, where each class corresponds to a discretized action. We show that CATS admits an online implementation that has low training and test time complexities per example, and enjoys statistical consistency guarantees under certain realizability assumptions. We also verify the efficiency and efficacy of CATS through large-scale experiments.
Abstract 31: Invited Talk 7 - Generalizing to Novel Tasks in the Low-Data Regime - Jure Leskovec in Workshop on eXtreme Classification: Theory and Applications, Leskovec: 02:15 PM

Developing algorithms that are able to generalize to a novel task given only a few labeled examples represents a fundamental challenge in closing the gap between machine- and human-level performance. The core of human cognition lies in the structured, reusable concepts that help us to rapidly adapt to new tasks and provide reasoning behind our decisions. However, existing meta-learning methods learn complex representations across prior labeled tasks without imposing any structure on the learned representations. In this talk I will discuss how meta-learning methods can improve generalization ability by learning to learn along human-interpretable concept dimensions. Instead of learning a joint unstructured metric space. We learn mappings of high-level concepts into semi-structured metric spaces, and effectively combine the outputs of independent concept learners. Experiments on diverse domains, including a benchmark image classification dataset and a novel single-cell dataset from a biological domain show significant gains over strong meta-learning baselines.

ICML 2020 Workshop on Computational Biology

Delassa Aghamirzaie, Alexander Anderson, Elham Azizi, Abdoulaye Banir© Diallo, Cassandra Burdziak, Jill Gallerher, Anshul Kundaje, Dana Pe’er, Sandhya Prabhakaran, Amine Remita, Mark Robertson-Tessi, Wesley Tansey, Julia Vogt, Yubin Xie

Fri Jul 17, 06:00 AM

The workshop will showcase recent research in the field of Computational Biology. Computational biology is an interdisciplinary field that develops and applies analytical methods, mathematical and statistical modeling and simulation to analyze and interpret vast collections of biological data, such as genetic sequences, cellular features or protein structures, and imaging datasets to make new predictions towards clinical response, discover new biology or aid drug discovery. The availability of high-dimensional data, at multiple spatial and temporal resolutions has made machine learning and deep learning methods increasingly critical for computational analysis and interpretation of the data. Conversely, biological data has also exposed unique challenges and problems that call for the development of new machine learning methods.

This workshop aims to bring together researchers working at the unique intersection of Machine Learning and Biology that include areas (and not limited to) such as computational genomics, neuroscience, pathology, radiology, evolutionary biology, population genomics, phenomics, ecology, cancer biology, causality, and representation learning and disentanglement to present recent advances and open questions to the ML community.

The workshop is a sequel to the WCB workshops we organized in the last four years ICML 2019, Long Beach , Joint ICML and IJCAI 2018, Stockholm , ICML 2017, Sydney and ICML 2016, New York as well as Workshop on Bioinformatics and AI at IJCAI 2015 Buenos Aires, IJCAI 2016 New York, IJCAI 2017 Melbourne which had excellent line-ups of talks and was well-received by the community. Every year, we received 60+ submissions. After multiple rounds of rigorous reviewing around 50 submissions were selected from which the best set of papers were chosen for Contributed talks and Spotlights and the rest were invited as Posters. We have a steadfast and growing base of reviewers making up the Program Committee. For the past two editions, a special issue of Journal of Computational Biology has been released with extended versions of a select set of accepted papers.

We invited Thomas Fuchs, Debora Marks, Fabian Theis, Olga Troyanskaya. We have received funding confirmation from PAIGE and Amazon Web Services that we intend to use for student travel awards. In past years, we have also been able to provide awards for the best poster/paper and partially contribute to the student registration fee.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:50 AM</td>
<td>Opening Remarks - Dana Pe’er</td>
</tr>
<tr>
<td>06:00 AM</td>
<td>Invited Talk 1: Fabian Theis CompBio</td>
</tr>
<tr>
<td>06:25 AM</td>
<td>Q&amp;A after invited talk 1 CompBio</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Contributed Talk 1 - Jacob C. Kimmel</td>
</tr>
<tr>
<td>06:45 AM</td>
<td>Q&amp;A after contributed talk 1</td>
</tr>
<tr>
<td>06:50 AM</td>
<td>Spotlight Set 1-1 : Neha Prasad</td>
</tr>
<tr>
<td>06:55 AM</td>
<td>Spotlight Set 1-2: Lingfei Wang</td>
</tr>
<tr>
<td>07:00 AM</td>
<td>Spotlight Set 1-3 : Kexin Huang</td>
</tr>
<tr>
<td>07:05 AM</td>
<td>Spotlight Set 1-4 : Leander Dony</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Spotlight Set 1-5 : Mohammad Lotfollahi</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>Break</td>
</tr>
<tr>
<td>07:30 AM</td>
<td>Invited Talk 2: Olga Troyanskaya</td>
</tr>
<tr>
<td>07:55 AM</td>
<td>Q&amp;A</td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Contributed Talk 2: Geoffrey Fudenberg</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Q&amp;A after contributed talk</td>
</tr>
<tr>
<td>08:20 AM</td>
<td>Highlight 1: Sanjiv Kumar Dwivedi</td>
</tr>
<tr>
<td>08:25 AM</td>
<td>Highlight 2: Iman Deznaby</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Highlight 3: Christian Matek</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Highlight 4: Serghei Mangul</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Highlight 5: Shibiao Wan</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Highlight 6: William Hsu</td>
</tr>
<tr>
<td>08:50 AM</td>
<td>Highlight 7: Jacob Schreiber</td>
</tr>
<tr>
<td>08:55 AM</td>
<td>Poster session and lunch break 1</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Poster session and lunch break 2</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited Talk 3: Thomas Fuchs</td>
</tr>
<tr>
<td>11:25 AM</td>
<td>Q&amp;A after invited talk 3</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Contributed Talk 3: Guillaume Jaume</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Q&amp;A after contributed talk 3</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Spotlight Set 3-1: Kathleen Lois Foster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:55 AM</td>
<td>Spotlight Set 3-2 : Jacob Schreiber</td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Spotlight Set 3-3 : Nic Fishman</td>
<td></td>
</tr>
<tr>
<td>12:05 PM</td>
<td>Spotlight Set 3-4 : Alex Karollus</td>
<td></td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Spotlight Set 3-5 : Pinar Demetci</td>
<td></td>
</tr>
<tr>
<td>12:15 PM</td>
<td>Poster Session and Break 3</td>
<td></td>
</tr>
<tr>
<td>01:15 PM</td>
<td>Invited Talk 4: Debora Marks</td>
<td></td>
</tr>
<tr>
<td>01:40 PM</td>
<td>Q&amp;A after invited talk 4</td>
<td></td>
</tr>
<tr>
<td>01:45 PM</td>
<td>Contributed Talk 4: Minxing Pang</td>
<td></td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Q&amp;A after contributed talk 4</td>
<td></td>
</tr>
<tr>
<td>02:05 PM</td>
<td>Award Ceremony &amp; Closing Remarks</td>
<td></td>
</tr>
</tbody>
</table>

Abstracts (27):

Abstract 2: Invited Talk 1: Fabian Theis in ICML 2020 Workshop on Computational Biology, CompBio 06:30 AM


Abstract 4: Contributed Talk 1 - Jacob C. Kimmel in ICML 2020 Workshop on Computational Biology, CompBio 06:30 AM


Abstract 6: Spotlight Set 1-1 : Neha Prasad in ICML 2020 Workshop on Computational Biology, CompBio 06:50 AM

[Optimal Transport using GANs for Lineage Tracing](https://slideslive.com/38931306/spotlight-set-11)

Abstract 7: Spotlight Set 1-2: Lingfei Wang in ICML 2020 Workshop on Computational Biology, CompBio 06:55 AM


Abstract 8: Spotlight Set 1-3 : Kexin Huang in ICML 2020 Workshop on Computational Biology, CompBio 07:00 AM

[scGNN: scRNA-seq Dropout Imputation via Induced Hierarchical Cell Similarity Graph](http://slideslive.com/38931308)

Abstract 9: Spotlight Set 1-4 : Leander Dony in ICML 2020 Workshop on Computational Biology, CompBio 07:05 AM
Abstract 10: Spotlight Set 1-5: Mohammad Lotfollahi in ICML 2020 Workshop on Computational Biology, CompBio 07:10 AM

[Variational autoencoders with flexible priors enable robust distribution learning on single-cell RNA sequencing data](https://slideslive.com/38931309)

Abstract 12: Invited Talk 2: Olga Troyanskaya in ICML 2020 Workshop on Computational Biology, CompBio 07:30 AM

[Decoding the genome with AI](https://slideslive.com/38930901/invited-talk-2)

Abstract 14: Contributed Talk 2: Geoffrey Fudenberg in ICML 2020 Workshop on Computational Biology, CompBio 08:00 AM

[Predicting 3D genome folding from DNA sequence](https://slideslive.com/38930902/predicting-3d-genome-folding-from-dna-sequence)

Abstract 16: Highlight 1: Sanjiv Kumar Dwivedi in ICML 2020 Workshop on Computational Biology, CompBio 08:20 AM

[Deriving Disease Modules from the Compressed Transcriptional Space Embedded in a Deep Autoencoder](https://slideslive.com/3893111/deriving-disease-modules-from-the-compressed-transcriptional-space-embedded-in-a-deep-autoencoder)

Abstract 17: Highlight 2: Iman Deznaby in ICML 2020 Workshop on Computational Biology, CompBio 08:25 AM


Abstract 18: Highlight 3: Christian Matek in ICML 2020 Workshop on Computational Biology, CompBio 08:30 AM

[A publicly available database for developing machine learning applications to differentiate leukocytes and recognise malignant cells in peripheral blood](https://slideslive.com/38931313/a-publicly-available-dataset-for-developing-machine-learning-applications-to-differentiate-leukocytes-and-recognise-malignant-cells-in-peripheral-blood)

Abstract 19: Highlight 4: Serghei Mangul in ICML 2020 Workshop on Computational Biology, CompBio 08:35 AM

[Profiling immunoglobulin repertoires across multiple human tissues using RNA Sequencing](https://slideslive.com/38931314/highlight-4)

Abstract 20: Highlight 5: Shibiao Wan in ICML 2020 Workshop on Computational Biology, CompBio 08:40 AM


Abstract 21: Highlight 6: William Hsu in ICML 2020 Workshop on Computational Biology, CompBio 08:45 AM


Abstract 22: Highlight 7: Jacob Schreiber in ICML 2020 Workshop on Computational Biology, CompBio 08:50 AM


Abstract 23: Poster session and lunch break 1 in ICML 2020 Workshop on Computational Biology, CompBio 08:55 AM

*Variational autoencoders with flexible priors enable robust distribution learning on single-cell RNA sequencing data : Leander Dony[](https://tum-conf.zoom.us/j/96157524115?pwd=M1VhZ1NqclNMWU9xV0U4RGc1enRLZz09)*

[Out-of-distribution prediction with disentangled representations for single-cell RNA sequencing data: Mohammad Lotfollahi](https://slideslive.com/3088198015?pwd=ZG1uUG9YdmJUWVIYVdHB3bHp)

*Cross Attentive Antibody-Antigen Interaction Prediction with Multi-task Learning: Kyohei Torn±](https://us02web.zoom.us/j/6491228470?pwd=VFNid3FDS3FDQm1KaHJrS3FDS2lqZz09)

*[Gene Expression Imputation with Generative Adversarial Imputation Nets: Ramon ViÁ±as](https://zoom.us/j/9040544941?pwd=bVE1TW5rYyt6OVRYa1RmRlFBVGFZdz09)*

*[Out-of-distribution prediction with disentangled representations for single-cell RNA sequencing data: Mohammad Lotfollahi](https://slideslive.com/3088198015?pwd=ZG1uUG9YdmJUWVIYVdHB3bHp)*

*[DNA folding features prediction with Recurrent Neural Networks using epigenetic data: Michal Rozenwald](https://zoom.us/j/92090285453)*

*[Deriving Cell Type-Specific Directed Weighted Signed Regulatory Networks from Single-Cell RNA Sequencing Data: Larisa Morales](https://kaust.zoom.us/j/3175037144)*

*[Learning Heat Diffusion for Network Alignment: Sisi Qu](https://kaust.zoom.us/j/7074930446)*

*[Identifying cross-tissue signaling between genes from biomedical literature: Aditya Jadhav](https://zoom.us/j/9040544941?pwd=VHZWWXZMekpIVUZwUi9icGJmaUZtUT09)*

*[Variational autoencoders with flexible priors enable robust distribution learning on single-cell RNA sequencing data: Leander Dony](https://tum-conf.zoom.us/j/96157524115?pwd=M1VhZ1NqclNMWU9xV0U4RGc1enRLZz09)*

*[Systematic characterization of generative models for de novo design of regulatory DNA: Nic Fishman](https://mit.zoom.us/j/8080355086?pwd=RlV2TnlOQXh1SUJnOHMRMFVoX09)*

*[GPU-Accelerated SVM Learning for Extremely Fast Massive-Scale Learning on Single-Cell RNA Sequencing Data: Leander Dony](https://tum-conf.zoom.us/j/96157524115?pwd=M1VhZ1NqclNMWU9xV0U4RGc1enRLZz09)*
Proteomics Classification: John Halloran]
https://zoom.us/j/6573983026?pwd=TXIMbHBkTXUzK3JuUzYzSTKvT2k2dz09

*[MTSplice predicts effects of genetic variants on tissue-specific splicing: Jen Chung][
https://us04web.zoom.us/j/3877262893?pwd=QTg4T2dUK2xvTTY0NS9ZU1It0ZV2dZ09

*[Auto-encoders with fibered latent spaces: A geometric approach to
batch correction: Tariq Daouda][
https://zoom.us/j/92812998341?pwd=b0h1aDcvVjxd0d4SGlhdTQ4bFR4UT09

*[Topological Methods for fMRI Data: Bastian Rieck][
https://ethz.zoom.us/j/97476472235)

*[Identification of Epitope-TCR Binding Using A Generative Adversarial
Network Model: Seojin Bang][
https://cmu.zoom.us/j/97196579441?pwd=aDI4SWN1SXBxMDk0YIVKQ0Q2ZVU2Zz09

*[Learning Cancer Progression Network from Mutation Allele
Frequencies: Amir Asiaee][ https://osu.zoom.us/j/94144845481

*[TENET: Gene network reconstruction using single cell transcriptomic
data reveals key factors for embryonic stem cell differentiation: Junil
Kim][https://ucph-ku.zoom.us/j/67574624734)

*[Determining causal interactions learned by genomic DL models with in
silico mutagenesis and Mendelian randomization: Stephen Malina][
https://zoom.us/j/9175502183?pwd=UmhoeTFkek4zOTkkNjIzZUE5ZUZz09

*[Gromov–Wasserstein Optimal Transport to Align Single-Cell
Multi-Omics Data: Rebecca Santorella][
https://brown.zoom.us/j/4492917142

*[Ledidi: Designing genome edits that induce functional activity Jacob
Schreiber][ (https://washington.zoom.us/j/2044872858)

*[Predicting Mean Ribosome Load for 5′UTR of any length using
Deep Learning: Alexander Karollus][
https://us02web.zoom.us/j/9208708398?pwd=VCR4b09NWkNoRTVNeEZLYZsvZ0dEdz09

*[Biologically-relevant transfer learning improves transcription factor
binding prediction: Manu Saraswat][
https://us04web.zoom.us/j/75724839393)

*[Slide-free MUSE Microscopy to H&E Histology Modality Conversion via
Unpaired Image-to-Image Translation GAN Models: Tanishq Abraham][
https://us02web.zoom.us/j/2585300743?pwd=ODJMNjVXVzdvM3dtk1kwdVF4Q)

*[DeepPurpose: a Deep Learning Based Drug Repurposing Toolkit:
Kexin Huang][ (https://harvard.zoom.us/j/3508373182)

*[Using deep learning on chest CT to track COVID-19 patients: Edward
Lee)][(https://brown.zoom.us/j/75724839393)

*[ChronoStrain: Sequence quality and time aware strain tracking with
shotgun metagenomic data: YounHun Kim][
https://mit.zoom.us/j/96343087302?pwd=aHJlRWhZaEF3eG5uN0FhYm5vZkE5ZUZz09

*[Supervised Tumor Cell Subtype Identification via SCAN: Russell
Kunes][https://columbiauniversity.zoom.us/j/93729574453?pwd=SGJGyH patrolsCVU2ozdz09)

*[Data-driven Variable-length Segmentation of Biological Sequences: Applications in Proteomics and Metagenomics: Ehsan
Asgar][https://berkeley.zoom.us/j/2690485065?pwd=VnBIZUzFeH9eYnVIZW4vVlIDh0
The designers of a machine learning (ML) system typically have far more power over the system than the individuals who are ultimately impacted by the system and its decisions. Recommender platforms shape the users’ preferences; the individuals classified by a model often do not have means to contest a decision; and the data required by supervised ML systems necessitates that the privacy and labour of many yield to the design choices of a few.

The fields of algorithmic fairness and human-centered ML often focus on centralized solutions, lending increasing power to system designers and operators, and less to users and affected populations. In response to the growing social-science critique of the power imbalance present in the research, design, and deployment of ML systems, we wish to consider a new set of technical formulations for the ML community on the subject of more democratic, cooperative, and participatory ML systems.

Our workshop aims to explore methods that, by design, enable and encourage the perspectives of those impacted by an ML system to shape the system and its decisions. By involving affected populations in shaping the goals of the overall system, we hope to move beyond just tools for enabling human participation and progress towards a redesign of power dynamics in ML systems.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:00 AM</td>
<td>Opening remarks</td>
</tr>
<tr>
<td>06:15 AM</td>
<td>Alâ€™s Contradiction: Kingâ€™s Radical Revolution in Values</td>
</tr>
<tr>
<td>06:45 AM</td>
<td>What does it mean for ML to be trustworthy?</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>Turning the tables on Facebook: How we audit Facebook using their own marketing tools</td>
</tr>
<tr>
<td>07:45 AM</td>
<td>Poster Session 1</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Breakout Sessions / Break</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Panel 1</td>
</tr>
</tbody>
</table>

**Abstracts (11):**

**Abstract 2: Alâ€™s Contradiction: Kingâ€™s Radical Revolution in Values in Participatory Approaches to Machine Learning, Petty 06:15 AM**

Dr. King called for a radical revolution of values in 1967. He understood that if we did not “begin the shift from a thing-oriented society to a person-oriented society,” and prioritize people over machines, computers and profit motives, we would be unable to undo the harms of racism, extreme materialism, and militarism. If we were to take Dr. King’s challenge seriously today, how might we deepen our questions, intervene in harmful technologies and slow down innovation for innovation’s sake?

**Abstract 3: What does it mean for ML to be trustworthy? in Participatory Approaches to Machine Learning, Papernot 06:45 AM**

The attack surface of machine learning is large: training data can be poisoned, predictions manipulated using adversarial examples, models exploited to reveal sensitive information contained in training data, etc. This is in large parts due to the absence of security and privacy considerations in the design of ML algorithms. Yet, adversaries have clear incentives to target these systems. Thus, there is a need to ensure that computer systems that rely on ML are trustworthy.

Fortunately, we are at a turning point where ML is still being adopted, which creates a rare opportunity to address the shortcomings of the technology before it is widely deployed. Designing secure ML requires that we have a solid understanding as to what we expect legitimate model behavior to look like. We structure our discussion around three directions, which we believe are likely to lead to significant progress.

The first encompasses a spectrum of approaches to verification and admission control, which is a prerequisite to enable fail-safe defaults in machine learning systems. The second seeks to design mechanisms for assembling reliable records of compromise that would help understand the degree to which vulnerabilities are exploited by adversaries, as well as favor psychological acceptability of machine learning applications. The third pursues formal frameworks for security and privacy in machine learning, which we argue should strive to align machine learning goals such as generalization with security and privacy desiderata like...
robustness or privacy. We illustrate these directions with recent work on model extraction, privacy-preserving ML and machine unlearning.

Abstract 4: Turning the tables on Facebook: How we audit Facebook using their own marketing tools in Participatory Approaches to Machine Learning, Sapiezynski 07:15 AM

Researchers and journalists have found many ways that advertisers can targetâ€œor excludeâ€œparticular groups of users seeing their ads on Facebook, comparatively little attention has been paid to the implications of the platform’s ad delivery process, where the platform decides which users see which ads. In this talk I will show how we audit Facebook’s delivery algorithms for potential gender and race discrimination using Facebook’s own tools designed to assist advertisers. Following these methods we find that Facebook delivers different job ads to men and women as well as white and Black users, despite inclusive targeting. We also identify how Facebook contributes to creating opinion filter bubbles by steering political ads towards users who already agree with their content.

Abstract 5: Poster Session 1 in Participatory Approaches to Machine Learning, 07:45 AM

Please check the details on our website: https://participatoryml.github.io/#poster-sessions
Discord sever: https://discord.gg/KSAwXKs

Abstract 6: Breakout Sessions / Break in Participatory Approaches to Machine Learning, 08:30 AM

Please check the details on our website: https://participatoryml.github.io/#breakout-sessions
Discord server: https://discord.gg/KSAwXKs

Abstract 8: Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services in Participatory Approaches to Machine Learning, Chouldechova 10:00 AM

Algorithmic decision-making systems are increasingly being adopted by government public service agencies. Researchers, policy experts, and civil rights groups have all voiced concerns that such systems are being deployed without adequate consideration of potential harms, disparate impacts, and public accountability practices. Yet little is known about the concerns of those most likely to be affected by these systems. In this talk I will discuss what we learned from a series of workshops conducted to better understand the concerns of affected communities in the context of child welfare services. Through these workshops we learned about the perspectives of families involved in the child welfare system, employees of child welfare agencies, and service providers.

Abstract 9: Actionable Recourse in Machine Learning in Participatory Approaches to Machine Learning, Ustun 10:30 AM

Machine learning models are often used to automate decisions that affect consumers: whether to approve a loan, a credit card application or provide insurance. In such tasks, consumers should have the ability to change the decision of the model. When a consumer is denied a loan by a credit score, for example, they should be able to alter its input variables in a way that guarantees approval. Otherwise, they will be denied the loan so long as the model is deployed, and â€œmore importantly â€œlack control over a decision that affects their livelihood. In this talk, I will formally discuss these issues in terms of a notion called recourse -- i.e., the ability of a person to change the decision of a model by altering actionable input variables. I will describe how machine learning models may fail to provide recourse due to standard practices in model development. I will then describe integer programming tools to verify recourse in linear classification models. I will end with a brief discussion on how recourse can facilitate meaningful consumer protection in modern applications of machine learning. This is joint work with Alexander Spangher and Yang Liu.

Abstract 10: Beyond Fairness and Ethics: Towards Agency and Shifting Power in Participatory Approaches to Machine Learning, Watson-Daniels 11:00 AM

When we consider power imbalances between those who craft ML systems and those most vulnerable to the impacts of those systems, what is often enabling that power is the localization of control in the hands of tech companies and technical experts who consolidate power using claims to perceived scientific objectivity and legal protections of intellectual property. At the same time, there is a legacy in the scientific community of data being wielded as an instrument of oppression, often reinforcing inequality and perpetuating injustice. At Data for Black Lives, we bring together scientists and community-based activists to take collective action using data for fighting bias, building progressive movements, and promoting civic engagement. In the ML community, people often take for granted the initial steps in the process of crafting ML systems that involve data collection, storage and access. Researchers often engage with datasets as if they appeared spontaneously with no social context. One method of moving beyond fairness metrics and generic discussions of ethics to meaningfully shifting agency to the people most marginalized is to stop ignoring the context, construction and implications of the datasets we use for research. I offer two considerations for shifting power in this way: Intentional data narratives and Data trusts - an alternative to current strategies of data governance.

Abstract 11: Panel 2 in Participatory Approaches to Machine Learning, Raji, Ustun, Chouldechova, Watson-Daniels 11:30 AM

Please check the details on our website: https://participatoryml.github.io/#poster-sessions
Discord sever: https://discord.gg/JtA55Gv

Abstract 12: Poster Session 2 in Participatory Approaches to Machine Learning, 12:15 PM

Please check the details on our website: https://participatoryml.github.io/#poster-sessions
Discord server: https://discord.gg/KSAwXKs

Abstract 13: Breakout Sessions in Participatory Approaches to Machine Learning, 01:00 PM

Please check the details on our website: https://participatoryml.github.io/#breakout-sessions
Discord server: https://discord.gg/KSAwXKs

Object-Oriented Learning: Perception, Representation, and Reasoning

Sungjin Ahn, Adam Kosiorrek, Jessica Hamrick, Sjoerd van Steenkiste, Yoshua Bengio
Objects, and the interactions between them, are the foundations on which our understanding of the world is built. Similarly, abstractions centered around the perception and representation of objects play a key role in building human-like AI, supporting high-level cognitive abilities like causal reasoning, object-centric exploration, and problem solving. Indeed, prior works have shown how relational reasoning and control problems can greatly benefit from having object descriptions. Yet, many of the current methods in machine learning focus on a less structured approach in which objects are only implicitly represented, posing a challenge for interpretability and the reuse of knowledge across tasks. Motivated by the above observations, there has been a recent effort to reinterpret various learning problems from the perspective of object-oriented representations.

In this workshop, we will showcase a variety of approaches in object-oriented learning, with three particular emphases. Our first interest is in learning object representations in an unsupervised manner. Although computer vision has made an enormous amount of progress in learning about objects via supervised methods, we believe that learning about objects with little to no supervision is preferable: it minimizes labeling costs, and also supports adaptive representations that can be changed depending on the particular situation and goal. The second primary interest of this workshop is to explore how object-oriented representations can be leveraged for downstream tasks such as reinforcement learning and causal reasoning. Lastly, given the central importance of objects in human cognition, we will highlight interdisciplinary perspectives from cognitive science and neuroscience on how people perceive and understand objects.

We have invited speakers whose research programs cover unsupervised and supervised 2-D and 3-D perception, reasoning, concept learning, reinforcement learning, as well as psychology and neuroscience. We will additionally source contributed works focusing on unsupervised object-centric representations, applications of such object-oriented representations (such as in reinforcement learning), and object-centric aspects of human cognition. To highlight and support research from a range of different perspectives, our invited speakers vary in their domain of expertise, institution, seniority, and gender. We will also encourage participation from underrepresented groups by providing travel grants courtesy of DeepMind and Kakao Brain. We are also planning to coordinate with the main conference and the speakers to provide remote access to the workshop.

### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:15 AM</td>
<td>Opening Remarks</td>
<td>Ahn</td>
</tr>
<tr>
<td>06:30 AM</td>
<td>What are Objects</td>
<td>Greff</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Attentive Grouping and Graph Neural Networks for Object-Centric Learning</td>
<td>Kipf</td>
</tr>
<tr>
<td>07:50 AM</td>
<td>Learning Affordances in Object-Centric Generative Models</td>
<td>Wu</td>
</tr>
<tr>
<td>08:10 AM</td>
<td>Learning Object-Centric Representations for High-Level Planning in Minecraft</td>
<td>James</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Capsule Networks: A Generative Probabilistic Perspective</td>
<td>Smith</td>
</tr>
<tr>
<td>08:20 AM</td>
<td>Unsupervised Object Keypoint Learning using Local Spatial Predictability</td>
<td>Gopalakrishnan</td>
</tr>
<tr>
<td>08:25 AM</td>
<td>Generative Adversarial Set Transformers</td>
<td>Stelzner</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Poster Session 1</td>
<td></td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Panel Discussion</td>
<td>Hamrick</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Object-level video understanding</td>
<td>Baradel</td>
</tr>
<tr>
<td>11:10 AM</td>
<td>“The treachery of images”: How the realness of objects affects brain activation and behavior</td>
<td>Culham</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Counterfactual Data Augmentation using Locally Factored Dynamics</td>
<td>Pits</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>12:40 PM</td>
<td>Object-Oriented Drawings</td>
<td>Dillon</td>
</tr>
<tr>
<td>01:20 PM</td>
<td>Learning from an infant’s point of view</td>
<td>Smith</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Implicit Neural Scene Representations</td>
<td>Sitzmann</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Energy-Based Models for Object-Oriented Learning</td>
<td>Mordatch</td>
</tr>
<tr>
<td>03:20 PM</td>
<td>Learning 3D Object-Oriented World Models from Unlabeled Videos</td>
<td>Crawford</td>
</tr>
<tr>
<td>03:40 PM</td>
<td>Rapid policy updating in human physical construction</td>
<td>McCarthy</td>
</tr>
<tr>
<td>03:45 PM</td>
<td>Amodal 3D Reconstruction for Robotic Manipulation via Stability and Connectivity</td>
<td>Agnew</td>
</tr>
<tr>
<td>03:50 PM</td>
<td>Hierarchical Decomposition and Generation of Scenes with Compositional Objects</td>
<td>Deng</td>
</tr>
<tr>
<td>03:55 PM</td>
<td>Conditional Set Generation with Transformers</td>
<td>Kosiorek</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Poster Session 2</td>
<td></td>
</tr>
</tbody>
</table>
Abstracts (22):

Abstract 2: What are Objects in Object-Oriented Learning: Perception, Representation, and Reasoning, Greff 06:30 AM

Recently, there has been a surge of interest for object-centric learning in neural network research. To many researchers, it seems clear that objects hold great potential for enabling more systematic generalisation, building compositional models of the world, and as grounding for language and symbolic reasoning. However, despite strong intuitions, a general definition of what constitutes an object is still lacking, and the precise notion of objects remains largely elusive. In this talk I aim to challenge some common intuitive conceptions about objects, and point to some of their subtle complexity. After that, I will present a few relevant findings from cognitive psychology regarding human object perception, and conclude by discussing a few challenges and promising approaches for incorporating objects into neural networks.

Abstract 3: Attentive Grouping and Graph Neural Networks for Object-Centric Learning in Object-Oriented Learning: Perception, Representation, and Reasoning, Kipf 07:10 AM

To enable explicit representation of objects in neural architectures, a core challenge lies in defining a mapping from input features (e.g., an image encoded by a CNN) to a set of abstract object representations. In this talk, I will discuss how attention mechanisms can be used in an iterative, competitive fashion to (a) efficiently group visual features into object slots and (b) segment temporal representations. I will further highlight how graph neural networks can be utilized to learn about interactions between objects and how object-centric models can be trained in a self-supervised fashion using contrastive losses.

Abstract 4: Learning Affordances in Object-Centric Generative Models in Object-Oriented Learning: Perception, Representation, and Reasoning, Wu 07:50 AM

Given visual observations of a reaching task together with a stick-like tool, we propose a novel approach that learns to exploit task-relevant object affordances by combining generative modelling with a task-based performance predictor. The embedding learned by the generative model captures the factors of variation in object geometry, e.g., length, width, and configuration. The performance predictor identifies sub-manifolds correlated with task success in a weakly supervised manner. Using a 3D simulation environment, we demonstrate that traversing the latent space in this task-driven way results in appropriate tool geometries for the task at hand. Our results suggest that affordances are encoded along smooth trajectories in the learned latent space. Given only high-level performance criteria (such as task success), accessing these emergent affordances via gradient descent enables the agent to manipulate learned object geometries in a targeted and deliberate way.

Abstract 5: Learning Object-Centric Representations for High-Level Planning in Minecraft in Object-Oriented Learning: Perception, Representation, and Reasoning, James 08:10 AM

We propose a method for autonomously learning an object-centric representation of a high-dimensional environment that is suitable for planning. Such abstractions can be immediately transferred between tasks that share the same types of objects, resulting in agents that require fewer samples to learn a model of a new task. We demonstrate our approach on a series of Minecraft tasks to learn object-centric representations - directly from pixel data - that can be leveraged to quickly solve new tasks. The resulting learned representations enable the use of a task-level planner, resulting in an agent capable of forming complex, long-term plans.


‘Capsule’ models try to explicitly represent the poses of objects, enforcing a linear relationship between an objects pose and those of its constituent parts. This modelling assumption should lead to robustness to viewpoint changes since the object-component relationships are invariant to the poses of the object. We describe a probabilistic generative model that encodes these assumptions. Our probabilistic formulation separates the generative assumptions of the model from the inference scheme, which we derive from a variational bound. We experimentally demonstrate the applicability of our unified objective, and the use of test time optimisation to solve problems inherent to amortised inference.

Abstract 7: Unsupervised Object Keypoint Learning using Local Spatial Predictability in Object-Oriented Learning: Perception, Representation, and Reasoning, Gopalakrishnan 08:20 AM

We propose a novel approach to representation learning based on object keypoints. It leverages the predictability of local image regions from spatial neighborhoods to identify salient regions that correspond to object parts, which are then converted to keypoints. Unlike prior approaches, this does not overly bias the keypoints to focus on a particular property of objects. We demonstrate the efficacy of our approach on Atari where we find that it learns keypoints corresponding to the most salient object parts and is more robust to certain visual distractors.

Abstract 8: Generative Adversarial Set Transformers in Object-Oriented Learning: Perception, Representation, and Reasoning, Stelzner 08:25 AM

Groups of entities are naturally represented as sets, but generative models usually treat them as independent from each other or as sequences. This either over-simplifies the problem, or imposes an order to the otherwise unordered collections, which has to be accounted for in loss computation. We therefore introduce GAST - a GAN for sets capable of generating variable-sized sets in a permutation-equivariant manner, while accounting for dependencies between set elements. It avoids the problem of formulating a distance metric between sets by using a permutation-invariant discriminator. When evaluated on a dataset of regular polygons and on MNIST point clouds, GAST outperforms graph-convolution-based GANs in sample fidelity, while showing good generalization to novel set sizes.

Abstract 9: Poster Session 1 in Object-Oriented Learning: Perception, Representation, and Reasoning, 08:30 AM

Please access the posters via the workshop website using Zoom room password: w00l

Abstract 10: Panel Discussion in Object-Oriented Learning: Perception, Representation, and Reasoning, Hamrick 09:30 AM

Suggest questions via the link below.
Understanding causes and effects in mechanical systems is an essential component of reasoning in the physical world. This work poses a new problem of counterfactual learning of object mechanics from visual input. We develop the CoPhy benchmark to assess the capacity of the state-of-the-art models for causal reasoning in a synthetic 3D environment and propose a model for learning the physical dynamics in a counterfactual setting. Having observed a mechanical experiment that involves, for example, a falling tower of blocks, a set of bouncing balls or colliding objects, we learn to predict how its outcome is affected by an arbitrary intervention on its initial conditions, such as displacing one of the objects in the scene. The alternative future is predicted given the altered past and a latent representation of the confounders learned by the model in an end-to-end fashion with no supervision. We compare against feedforward video prediction baselines and show how observing alternative experiences allows the network to capture latent physical properties of the environment, which results in significantly more accurate predictions at the level of super human performance.


Two-dimensional images are commonly used to study and model perceptual and cognitive processes because of the convenience and ease of experimental control they provide. However, real objects differ from pictures in many ways, including the potential for interaction and richer information about distance and size. Across a series of neuroimaging studies and behavioral experiments in adults, we have shown different responses to real objects than pictures. Moreover, we have found behavioral differences between real objects and pictures even in infants, suggesting that realness plays an important role in learning about objects. These results can inform the next generation of computational models as to how human brains learn to process objects in the real world.


Many dynamic processes, including common scenarios in robotic control and reinforcement learning (RL), involve a set of interacting subprocesses. Though the subprocesses are not independent, their interactions are often sparse, and the dynamics at any given time step can often be decomposed into locally independent causal mechanisms. Such local causal structures can be leveraged to improve the sample efficiency of sequence prediction and off-policy reinforcement learning. We formalize this by introducing local causal models (LCMs), which are induced from a global causal model by conditioning on a subset of the state space. We propose an approach to inferring these structures given an object-oriented state representation, as well as a novel algorithm for model-free Counterfactual Data Augmentation (CoDA). CoDA uses local structures and an experience replay to generate counterfactual experiences that are causally valid in the global model. We find that CoDA significantly improves the performance of RL agents in locally factored tasks, including the batch-constrained and goal-conditioned settings.

Abstract 14: Object-Oriented Drawings in Object-Oriented Learning: Perception, Representation, and Reasoning. Dillon 12:40 PM

Objects elicit attention in many everyday contexts, even from infancy. Objects also serve as the referents for humans’ earliest symbolic learning: language. In this talk, Iâ€™ll present my lab’s recent work with young children suggesting that objects are also prioritized in another early emerging and uniquely human symbolic expression: drawing. Iâ€™ll conclude my talk by suggesting that researchers interested in artificial intelligence may look for inspiration in human intelligence, especially when it comes to the way that humans attend to and represent objects.

Abstract 15: Learning from an infant’s point of view in Object-Oriented Learning: Perception, Representation, and Reasoning. Smith 01:20 PM

Learning depends on both the learning mechanism and the structure of the training data, yet most research in human learning and efforts in machine learning concentrate on the learning mechanisms. I will present evidence on the everyday-day ego-centric visual experiences of infants. The regularities differ fundamentally and in multiple inter-related from current approaches to training in machine learning and perhaps will offer inspiration to more powerful, more incremental, and more autonomous machine learning.

Abstract 16: Implicit Neural Scene Representations in Object-Oriented Learning: Perception, Representation, and Reasoning. Stitzmann 02:00 PM

How we represent signals has major implications for the algorithms we build to analyze them. Today, most signals are represented discretely: Images as grids of pixels, shapes as point clouds, audio as grids of amplitudes, etc. If images weren’t pixel grids - would we be using convolutional neural networks today? What makes a good or bad representation? Can we do better? I will talk about leveraging emerging implicit neural representations for complex & large signals, such as room-scale geometry, images, audio, video, and physical signals defined via partial differential equations. By embedding an implicit scene representation in a neural rendering framework and learning a prior over these representations, I will show how we can enable 3D reconstruction from only a single posed 2D image. Finally, I will show how gradient-based meta-learning can enable fast inference of implicit representations, and how the features we learn in the process are already useful to the downstream task of semantic segmentation.

Abstract 17: Energy-Based Models for Object-Oriented Learning in Object-Oriented Learning: Perception, Representation, and Reasoning. Mordatch 02:40 PM

Energy-based models are undergoing a resurgence of interest, but their applications have largely focused on generative modeling and density estimation. In this talk I will discuss application of energy-based models to object or concept oriented learning and reasoning. These models offer an elegant approach to concept composition, continual and unsupervised learning, and usage of concepts in multiple contexts. I will show examples of these advantages, and conclude with a set of future research directions.

Abstract 18: Learning 3D Object-Oriented World Models from Unlabeled Videos in Object-Oriented Learning: Perception, Representation, and Reasoning. Crawford 03:20 PM

Objects elicit attention in many everyday contexts, even from infancy. Objects also serve as the referents for humans’ earliest symbolic learning: language. In this talk, Iâ€™ll present my lab’s recent work with young children suggesting that objects are also prioritized in another early emerging and uniquely human symbolic expression: drawing. Iâ€™ll conclude my talk by suggesting that researchers interested in artificial intelligence may look for inspiration in human intelligence, especially when it comes to the way that humans attend to and represent objects.
The physical world can be decomposed into discrete 3D objects. Reasoning about the world in terms of these objects may provide a number of advantages to learning agents. For example, objects interact compositionally, and this can support a strong form of generalization. Knowing properties of individual objects and rules for how those properties interact, one can predict the effects that objects will have on one another even if one has never witnessed an interaction between the types of objects in question. The promise of object-level reasoning has fueled a recent surge of interest in systems capable of learning to extract object-oriented representations from perceptual input without supervision. However, the vast majority of such systems treat objects as 2D entities, effectively ignoring their 3D nature. In the current work, we propose a probabilistic, object-oriented model equipped with the inductive bias that the world is made up of 3D objects moving through a 3D world, and make a number of structural adaptations which take advantage of that bias. In a series of experiments we show that this system is capable not only of segmenting objects from the perceptual stream, but also of extracting 3D information about objects (e.g. depth) and of tracking them through 3D space.

Abstract 20: Rapid policy updating in human physical construction in Object-Oriented Learning: Perception, Representation, and Reasoning. McCarthy 03:40 PM

The ability to build a wide array of physical structures, from sand castles to skyscrapers, is a hallmark of human intelligence. What computational mechanisms enable humans to reason about how such structures are built? Here we conduct an empirical investigation of how people solve challenging physical assembly problems and update their policies across repeated attempts. Participants viewed silhouettes of 8 unique towers in a 2D virtual environment simulating rigid-body physics, and aimed to reconstruct each one using a fixed inventory of rectangular blocks. We found that people learned to build each target tower more accurately across repeated attempts, and that these gains reflect both group-level convergence upon a smaller set of viable policies, as well as error-dependent updating of each individual’s policy. Taken together, our study provides a novel benchmark for evaluating how well algorithmic models of physical reasoning and planning correspond to human behavior.


Learning-based 3D object reconstruction enables single- or few-shot estimation of 3D object models. For robotics this holds the potential to allow model-based methods to rapidly adapt to novel objects and scenes. Existing 3D reconstruction techniques optimize for visual reconstruction fidelity, typically measured by chamfer distance or voxel IOU. We find that when applied to realistic, cluttered robotic environments these systems produce reconstructions with low physical realism, resulting in poor task performance when used for model-based control. We propose ARM an amodal 3D reconstruction system that introduces (1) an object stability prior over the shapes of groups of objects, (2) an object connectivity prior over object shapes, and (3) a multi-channel input representation and reconstruction objective that allows for reasoning over relationships between groups of objects. By using these priors over the physical properties of objects, our system improves reconstruction quality not just by standard visual metrics, but also improves performance of model-based control on a variety of robotics manipulation tasks in challenging, cluttered environments.

Abstract 22: Hierarchical Decomposition and Generation of Scenes with Compositional Objects in Object-Oriented Learning: Perception, Representation, and Reasoning. Deng 03:50 PM

Compositional structures between parts and objects are inherent in natural scenes. Recent work on representation learning has succeeded in modeling scenes as composition of objects, but further decomposition of objects into parts and subparts has largely been overlooked. In this paper, we propose RICH, the first deep latent variable model for learning Representation of Interpretable Compositional Hierarchies. At the core of RICH is a latent scene graph representation that organizes the entities of a scene into a tree according to their compositional relationships. During inference, RICH takes a top-down approach, allowing higher-level representation to guide lower-level decomposition in case there is compositional ambiguity. In experiments on images containing multiple compositional objects, we demonstrate that RICH is able to learn the latent compositional hierarchy, generate imaginary scenes, and improve data efficiency in downstream tasks.


A set is an unordered collection of unique elements and yet many machine learning models that generate sets impose an implicit or explicit ordering. Since model performance can depend on the choice of ordering, any particular ordering can lead to sub-optimal results. An alternative solution is to use a permutation-equivariant set generator, which does not specify an order-ing. An example of such a generator is the DeepSet Prediction Network (DSPN). We introduce the Transformer Set Prediction Network (TSPN), a flexible permutation-equivariant model for set prediction based on the transformer, that builds upon and outperforms DSPN in the quality of predicted set elements and in the accuracy of their predicted sizes. We test our model on MNIST-as-point-clouds (SET-MNIST) for point-cloud generation and on CLEVR for object detection.

Abstract 24: Poster Session 2 in Object-Oriented Learning: Perception, Representation, and Reasoning. 04:00 PM

Please access the posters via the workshop website using Zoom room password: w00l
This workshop aims to highlight recent theoretical contributions, with an emphasis on addressing significant challenges on the road ahead. Such theoretical understanding is important in order to design algorithms that have robust and compelling performance in real-world applications. As part of the ICML 2020 conference, this workshop will be held virtually. It will feature keynote talks from six reinforcement learning experts tackling different significant facets of RL. It will also offer the opportunity for contributed material (see below the call for papers and our outstanding program committee). The authors of each accepted paper will prerecord a 10-minute presentation and will also appear in a poster session. Finally, the workshop will have a panel discussing important challenges in the road ahead.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Talk</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:30 AM</td>
<td>Exploration, Policy Gradient Methods, and the Deadly Triad</td>
<td>Kakade</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>A Unifying View of Optimism in Episodic Reinforcement Learning</td>
<td>Neu</td>
</tr>
<tr>
<td>08:10 AM</td>
<td>Poster Session 1</td>
<td></td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Speaker Panel</td>
<td>Szepesvari, White, Kakade, Neu, Agrawal, Krishnamurthy</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>An Off-policy Policy Gradient Theorem: A Tale About Weightings</td>
<td>White</td>
</tr>
<tr>
<td>11:20 AM</td>
<td>Short Talk 1 - Crush Optimism with Pessimism: Structured Bandits Beyond</td>
<td>Jun</td>
</tr>
<tr>
<td>11:35 AM</td>
<td>Short Talk 2 - Adaptive Discretization for Model-Based Reinforcement Learning</td>
<td>Sinclair</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Short Talk 3 - A Kernel-Based Approach to Non-Stationary Reinforcement Learning in Metric Spaces</td>
<td>Danwiche Domingues</td>
</tr>
<tr>
<td>12:05 PM</td>
<td>Short Talk 4 - Adaptive Regret for Online Control</td>
<td>Minasyan</td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Short Talk 5 - Near-Optimal Reinforcement Learning with Self-Play</td>
<td>Yu</td>
</tr>
<tr>
<td>12:35 PM</td>
<td>Short Talk 6 - Preference learning along multiple criteria: A game-theoretic perspective</td>
<td>Bhatia</td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Poster Session 2</td>
<td></td>
</tr>
</tbody>
</table>

Abstracts (12):

Abstract 1: Exploration, Policy Gradient Methods, and the Deadly Triad - Sham Kakade in Theoretical Foundations of Reinforcement Learning. Kakade 06:30 AM

Practical reinforcement learning algorithms often face the "deadly triad" [Rich Sutton, 2015]: function approximation, data efficiency (e.g. by bootstrapping value function estimates), and exploration (e.g. by off-policy learning). Algorithms which address two without the third are often ok, while trying to address all three leads to highly unstable algorithms in practice. This talk considers a policy gradient approach to alleviate these issues. In particular, we introduce the Policy Cover-Policy Gradient (PC-PG) algorithm, which provably balances the exploration vs. exploitation tradeoff, with polynomial sample complexity, using an ensemble of learned policies (the policy cover). We quantify how the relevant notion of function approximation is based on an approximation error term, under distribution shift. Furthermore, we will provide simple examples where a number of standard (and provable) RL approaches are less robust when it comes to function approximation. Time permitting, we will discuss the implications this has for more effective data re-use.

Joint work with: Alekh Agarwal, Mikael Henaff, and Wen Sun.


The principle of optimism in the face of uncertainty underpins many theoretically successful reinforcement learning algorithms. In this paper we provide a general framework for designing, analyzing and implementing such algorithms in the episodic reinforcement learning problem. This framework is built upon Lagrangian duality, and demonstrates that every model-optimistic algorithm that constructs an optimistic MDP has an equivalent representation as a value-optimistic dynamic programming algorithm. Typically, it was thought that these two classes of algorithms were distinct, with model-optimistic algorithms benefiting from a cleaner probabilistic analysis while value-optimistic algorithms are easier to implement and thus more practical. With the framework developed in this paper, we show that it is possible to get the best of both worlds by providing a class of algorithms which have a computationally efficient dynamic-programming implementation and also a simple probabilistic analysis. Besides being able to capture many existing algorithms in the tabular setting, our framework can also address largescale problems under realizable function approximation, where it
enables a simple model-based analysis of some recently proposed methods.


The goal of the talk is to discuss our recent work on an off-policy policy gradient theorem, and how it can help leverage theory for the on-policy setting for use in the off-policy setting. The key insight is to provide a more general objective for the off-policy setting, that encompasses the on-policy episodic objective. These simple generalizations make it straightforward to port and generalize ideas.

Abstract 6: Short Talk 1 - Crush Optimism with Pessimism: Structured Bandits Beyond Asymptotic Optimality in Theoretical Foundations of Reinforcement Learning, Jun 11:20 AM

We study regret minimization in stochastic structured bandits. The fact that the popular optimistic algorithms do not achieve the asymptotic instance-dependent regret optimality has recently allured researchers. On the other hand, it is known that one can achieve a bounded regret (i.e., does not grow indefinitely with ) in certain instances. Unfortunately, existing asymptotically optimal algorithms rely on forced sampling that introduces an term w.r.t. the time horizon in their regret, failing to adapt to the “easiness” of the instance. In this paper, we focus on the finite hypothesis class case and ask if one can achieve the asymptotic optimality while enjoying bounded regret whenever possible. We provide a positive answer via a new algorithm called CRush Optimism with Pessimism (CROP). Our analysis shows that CROP achieves a constant-factor asymptotic optimality and, thanks to the forcedexploration-free design, adapts to bounded regret, and its regret bound scales not with the number of arms but with an effective number of arms that we introduce. We also show that CROP can be exponentially better than existing algorithms in the (nonasymptotic) regimes. Finally, we observe that even a clairvoyant oracle who plays according to the asymptotically optimal arm pull scheme may suffer a linear worst-case regret, indicating that it may not be the end of optimism. We believe our work may inspire a new family of algorithms for bandits and reinforcement learning.

Kwang-Sung Jun, Chicheng Zhang

Abstract 7: Short Talk 2 - Adaptive Discretization for Model-Based Reinforcement Learning in Theoretical Foundations of Reinforcement Learning, Sinclair 11:35 AM

We introduce the technique of adaptive discretization to design efficient model-based episodic reinforcement learning algorithms in large (potentially continuous) state-action spaces. Our algorithm is based on optimistic one-step value iteration extended to maintain an adaptive discretization of the space. From a theoretical perspective, we provide worst-case regret bounds for our algorithm, which are competitive compared to the state-of-the-art RL algorithms; moreover, our bounds are obtained via a modular proof technique, which can potentially extend to incorporate additional structure on the problem. Our algorithm has much lower storage and computational requirements, due to maintaining a more efficient partition of the state and action spaces. We illustrate this via experiments on several canonical control problems, which shows that our algorithm empirically performs significantly better than fixed discretization in terms of both faster convergence and lower memory usage.

Sean R. Sinclair, Tianyu Wang, Gauri Jain, Sid Banerjee, Christina Yu

Abstract 8: Short Talk 3 - A Kernel-Based Approach to Non-Stationary Reinforcement Learning in Metric Spaces in Theoretical Foundations of Reinforcement Learning, Darwiche Domingues 11:50 AM

In this work, we propose KeRNS: an algorithm for episodic reinforcement learning in non-stationary Markov Decision Processes (MDPs) whose state-action set is endowed with a metric. Using a non-parametric model of the MDP built with time-dependent kernels, we prove a regret bound that scales with the covering dimension of the state-action space and the total variation of the MDP with time, which quantifies its level of non-stationarity. Our method generalizes previous approaches based on sliding windows and exponential discounting used to handle changing environments.

Omar Darwiche Domingues, Pierre MENARD, Matteo Pirotta, Emilie Kaufmann, Michal Valko

Abstract 9: Short Talk 4 - Adaptive Regret for Online Control in Theoretical Foundations of Reinforcement Learning, Minasyan 12:05 PM

We consider regret minimization for online control with time-varying linear dynamical systems. The metric of performance we study is adaptive policy regret, or regret compared to the best policy on (it any interval in time). We give an efficient algorithm that attains first-order adaptive regret guarantees for the setting of online convex optimization with memory, subsequently used to derive a controller with such guarantees. We show that these bounds are nearly tight and validate these theoretical findings experimentally on simulations of time-varying dynamics and disturbances.

Paula Gradu, Eldad Hazan, Edgar Minasyan

Abstract 10: Short Talk 5 - Near-Optimal Reinforcement Learning with Self-Play in Theoretical Foundations of Reinforcement Learning, Yu 12:20 PM

This paper considers the problem of designing optimal algorithms for reinforcement learning in two-player zero-sum games. We focus on self-play algorithms which learn the optimal policy by playing against itself without any direct supervision. In a tabular episodic Markov game with states, max-player actions and min-player actions, the best existing algorithm for finding an approximate Nash equilibrium requires steps of game playing, when only highlighting the dependency on . In contrast, the best existing lower bound scales as and has a significant gap from the upper bound. This paper closes this gap for the first time: we propose an optimistic variant of the (Nash Q-learning) algorithm with sample complexity , and a new (Nash V-learning) algorithm with sample complexity . The latter result matches the information-theoretic lower bound in all problem-dependent parameters except for a polynomial factor of the length of each episode. We complement our upper bounds with a computational hardness result for achieving sublinear regret when playing against adversarial opponents in Markov games.

Yu Bai, Chi Jin, Tiancheng Yu
Abstract 11: Short Talk 6 - Preference learning along multiple criteria: A game-theoretic perspective in Theoretical Foundations of Reinforcement Learning, Bhatia 12:35 PM

The literature on ranking from ordinal data is vast, and there are several ways to aggregate overall preferences from pairwise comparisons between objects. In particular, it is well-known that any Nash equilibrium of the zero-sum game induced by the preference matrix defines a natural solution concept (winning distribution over objects) known as a von Neumann winner. Many real-world problems, however, are inevitably multi-criteria, with different pairwise preferences governing the different criteria. In this work, we generalize the notion of a von Neumann winner to the multi-criteria setting by taking inspiration from Blackwell’s approachability. Our framework allows for non-linear aggregation of preferences across criteria, and generalizes the linearization-based approach from multi-objective optimization.

From a theoretical standpoint, we show that the Blackwell winner of a multi-criteria problem instance can be computed as the solution to a convex optimization problem. Furthermore, given random samples of pairwise comparisons, we show that a simple, "plug-in" estimator achieves (near-)optimal minimax sample complexity. Finally, we showcase the practical utility of our framework in a user study on autonomous driving, where we find that the Blackwell winner outperforms the von Neumann winner for the overall preferences.

Kush Bhatia, Ashwin Pananjady, Peter Bartlett, Anca Dragan, Martin Wainwright

Abstract 13: Representation learning and exploration in reinforcement learning - Akshay Krishnamurthy in Theoretical Foundations of Reinforcement Learning, Krishnamurthy 02:20 PM

I will discuss new provably efficient algorithms for reinforcement in rich observation environments with arbitrarily large state spaces. Both algorithms operate by learning succinct representations of the environment, which they use in an exploration module to acquire new information. The first algorithm, called Homer, operates in a block MDP model and uses a contrastive learning objective to learn the representation. On the other hand, the second algorithm, called FLAMBE, operates in a much richer class of low rank MDPs and is model-based. Both algorithms accommodate nonlinear function approximation and enjoy provable sample and computational efficiency guarantees.

Abstract 14: Learning to price under the Bass model for dynamic demand - Shipra Agrawal in Theoretical Foundations of Reinforcement Learning, Agrawal 03:10 PM

We consider a novel formulation of the dynamic pricing and demand learning problem, where the evolution of demand in response to posted prices is governed by a stochastic variant of the popular Bass model with parameters that are linked to the so-called “innovation” and “imitation” effects. Unlike the more commonly used i.i.d. demand models, in this model the price posted not only affects the demand and the revenue in the current round but also the evolution of demand, and hence the fraction of market potential that can be captured, in future rounds. Finding a revenue-maximizing dynamic pricing policy in this model is non-trivial even when model parameters are known, and requires solving for optimal non-stationary policy of a continuous-time, continuous-state MDP. In this paper, we consider a more challenging problem where dynamic pricing is used in conjunction with learning the model parameters, with the objective of optimizing the cumulative revenues over a given selling horizon. Our main contribution is an algorithm with a regret guarantee of $O(m^{2/3})$, where $m$ is mnemonic for the (known) market size, along with a matching lower bound.

Abstract 15: Efficient Planning in Large MDPs with Weak Linear Function Approximation - Csaba Szepesvari in Theoretical Foundations of Reinforcement Learning, Szepesvari 04:00 PM

Large-scale Markov decision processes (MDPs) require planning algorithms with runtime independent of the number of states of the MDP. We consider the planning problem in MDPs using linear value function approximation with only weak requirements: low approximation error for the optimal value function, and a small set of "core" states whose features span those of other states. In particular, we make no assumptions about the representability of policies or value functions of non-optimal policies. Our algorithm produces almost-optimal actions for any state using a generative oracle (simulator) for the MDP, while its computation time scales polynomially with the number of features, core states, and actions and the effective horizon. I will discuss how this is achieved, some selected part of the vast related literature and what remains open.

Joint work with Roshan Shariff

ML Interpretability for Scientific Discovery

Subhashini Venugopalan, Michael Brenner, Scott Linderman, Been Kim

Fri Jul 17, 06:50 AM

ML has shown great promise in modeling and predicting complex phenomenon in many scientific disciplines such as predicting cardiovascular risk factors from retinal images, understanding how electrons behave at the atomic level [3], identifying patterns of weather and climate phenomena, etc. Further, models are able to learn directly (and better) from raw data as opposed to human selected features. The ability to interpret the model and find significant predictors could provide new scientific insights. Traditionally, the scientific discovery process has been based on careful observations of natural phenomenon, followed by systematic human analysis (of hypothesis generation and ex-perimental validation). ML interpretability has the potential to bring a radically different yet principled approach. While general interpretability relies on human parsing (common sense), scientific domains have semi-structured and highly structured bases for interpretation. Thus, despite differences in data modalities and domains, be it brain sciences, the behavioral sciences, or material sciences, there is a need for a common set of tools that address a similar flavor of problem, one of interpretability or fitting models to a known structure. This workshop aims to bring together members from the ML and physical sciences communities to introduce exciting problems to the broader community, and stimulate the production of new approaches towards solving open scientific problems.
07:00 AM Invited Talk - Eun-Ah Kim  
Kim, Venugopalan

07:50 AM Invited Talk - Barbara Engelhardt  
Engelhardt, Venugopalan

08:00 AM Q&A - Barbara Engelhardt  
Engelhardt, Venugopalan

08:40 AM Invited Talk - Sendhil Mullainathan, Machine Learning for Scientific Discovery (recorded)  
Mullainathan, Venugopalan

09:00 AM Q&A - Sendhil Mullainathan  
Mullainathan, Venugopalan

09:30 AM Invited Talk - Arun Narayanaswamy, ML Driven Scientific Discovery (recorded)  
Narayanaswamy, Venugopalan

10:10 AM Q&A - Arun Narayanaswamy  
Narayanaswamy, Venugopalan

10:50 AM Invited Talk - Katie Bouman  
Bouman, Venugopalan

11:40 AM Panel Discussion - setup  
Venugopalan

11:50 AM Panel Discussion  
Venugopalan, Kim, Engelhardt, Mullainathan, Narayanaswamy, Bouman

01:10 PM Poster Session (recorded)  
Venugopalan

01:35 PM Parallel poster session info, Closing remarks  
Venugopalan

01:45 PM [Session 1] P#13 Learning about learning by many-body systems

01:45 PM [Session 1] P#10 Explaining Chemical Toxicity Using Missing Features

01:45 PM [Session 1] P#11 Modeling Brain Microarchitecture with Deep Representation Learning

01:45 PM [Session 1] P#14 DisCont: Self-Supervised Visual Attribute Disentanglement using Context Vectors

01:45 PM [Session 1] P#08 (Re)Discovering Protein Structure and Function Through Language Modeling

01:45 PM Parallel poster session

01:45 PM [Session 1] P#12 Feature Extraction on Synthetic Black Hole Images

01:45 PM [Session 1] P#06 Unpacking Chemical Reaction Prediction Models Using Integrated Gradients

01:45 PM [Session 1] P#07 Inverse Problems, Deep Learning, and Symmetry Breaking

01:45 PM [Session 1] P#05 Actionable Attribution Maps for Scientific Machine Learning

02:15 PM [Session 2] P#17 In-Distribution Interpretability for Challenging Modalities

02:15 PM [Session 2] P#29 Unsupervised Attention-Guided Atom-Mapping

02:15 PM [Session 2] P#18 Extracting Interpretable Physical Parameters from Spatiotemporal Systems using Unsupervised Learning

02:15 PM [Session 2] P#27 End to End learning for Phase Retrieval

02:15 PM [Session 2] P#24 Interpreting Stellar Spectra with Unsupervised Domain Adaptation

02:15 PM [Session 2] P#16 Attribution Methods Reveal Flaws in Fingerprint-Based Virtual Screening

02:15 PM [Session 2] P#15 Look at the Loss: Towards Robust Detection of False Positive Feature Interactions Learned by Neural Networks on Genomic Data
Abstract 5: **Invited Talk - Sendhil Mullainathan, Machine Learning for Scientific Discovery (recorded)** in ML Interpretability for Scientific Discovery, Mullainathan, Venugopalan 08:40 AM

[Sendhil's talk is recorded and can be viewed here](https://slideslive.com/38930792/machine-learning-for-scientific-discovery?ref=search)

Abstract 7: **Invited Talk - Arun Narayanaswamy, ML Driven Scientific Discovery (recorded)** in ML Interpretability for Scientific Discovery, Narayanaswamy, Venugopalan 09:30 AM

[Arun's recorded talk can be viewed here](https://slideslive.com/38930788/ml-driven-scientific-discovery?ref=search)

Abstract 9: **Break - Poster videos (recorded)** in ML Interpretability for Scientific Discovery, Venugopalan 10:20 AM

Break + Watch some poster presentation videos

[Watch poster videos here](https://slideslive.com/38930794/break?ref=search)

Abstract 13: **Poster Session (recorded)** in ML Interpretability for Scientific Discovery, Venugopalan 01:10 PM

[Watch poster videos here](https://slideslive.com/38930794/break?ref=search)

Accepted papers pdf and videos are also here: [https://sites.google.com/corp/view/mli4sd-icml2020/program](https://sites.google.com/corp/view/mli4sd-icml2020/program)

Abstract 14: **Parallel poster session info, Closing remarks** in ML Interpretability for Scientific Discovery, Venugopalan 01:35 PM

Closing remarks + Logistics for parallel poster sessions and break out.

Abstract 15: **[Session 1] P#13 Learning about learning by many-body systems** in ML Interpretability for Scientific Discovery, 01:45 PM

Learning about learning by many-body systems

Weishun Zhong (Massachusetts Institute of Technology)*; Jacob Gold (Massachusetts Institute of Technology); Sarah Marzen (Massachusetts Institute of Technology; Claremont Colleges); Jeremy L England (GlaxoSmithKline); Nicole Yunger Halpern (Harvard University; Massachusetts Institute of Technology)


Video: [https://www.youtube.be/2gyLEPnTrI1M](https://www.youtube.be/2gyLEPnTrI1M)

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 16: **[Session 1] P#10 Explaining Chemical Toxicity Using Missing Features** in ML Interpretability for Scientific Discovery, 01:45 PM

Explaining Chemical Toxicity Using Missing Features

Kar Wai Lim (IBM Singapore)*; Bhanushee Sharma (Rensselaer Polytechnic Institute); Vijil Chentharamarakan (IBM AI Research); Payel Das (IBM Research); Jonathan S. Dordick (Rensselaer Polytechnic Institute)


Abstract 17: **[Session 1] P#11 Modeling Brain Microarchitecture with Deep Representation Learning** in ML Interpretability for Scientific Discovery, 01:45 PM

Modeling Brain Microarchitecture with Deep Representation Learning

Aishwarya H. Balwani (Georgia Institute of Technology)*; Eva Dyer (Georgia Tech)

Paper: [https://aishwaryaab.github.io/docs/papers/Balwani\_ICML\_Interpretability\_Workshop.pdf](https://aishwaryaab.github.io/docs/papers/Balwani\_ICML\_Interpretability\_Workshop.pdf)

Video: [https://www.dropbox.com/s/ra3o19eu4xdyywj/Modeling%20Brain%20Microarchitecture.mp4?dl=0](https://www.dropbox.com/s/ra3o19eu4xdyywj/Modeling%20Brain%20Microarchitecture.mp4?dl=0)

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 18: **[Session 1] P#14 DisCont: Self-Supervised Visual Attribute Disentanglement using Context Vectors** in ML Interpretability for Scientific Discovery, 01:45 PM

DisCont: Self-Supervised Visual Attribute Disentanglement using Context Vectors

Sarthak Bhagat (IIIT-Delhi)*; Vishaal Udandarao (IIIT Delhi); Shagun Uppal (IIIT-Delhi)


Video: [https://youtu.be/2gyLEPnTrI1M](https://youtu.be/2gyLEPnTrI1M)

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 19: **[Session 1] P#08 (Re)Discovering Protein Structure and Function Through Language Modeling** in ML Interpretability for Scientific Discovery, 01:45 PM

(Re)Discovering Protein Structure and Function Through Language Modeling

Jesse Vig (Salesforce)*; Ali Madani (Salesforce Research); Lav Varshney (UIUC: ECE); Nazneen Fatema Rajani (Salesforce Research)


Video: [https://vimeo.com/434882244](https://vimeo.com/434882244)

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

---

**Networking Town**

Abstracts (25):
Abstract 20: Parallel poster session in ML Interpretability for Scientific Discovery, 01:45 PM

Poster and Networking session

How does the poster session work:
Each poster has a zoom room where one of the authors will be present to answer questions. Click on the poster to see the zoom link information in the abstract.

Use the gather.town link below to network, and take your poster sessions outside of zoom after your session ends.
https://gather.town/wMniMF64Uloi89ND/mli4sd-icml2020-town

Abstract 21: [Session 1] P#12 Feature Extraction on Synthetic Black Hole Images in ML Interpretability for Scientific Discovery, 01:45 PM

Feature Extraction on Synthetic Black Hole Images

Joshua Yao-Yu Lin (Physics department, University of Illinois at Urbana-Champaign)*; George Wong (University of Illinois at Urbana-Champaign); Ben Prather (University of Illinois at Urbana-Champaign); Charles Gammie


Video: https://www.youtube.com/watch?v=D4QNmjXtp3Q

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 22: [Session 1] P#06 Unpacking Chemical Reaction Prediction Models Using Integrated Gradients in ML Interpretability for Scientific Discovery, 01:45 PM

Unpacking Chemical Reaction Prediction Models Using Integrated Gradients

William McCorkindale (University of Cambridge); David P Kovacs (University of Cambridge)*; Alpha Lee (University of Cambridge)


Video: https://drive.google.com/file/d/1j011oldfLa_Hry1ws-ZasdA5Y0A-STt/view

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 23: [Session 1] P#07 Inverse Problems, Deep Learning, and Symmetry Breaking in ML Interpretability for Scientific Discovery, 01:45 PM

Inverse Problems, Deep Learning, and Symmetry Breaking

Kshitij Tayal (University of Minnesota)*; Chieh-Hsin Lai (University of Minnesota, Twin Cities); Raunak Manekar (University of Minnesota); Vipin Kumar (University of Minnesota); Ju Sun (University of Minnesota)


Video: https://drive.google.com/drive/folders/1RT7xQ6QAuvUhrb7E0xytkvOuNXj1Zdth_?usp=sharing

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 24: [Session 1] P#05 Actionable Attribution Maps for Scientific Machine Learning in ML Interpretability for Scientific Discovery, 01:45 PM

Actionable Attribution Maps for Scientific Machine Learning

Shusen Liu (Lawrence Livermore National Laboratory)*; Bhavya Kaikthur (Lawrence Livermore National Laboratory); Jize Zhang (Lawrence Livermore National Laboratory); Anna Hiszpanski (Lawrence Livermore National Laboratory); Emily Robertson (Lawrence Livermore National Laboratory); Donald Loveland (Lawrence Livermore National Laboratory); T. Yong-Jin Han (LLNL)


Video: https://youtu.be/IMfWZ82Dcow

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 25: [Session 2] P#28 Learning Cell State Representations From Barcoded Gene-Expression Trajectories in ML Interpretability for Scientific Discovery, 02:15 PM

Learning Cell State Representations From Barcoded Gene-Expression Trajectories

Yu Wu (Princeton University)*; Le Cong (Stanford University); Mengdi Wang (Princeton University/DeepMind)


Video: https://youtu.be/4tjD6_7Eclg

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 26: [Session 2] P#18 Extracting Interpretable Physical Parameters from Spatiotemporal Systems using Unsupervised Learning in ML Interpretability for Scientific Discovery, 02:15 PM

Extracting Interpretable Physical Parameters from Spatiotemporal Systems using Unsupervised Learning

Peter Y Lu (MIT)*; Samuel Kim (MIT); Marin Soljacic (MIT)


Video: https://youtu.be/oWgj flaggedX.png

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 27: [Session 2] P#24 Interpreting Stellar Spectra with Unsupervised Domain Adaptation in ML Interpretability for Scientific Discovery, 02:15 PM

Interpreting Stellar Spectra with Unsupervised Domain Adaptation

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.
Abstract 28: [Session 2] P#29 Unsupervised Attention-Guided Atom-Mapping for Scientific Discovery, 02:15 PM

Unsupervised Attention-Guided Atom-Mapping

Philippe PS Schwaller (IBM Research Europe / University of Bern)\(^*\); Benjamin Hoover (IBM Research); Jean-Louis Reymond (University of Bern); Hendrik Strobelt (IBM Research); Teodoro Laino (IBM Research Europe)


Video: https://vimeo.com/434757113

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 29: [Session 2] P#17 In-Distribution Interpretability for Challenging Modalities in ML Interpretability for Scientific Discovery, 02:15 PM

In-Distribution Interpretability for Challenging Modalities

Cosmas Heiss (TU Berlin)\(^*\); Ron Levie (TU Berlin); Cinjon Resnick (NYU); Gitta Kutyniok (Technische Universität Berlin); Joan Bruna (Courant Institute of Mathematical Sciences, NYU, USA)


Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 30: [Session 2] P#20 Deep Interpretability for GWAS in ML Interpretability for Scientific Discovery, 02:15 PM

Deep Interpretability for GWAS

Deepak Sharma (MILA, McGill University)\(^*\); Audrey Durand (Université Laval); Marc-Andre Legault (Université de Montréal); Louis-Philippe Lemieux (Montreal Heart Institute)


Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 31: [Session 2] P#27 End to End learning for Phase Retrieval in ML Interpretability for Scientific Discovery, 02:15 PM

End to End learning for Phase Retrieval

Raunak Manekar (University of Minnesota); Kshitij Tayal (University of Minnesota)\(^*\); Vipin Kumar (University of Minnesota); Ju Sun (University of Minnesota)


Video: https://drive.google.com/drive/folders/10To09kLk0PP-M6kCqe5T4sJxzlnHuFz1?usp=sharing

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 32: [Session 2] P#16 Attribution Methods Reveal Flaws in Fingerprint-Based Virtual Screening in ML Interpretability for Scientific Discovery, 02:15 PM

Attribution Methods Reveal Flaws in Fingerprint-Based Virtual Screening

Vikram Sundar (Google); Lucy Colwell (Google)\(^*\)


Video: https://www.youtube.com/watch?v=2Vv_e53DuFs

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 33: [Session 2] P#15 Look at the Loss: Towards Robust Detection of False Positive Feature Interactions Learned by Neural Networks on Genomic Data in ML Interpretability for Scientific Discovery, 02:15 PM

Look at the Loss: Towards Robust Detection of False Positive Feature Interactions Learned by Neural Networks on Genomic Data

Mara R Finkelstein (Stanford University); Avanti Shrikumar (Stanford University)\(^*\); Anshul Kundaje (Stanford University)

Paper: https://github.com/kundajelab/feature_interactions/blob/master/Look\_At\_The\_Loss\_MLiSD\_ICML\_Workshop.pdf

Video: https://drive.google.com/drive/folders/127QlmcQZpfCNWmhtOuA79yRCpoqOtQnWV?usp=sharing

Please join poster Q&A zoom via link below. Please DO NOT SHARE zoom link externally.

Abstract 34: Networking Town in ML Interpretability for Scientific Discovery, N/A

Gather town:

https://gather.town/wMniMF64Uloi89ND/ml4sd-icml2020-town

Uncertainty and Robustness in Deep Learning Workshop (UDL)
There has been growing interest in rectifying deep neural network instabilities. Challenges arise when models receive samples drawn from outside the training distribution. For example, a neural network tasked with classifying handwritten digits may assign high confidence predictions to cat images. Anomalies are frequently encountered when deploying ML models in the real world. Well-calibrated predictive uncertainty estimates are indispensable for many machine learning applications, such as self-driving vehicles and medical diagnosis systems. Generalization to unseen and worst-case inputs is also essential for robustness to distributional shift. In order to have ML models reliably predict in open environment, we must deepen technical understanding in the emerging areas of: (1) learning algorithms that can detect changes in data distribution (e.g. out-of-distribution examples); (2) mechanisms to estimate and calibrate confidence produced by neural networks in typical and unforeseen scenarios; (3) methods to improve out-of-distribution generalization, including generalization to temporal, geographical, hardware, adversarial, and image-quality changes; (4) benchmark datasets and protocols for evaluating model performance under distribution shift; and (5) key applications of robust and uncertainty-aware deep learning (e.g., computer vision, robotics, self-driving vehicles, medical imaging) as well as broader machine learning tasks.

This workshop will bring together researchers and practitioners from the machine learning communities, and highlight recent work that contributes to addressing these challenges. Our agenda will feature contributed papers with invited speakers. Through the workshop we hope to help identify fundamentally important directions on robust and reliable deep learning, and foster future collaborations.

Schedule

07:30 AM Opening Remarks Li
07:40 AM Keynote #1 Matthias Hein Hein
08:10 AM Spotlight Talk 1: Likelihood Regret: An Out-of-Distribution Detection Score For Variational Auto-encoder Xiao
08:15 AM Spotlight Talk 2: A Closer Look at Accuracy vs. Robustness Yang
08:20 AM Spotlight Talk 3: Depth Uncertainty in Neural Networks Antorájn, Allingham
08:25 AM Spotlight Talk 4: Few-shot Out-of-Distribution Detection Wang
08:30 AM Spotlight Talk 5: Detecting Failure Modes in Image Reconstructions with Interval Neural Network Uncertainty Oala
08:35 AM Spotlight Talk 6: On using Focal Loss for Neural Network Calibration Mukhoti
08:40 AM Spotlight Talk 7: AutoAttack: reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks Croce
08:45 AM Spotlight Talk 8: Calibrated Top-1 Uncertainty estimates for classification by score based models Oberman
09:00 AM Poster Session (click to see links)
10:00 AM Coffee Break
10:30 AM Keynote #2 Finale Doshi-Velez
11:00 AM Keynote #3 Percy Liang Liang
11:30 AM Panel Discussion
12:30 PM Lunch Break
01:30 PM Keynote #4 Raquel Urtasun Urtasun
02:00 PM Contributed Talk 1: Confidence-Calibrated Adversarial Training: Generalizing to Unseen Attacks Stutz
02:10 PM Contributed Talk 2: Improving robustness against common corruptions by covariate shift adaptation Schneider
02:20 PM Contributed Talk 3: A Unified View of Label Shift Estimation Garg
02:30 PM Keynote #5 Justin Gilmer Gilmer
03:00 PM Coffee Break
03:30 PM Contributed Talk 4: A Benchmark of Medical Out of Distribution Detection Cohen
03:40 PM Contributed Talk 5: Neural Ensemble Search for Performant and Calibrated Predictions Zaidi

https://slideslive.com/38930947

Abstract 7: Spotlight Talk 5: Detecting Failure Modes in Image Reconstructions with Interval Neural Network Uncertainty in Uncertainty and Robustness in Deep Learning Workshop (UDL), Oala 08:30 AM

https://slideslive.com/38930948

Abstract 8: Spotlight Talk 6: On using Focal Loss for Neural Network Calibration in Uncertainty and Robustness in Deep Learning Workshop (UDL), Mukhoti 08:35 AM

https://slideslive.com/38930949

Abstract 9: Spotlight Talk 7: AutoAttack: reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks in Uncertainty and Robustness in Deep Learning Workshop (UDL), Croce 08:40 AM

https://slideslive.com/38930950

Abstract 10: Spotlight Talk 8: Calibrated Top-1 Uncertainty estimates for classification by score based models in Uncertainty and Robustness in Deep Learning Workshop (UDL), Oberman 08:45 AM

https://slideslive.com/38930951

Abstract 11: Poster Session (click to see links) in Uncertainty and Robustness in Deep Learning Workshop (UDL), 09:00 AM

1 Confidence-Calibrated Adversarial Training: Generalizing to Unseen Attacks David Stutz, Matthias Hein and Bernt Schiele
https://zoom.us/j/99678534052

2 Improving robustness against common corruptions by covariate shift adaptation Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel and Matthias Bethge
https://meet.google.com/qjt-atoh-wup

3 A Unified View of Label Shift Estimation Saurabh Garg, Yifan Wu, Sivaraman Balakrishnan and Zachary Lipton
https://cmu.zoom.us/j/94831507038?pwd=RVVyVy96YXNiJUJ5dVcvQT09

4 A Benchmark of Medical Out of Distribution Detection Tianshi Cao, David Yu-Tung Hui, Chin-Wei Huang and Joseph Paul Cohen
https://meet.google.com/jkx-paul-gao

5 Neural Ensemble Search for Performant and Calibrated Predictions Sheheryar Zaidi, Arber Zela, Thomas Elsken, Frank Hutter and Yee Whye Teh https://meet.google.com/uxs-uuxx-uwo

6 Bayesian model averaging is suboptimal for generalization under model misspecification Andres Masegosa
https://meet.google.com/mpf-chva-pgy

7 Likelihood Regret: An Out-of-Distribution Detection Score For Variational Auto-encoder Zhisheng Xiao, Qing Yan and Yali Amit
https://uchicago.zoom.us/j/98748547880?pwd=Q3Y0dUVPUFhbG4NmNJ2hwZndkWn09

Password: 370663

8 A Closer Look at Accuracy vs. Robustness Yao-Yuan Yang, Cyrus Rashtchian, Hongyang Zhang, Ruslan Salakhudinov and Kamalika Chaudhuri
https://ucsd.zoom.us/j/92778085557?pwd=Q3Y0dUVPUFhbG4NmNJ2hwZndkWn09

9 Depth Uncertainty in Neural Networks Javier Antorán, James Urquhart Allingham and JosÃ© Miguel HernÃ¡ndez-Lobato
https://us02web.zoom.us/j/5419103161?pwd=eXdqWURLc3o4SktwQWZOU2pZNhl6QT09

10 Few-shot Out-of-Distribution Detection Kuan-Chieh Wang, Paul Virol, Eleni Triantafillou and Richard Zemel
https://vectorinstitute.zoom.us/j/88251013741?pwd=ZnNlOTk5b21OdEl2aS9zaE1TSHVpMF09

11 Detecting Failure Modes in Image Reconstructions with Interval Neural Network Uncertainty Luis Oala, Cosmas HeiÁE, Jan Macdonald, Maximilian MÃÂrrz, Gitta Kutyniok and Wojiectch Samek
https://us02web.zoom.us/j/88251013741?pwd=TzIOUXeExzUVdSUDE5RXFkTytWZz09

18 Redundant features can hurt robustness to distribution shift Guillermo Ortiz-Jimenez, Apostolos Modas, Seyed-Mohsen Moosavi-Dezfooli and Pascal Frossard  
https://epfl.zoom.us/j/9895605194?pwd=aG5yaTBFbU5rQkxjQVdSa3pVRV9hUT09  
548657

19 Scalable Training with Information Bottleneck Objectives Andreas Kirsch, Clare Lyle and Yarin Gal  
https://meet.google.com/mzn-pjuh-kia  

17 Measuring Robustness to Natural Distribution Shifts in Image Classification Rohan Taori, Achai Dave, Vaishaal Shankar, Nicholas Carlini, Benjamin Recht and Ludwig Schmidt  
https://berkeley.zoom.us/j/9751483859?pwd=SXdiOSthMzNU01QYUSVEVFTTM2aV09  

https://berkeley.zoom.us/j/4219480859?pwd=QkNIYmNCZjJobEUXq2Q5TzR4Om1QUT09  

15 Bayesian Deep Ensembles via the Neural Tangent Kernel Bobby He, Balaji Lakshminarayanan and Yee Whye Teh  
https://meet.google.com/exb-jkju-vbr  

14 Calibrated Top-1 Uncertainty estimates for classification by score based models Adam Oberman, Chris Finlay, Alexander Lannuantouo and Tiago Salvador  
https://mogill.zoom.us/j/91308261627  

13 AutoAttack: reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks Francesco Croce and Matthias Hein  
https://zoom.us/j/98193746046?pwd=RWNYVEJMqFzGRlUDgr1c1NmtVdz09  

12 Learning Representations with Score Invariant Learning Daksh Idnani and Jonathan Kao  
https://ucla.zoom.us/j/98883773408?pwd=Y3M2WmxNTW1HOS9hMIBZYmJSH20bOGJqbec  

11 Understanding and Improving Fast Adversarial Training Maksym Papailiopoulos, Isabela Albuquerque, Joao Monteiro, Mohammad Darvishi, Tiago Falk and Ioannis Mitliagkas  
https://zoom.us/j/98309028439?pwd=vdTVFUMdZllIZFVEIIdDbkdCL2QvZz09  

10 On the relationship between class selectivity, dimensionality, and robustness Matthew Leavitt and Ari Morcos  
https://utoronto.zoom.us/j/6995996022?pwd=UUIxNUNnQThLVMhLV0h5RDVJRG92dz09  

https://berkeley.zoom.us/j/4219480859?pwd=QkNIYmNCZjJobEUXq2Q5TzR4Om1QUT09  

8 By-CUy  

7 Improving out-of-distribution generalization via multi-task self-supervised pretraining Isabela Albuquerque, Nikhil Naik, Junnan Li, Shrish Keskar and Richard Socher  
https://zoom.us/j/9639222189?pwd=bVg5am4vb0FXVZMNETiPWER4SURyUT09  
6byCly

https://berkeley.zoom.us/j/4219480859?pwd=QkNIYmNCZjJobEUXq2Q5TzR4Om1QUT09  

548657

5 Redundant features can hurt robustness to distribution shift Guillermo Ortiz-Jimenez, Apostolos Modas, Seyed-Mohsen Moosavi-Dezfooli and Pascal Frossard  
https://epfl.zoom.us/j/9895605194?pwd=aG5yaTBFbU5rQkxjQVdSa3pVRV9hUT09  

4 Consistency Regularization for Certified Robustness of Smoothed Classifiers Jongheon Jeong and Jinwoo Shin  
https://zoom.us/j/9802066306?pwd=MGJyRzR4Q2FXSWM4c3c3NURrR3NiUT09  

https://berkeley.zoom.us/j/4219480859?pwd=QkNIYmNCZjJobEUXq2Q5TzR4Om1QUT09  

2 Learning Representations with Score Invariant Learning Daksh Idnani and Jonathan Kao  
https://ucla.zoom.us/j/98883773408?pwd=Y3M2WmxNTW1HOS9hMIBZYmJSH20bOGJqbec  

1 AutoAttack: reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks Francesco Croce and Matthias Hein  
https://zoom.us/j/98193746046?pwd=RWNYVEJMqFzGRlUDgr1c1NmtVdz09

40 Revisiting One-vs-All Classifiers for Predictive Uncertainty and Out-of-Distribution Detection in Neural Networks Shreyas Padhy, Zachary Nado, Jie Ren, Jeremiah Liu, Jasper Snoek and Balaji Lakshminarayanan
https://meet.google.com/cym-bnc-mbb

41 Tilted Empirical Risk Minimization Tian Li, Ahmad Beirami, Maziar Sanjabi and Virginia Smith
https://cmu.zoom.us/j/96622035970?pwd=M211WMTIIV0tuWkpmUAIHMDiBmpFdz09

42 Towards Robust Classification with Deep Generative Forests Alvaro Henrique Chaim Correia, Robert Peharz and Cassio de Campos
https://zoom.us/j/98359572057?pwd=VDEzUHE5Q1JkVFBoK1RQUTJTUTU0TG09

43 Riemannian Continuous Normalizing Flows Emile Mathieu
https://meet.google.com/vts-axeh-doo

44 Improving Calibration of BatchEnsemble with Data Augmentation Yeming Wen, Ghassen Jerfel, Rafael Muller, Michael Dusenbery, Jasper Snoek, Balaji Lakshminarayanan and Dustin Tran
https://us04web.zoom.us/j/7911629966?pwd=eVNEZ2NRVGY4NExU1Kd1Yk1lTzB1dE09

45 Environment Inference for Invariant Learning Elliot Creager, JÁEf-henrik Jacobsen and Richard Zemel
https://vectorinstitute.zoom.us/j/9861353513?pwd=MK5xa2VzLzBV0JzdjZINXNVl나LLR09
470387

46 Nonlinear Gradient Estimation for Query Efficient Blackbox Attack Huichen Li, Linyi Li, Xiaojun Xu, Xiaolu Zhang, Shuang Yang and Bo Li
https://illinois.zoom.us/j/92837610121?pwd=Wi45QXRvaeNIWfXNOjJdR3UmcG1UQm09

47 ImageNet performance correlates with pose estimation robustness and generalization on out-of-domain data Alexander Mathis, Thomas Blasi, Mert YÄ¼ksegÅ¶nÅ¼, Byron Rogers, Matthias Bethge and Mackenzie Mathis
https://harvard.zoom.us/j/96654497889?pwd=Z2NrL05LQlozZkdSMjZ3TTJUQzUxQT09

48 CRUDE: Calibrating Regression Uncertainty Distributions Empirically Eric Zelikman, Christopher Healy, Sharon Zhou and Anand Avati
https://harvard.zoom.us/j/98541353638?pwd=QJuoTjVkJYm4mR1NSamw1VUJ3eWVJZz09

49 Positive-Unlabeled Learning with Arbitrarily Non-Representative Labeled Data Zayd Hammoudeh and Daniel Lowd
https://uoregon.zoom.us/j/99760928750?pwd=dmVCULI3WUlxNldDRGV5dmp5FjFCcWvyUTQ0
359623

50 Probabilistic Robustness Estimates for Deep Neural Networks Nicolas Couellan
https://zoom.us/j/92492909798?pwd=NUo4QzIPUMJlVnZCTW9HOGdpZ3lidz09

51 Estimating Risk and Uncertainty in Deep Reinforcement Learning William Clements, Bastien Van Delft, BenoÃ®t-Marie Robaglia, Reda Bahi Staoui and SÃ©bastien Toth
https://us02web.zoom.us/j/84543888187?pwd=cvDlSI2ZQUVAzTNkrRlRT2ITVxNCZz09

52 An Empirical Study of Invariant Risk Minimization Yo Joong Choe, Jiyeon Ham and Kyubyong Park
https://cmu.zoom.us/j/9718408310?pwd=MGJ0bEYzQXg3RVH5SksxTTWY2RnBHez09

53 Information-Bottleneck under Mean Field Initialization Vinayak Abrol and Jared Tanner
https://zoom.us/j/99737001164?pwd=ZnNLDX14SkFTmlkdvVYzSzVkc5Zz09

https://berkeley.zoom.us/j/96564472302?pwd=TXIPWHFzEHtVzTROYnB4KzJSVJhLU

55 Evaluating Prediction-Time Batch Normalization for Robustness under Covariate Shift Zachary Nado, Shreyas Padhy, D. Sculley, Alexander Dâ€™Amore, Balaji Lakshminarayanan and Jasper Snoek
https://meet.google.com/zxr-qijn-xkk

56 Failures of Variational Autoencoders and their Effects on Downstream Tasks Yaniv Yacoby, Weiwei Pan and Finale Doshi-Velez
https://harvard.zoom.us/j/98541353638?pwd=QJuoTjVkJYm4mR1NSamw1VUJ3eWVJZz09

57 Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness Jeremiah Liu, Zi Lin, Shreyas Padhy, Bahi Slaoui and Sébastien Toth
https://zoom.us/j/996562904311?pwd=QUx1bzdJL0dCcG5QcDdB0211OOGppZz09

58 On the Role of Dataset Quality and Heterogeneity in Model Confidence Yuan Zhao, Jiasi Chen and Samet Oymak
https://us04web.zoom.us/j/7637416589?pwd=L09lNIBDUXcvNHZ6MJV3TTJUQzUxQT09

59 On the Dataset of Dataset Quality and Heterogeneity in Model Confidence Yuan Zhao, Jiasi Chen and Samet Oymak
https://us04web.zoom.us/j/7637416589?pwd=L09lNIBDUXcvNHZ6MJV3TTJUQzUxQT09

60 Certified Adversarial Robustness via Randomized Smoothing: a Case Study for Laplace Noises Jiayi Teng, Guanghee Lee and Yang Yuan
https://zoom.us/j/94122377310?pwd=SGZoRXgxQ1AU1dEYkVIZndnM3Y3UT09
12341234

61 QUEST for MEDIXS: Quasi-norm based Uncertainty ESTimation for MEDical Image SYNthesis Uddeshya Upadhyay, Viswanath P. Sudarshan and Suyash P. Aawate
https://stanford.zoom.us/j/92845303885?pwd=Kzd6d2tjMGdXL0FsdjzDlE5DTrzNJ09
519357

62 On uncertainty estimation in active learning for image segmentation Bo Li and Tommy Sonne AlstrÃœm
https://zoom.us/j/65926847472?pwd=NGU0cXlVcGxVQm5hM3IcN3FtRjWSZz09
519357

63 Bayesian Autoencoders: Analysing and Fixing the Bernoulli likelihood for Out-of-Distribution Detection Bang Xiang Yong, Tim Pearce and Alexandra Brintrup meet.google.com/uxr-hzin-coj

64 A Comparison of Bayesian Deep Learning for Out of Distribution Detection and Uncertainty Estimation John Mitros, Arjun Pakrashi and Brian Mac Namee
https://meet.google.com/puu-yysf-owt?hs=122&authuser=0

65 Practical Bayesian Neural Networks via Adaptive Subgradient Optimization Methods Samuel Kessler, Arnold Salas, Vincent Tan Weng and Jared Tanner
https://zoom.us/j/99737001164?pwd=ZnNLDX14SkFTmlkdvVYzSzVkc5Zz09

ICML 2020 Workshop book

66 Bayesian Few-Shot Classification with One-vs-Each Priors-Gamma
Augmented Gaussian Processes Jake Snell and Richard Zemel
https://vectorinstitute.zoom.us/j/98053290381?pwd=T2FaXx1aUhWcS9iSFFhTW9iY1ZtOGUT09

67 Characteristics of Monte Carlo Dropout in Wide Neural Networks
Joachim Sicking, Maram Akila, Tim Wirtz, Sebastian Houben and Asja Fischer
https://us02web.zoom.us/j/98053290381?pwd=T2FaXx1aUhWcS9iSFFhTW9iY1ZtOGUT09

68 On Power Laws in Deep Ensembles Ekaterina Lobacheva, Nadezhda Chirkova, Maxim Kodryan and Dmitry Vetroy
https://zoom.us/j/95153236006?pwd=bTJDM3M0NIbVelZXRJHay9XZHRQ0T9

69 Outlier Detection through Null Space Analysis of Neural Networks
Matthew Cook, Alina Zare and Paul Gader
https://ufl.zoom.us/j/9456341852?pwd=MXR0uVw5M3BUuMFrYcjiGTDBVdDhQ0T6

70 In a forward direction: Analyzing distribution shifts in machine translation test sets over time
Thomas Liao, Benjamin Recht and Ludwig Schmidt
https://berkeley.zoom.us/j/96864204218?pwd=el3ZRVVOQklkJDZGTU1PZTYoDJ2P0UT0

71 Characterizing Adversarial Transferability via Gradient Orthogonality and Smoothness
Zhudi Yang, Linyi Li, Xiaojun Xu, Kaizhao Liang, Shiliang Zuo, Qian Chen, Benjamin Rubinstein, Ce Zhang and Bo Li
https://illinois.zoom.us/j/95161238471?pwd=NHRrZ1c0ZlAwMVp6U25vNzkyRTd6UT09

72 Untapped Potential of Data Augmentation: A Domain Generalization Viewpoint
Vihari Piratla and Shiv Shankar
https://meet.google.com/xya-jmcm-htd

73 Classifying Perturbation Types for Adversarial Robustness Against Multiple Threat Models Pratyush Maini, Xinyun Chen, Bo Li and Dawn Song
https://berkeley.zoom.us/j/9635581753?pwd=ezFr0hMVmQlRU5mNzNsb0lXcHlXUT09

https://berkeley.zoom.us/j/92010569509?pwd=OEInbXJ5TdRza2FzSFSGZ2b09HTl6deld

75 You won't believe you can learn CIFAR-10 with this: Another take on Information Bottleneck Objectives
Andreas Kirsch, Clare Lyle and Yarin Gal
https://meet.google.com/xte-ywjb-paj

76 Principled Uncertainty Estimation for High Dimensional Data
JosÃ© Miguel HernÃ¡ndez-Lobato and Yarin Gal
https://meet.google.com/xte-ywjb-paj

77 Rethink Autoencoders: Robust Manifold Learning Tahuir Li, Rishabh Mehta, Zeqh Chian and Ju Sun
https://umn.zoom.us/j/92216511031?pwd=UjN1K2IvRFVvaDpjVUJXWFQtMURZUT09

78 Continuous-Depth Bayesian Neural Networks Winnie Xu, Ricky T.Q. Chen and David Duvenaud
https://utoronto.zoom.us/j/7553398989

79 Robust Temporal Point Event Localization through Smoothing and Counting
Julien Schroeter, Kirill Sidorov and David Marshall
https://cardiff.zoom.us/j/96449530903?pwd=RjBTNDFmY0ZIT1EyMFJRZjdNcDHoUT09

80 Chi-square Information for Invariant Learning Prasanna Sattigeri, Vikash Sehwag, Arjun Nitin Bhagoji and Prateek Mittal
https://mit.zoom.us/j/92724256949?pwd=K3JRdHhVtG1xV2tRbE81WXJQUGZsdz09

81 Robust Out-of-distribution Detection via Informative Outlier Mining
Jiefeng Chen, Sharon Li, Xi Yingli and Somesh Jha
https://us04web.zoom.us/j/78581014579?pwd=MFIeXRKZTBjdWw4eDNoeTJEcGZS7

82 An Empirical Analysis of the Impact of Data Augmentation on Distillation
Deepan Das, Haley Massa, Abhimanyu Kulkarni and Theodoros Rekatsinas
https://meet.google.com/yzd-nbjv-qyx

83 It is Likely That Your Loss Should be a Likelihood Mark Hamilton, Evan Shelhamer and William Freeman
https://berkeley.zoom.us/j/98053290381?pwd=T2FaXx1aUhWcS9iSFFhTW9iY1ZtOGUT09

84 Self-Adaptive Training: beyond Empirical Risk Minimization
Chao Zhang and Hongyang Zhang
https://meet.google.com/cek-gsff-ebv

85 BCaOUN: Bayesian Classifiers with Out-of-Distribution Uncertainty
Doshi-Velez and Daiwei Pan
https://ufl.zoom.us/j/94495842174?pwd=czMfNvaQrMrMzGRyvdjEczXjWUJwUT09

86 Bayesian active learning for production, a systematic study and a reusable library Parmida Atighehchian, FrÃ©do Brachand-Charron and Alexandre Lacoste
https://elementai.zoom.us/j/92111246172?pwd=Nv92V2MyJ1lmKzFaXWICa1dpNHZM7

87 Certainty as Supervision for Test-Time Adaptation Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen and Trevor Darrell
https://berkeley.zoom.us/j/92105695905?pwd=cEk0TOU5Y0RDb0ZFSGZ2b09HTl6deld

88 Robust Deep Reinforcement Learning through Adversarial Loss Tuomas Oikarinen, Tsui-Wei Weng and Luca Daniel
https://mit.zoom.us/j/92724256949?pwd=K3JRdHhVtG1xV2tRbE81WXJQUGZsdz09

89 Cold Posteriors and Aleatoric Uncertainty Ben Adlam, Sam Smith and Tuomas Oikarinen
https://meet.google.com/bdq-nckv-tyq

90 Ensemble Mean vs. Ensemble Variance: Which is a Better Uncertainty Metric for Incipient Disease Detection? Baihong Jin, Yingshui Tan, Xiangyu Yue, Yue Chen and Alberto Sangiovanni-Vincentelli
https://berkeley.zoom.us/j/95691253740?pwd=SDFlR2hXYkRBQWFJZ3cydXlwRnEyZz09

91 Simplicity Bias and the Robustness of Neural Networks Harsh Shah, Kaustav Tamuly, Aditi Raghunathan, Prateek Jain and Praneeth Netrapalli
https://meet.google.com/bdz-nkvy-tyq

92 Exact posterior distributions of wide Bayesian neural networks Jiri Hron, Yasaman Bahri, Roman Novak, Jeffrey Pennington and Jascha Sohl-Dickstein
https://meet.google.com/pvb-brkx-png

93 A Simulation-based Framework for Characterizing Predictive Distributions for Deep Learning Jessica Ai, Beliz Gokkaya, Ilknur Emamjomeh-Zadeh, Hannah Li, Li Chen, Neamah Hussein, Ousmane Dia and Seba Baloglu
https://us04web.zoom.us/j/72581014579?pwd=MFlJeXRKZTBjdWw4eDNoeTJEcGZS7

94 In a forward direction: Analyzing distribution shifts in machine translation test sets over time
https://berkeley.zoom.us/j/96864204218?pwd=el3ZRVVOQklkJDZGTU1PZTYoDJ2P0UT09

95 Outlier Detection through Null Space Analysis of Neural Networks
Matthew Cook, Alina Zare and Paul Gader
https://ufl.zoom.us/j/9456341852?pwd=MXR0uVw5M3BUuMFrYcjiGTDBVdDhQ0T6

96 Robust Temporal Point Event Localization through Smoothing and Counting
Julien Schroeter, Kirill Sidorov and David Marshall
https://cardiff.zoom.us/j/96449530903?pwd=RjBTNDFmY0ZIT1EyMFJRZjdNcDHoUT09
94 Structured Weight Priors for Convolutional Neural Networks
Tim Pearce, Andrew Y.K. Foong and Alexandra Brintrup
meet.google.com/rgz-vdp-xir

95 Learning Generative Models from Classifier Uncertainties Siddharth Narayanaswamy and Brooks Paige https://meet.google.com/gjz-innd-iyt

https://asu.zoom.us/j/95489346758

https://cmu.zoom.us/j/95091389695?pwd=K2RXbzZVSItzRVZadHB4eHZaZmIDZz09

98 Transferable Adversarial Examples for Atari 2600 Games Damian Stachura and MichaÅ‡. ZajÄ…c https://meet.google.com/jfl-nqyd-fly

99 Robustness to Distribution Shifts using Multiple Environments Andris Andreassen, Rebecca Roeolfs and Behnam Neyshabur
https://zoom.us/j/91459170926?pwd=M25Sa01jWjZJeWlMNG5OSlV5Zmd6QT09

100 Our Evaluation Metric Needs an Update to Encourage Generalization Swaroop Mishra, Anjana Arunkumar and Chris Bryan
https://asu.zoom.us/j/92127915576

101 Harder or Different? A Closer Look at Distribution Shift in Dataset Reproduction Shangyun Lu, Bradley Nott, Aaron Olson, Alberto Todeschini, Puya Vahabi, Yair Carmon and Ludwig Schmidt
https://stanford.zoom.us/j/98068202076?pwd=LytZdjg5bTR1sedVcFFHQVpSaHrzZ3O9

102 Empirical Scoring Rule Decomposition in Deep Learning Tony Duan
https://teams.microsoft.com/l/meetup-join/19%3ameeting_NT8bknxzE2OTY6DQwOS00Mj01Njy0Q3EyN2M1ZWNj%40thread.v2?context=%7b%22Tid%22%3a%2272f988bf-86f1-41af-91ab-2d7cd011db47%22%2c%22Oid%22%3a%2233328e52-be2d-485a-a0c5-7a8d97aaa4e2%22%7d

103 Failure Prediction by Confidence Estimation of Uncertainty-Aware Dirichlet Networks Theodoros Tsigkariadis
https://mit.zoom.us/j/95278652551?pwd=WHnWnesuUUvOod0FSaUVbkJaHNp5QT09

104 RayS: A Ray Searching Method for Hard-label Adversarial Attack Jinghui Chen and Quanquan Gu
https://ucla.zoom.us/j/92290062678?pwd=MThrUVVDsJnElU2FNYWFlOTJyNjJ1cFJSbGd0

105 Joint Energy-Based Models for Semi-Supervised Classification
Stephen Zhao, Joern-Henrik Jacobsen and Will Grathwohl
https://us02web.zoom.us/j/98327646584?pwd=ZkxBcDNRMU9jWU0OJD6SDhoYjK5z231335

106 Single Shot MC Dropout Approximation Kai Brach, Beate Sick and Oliver Duerr
https://us02web.zoom.us/j/89892013425?pwd=VDZQN2JLSy9hTVg3VpqoFhBdVJ2dz8

107 Reliable Uncertainties for Bayesian Neural Networks using Alpha-divergences Hector Javier Hortua, Luigi Malago and Riccardo Volpi
https://zoom.us/j/98397444333?pwd=b0JCTGhiQEFWThTDYd6akpDeXFzd098a3QWf

108 Evaluating Uncertainty Estimation Methods on 3D Semantic Segmentation of Point Clouds Swaroop Bhandary, Nico Hochgeschwender, Paul Pfäger and Matias Valdenegro-Toro
https://zoom.us/j/99779971867?pwd=Qm9aOzJXYrJvZGzwNmk4SU9BbFJQzd09

109 Improving predictions of Bayesian neural networks via local linearization Alexander Immer, Maciej Korzepa and Matthias Bauer
https://ethz.zoom.us/j/98684866685?pwd=aVFJS3VNdSsyMGVzOFQvbW5qSisrZz09

110 URSABench: Comprehensive Benchmarking of Approximate Bayesian Inference Methods for Deep Neural Networks Meet Vadera, Adam Cobb, Brian Jalaian and Benjamin Martin
https://umass-amherst.zoom.us/j/99918090662?pwd=N1E1Mm9LcEQ4bnhjNWIWdTY9

111 Robustness to Distribution Shifts using Multiple Environments Anders Andreassen, Rebecca Roeolfs and Behnam Neyshabur
https://youtube.com/98295724571?pwd=cW5QQ2EzL2RaUW15OWls3T3JrTkZ2

112 Robust Classification under Class-Dependent Domain Shift Tigran Galstyan, Hrant Khachatryan, Greg Ver Steeg and Aram Galstyan
https://us02web.zoom.us/j/82576216178?pwd=RzVGQizUNE95SN1RkA5kXWTXMrz736349

113 You Need Only Uncertain Data: Efficient Multilingual Question Answering Zhizhou Lyu, Danyu Dulongxi, Bowei Dai, Yuan Yao, Pasquale Minervini, Tim Z. Xiao and Yanir Gal
https://us02web.zoom.us/j/88327646584?pwd=ZkxBcDNRMU9jWU0OJD6SDhoYjK5z231335

114 Diverse Ensembles Improve Calibration Asa Cooper Stickland and Iain Murray
https://zoom.us/j/6892133874?pwd=OnBWS2FzWcKgQyQyvJrJaDRsbFyJUT09

115 Exploring the Uncertainty Properties of Neural Networksâ€™ Implicit Priors in the Infinite-Width Limit Ben Adlam, Jaehoon Lee, Lechao Xiao, Jeffrey Pennington and Jasper Snoek meet.google.com/agb-migr-qdt

116 Provably Robust Learning Based on Transformation-Specific Smoothing Linyi Li, Maurice Weber, Xiaojun Xu, Luka Rimanic, Shuang Yang, Tao Xie, Ce Zhang and Bo Li
https://ethz.zoom.us/j/98321715637?pwd=Zk1hS0J1NUUteWYxYmU5bVbOZD09

117 Simple and Effective VAE Training with Calibrated Decoders Oleh Rybkin, Kosants Danilidis and Sergey Levine
meet.google.com/hvr-sdbx-syc

118 Relativized Conditional Normalized Maximum Likelihood Aurick Zhou and Sergey Levine
https://berkeley.zoom.us/j/4913207525?pwd=VGN5bXRFQji9HRXJEkF65Jz3a29id09

119 Neural Networks with Recurrent Generative Feedback Yujia Huang, Chaitanya Dwivedi, Bhavya Kailkhura, Gauri Joshi and T. Yong-Jin Han
https://zoom.us/j/98397444333?pwd=b0JCTGhiQEFWThTDYd6akpDeXFzd098a3QWf

120 Robustness of Latent Representations of Variational Autoencoders Andrea Karlova https://meet.google.com/hmk-kory-kba

121 Learning approximate invariance requires far fewer data Jean Michel

Amath Sarr, Alessane Bah and Christophe Cambier
https://us04web.zoom.us/j/75476163865?pwd=Z3RES1dMWVWZkE4YjM2RzRLd2ZvZ0dp

122 MCU-Net: A framework towards uncertainty representations for decision support system patient referrals in healthcare contexts Nabeel Seedat
https://zoom.us/j/91976554789?pwd=NCxJUDJDOUlZajkxUTV6VlwxUK0kzQT09

123 Uncertainty-sensitive Learning and Planning with Ensembles Piotr Milos, Lukasz Kucinski, Konrad Czechowski, Piotr Kozakowski and Maciej Klimek
https://us02web.zoom.us/j/89309913614?pwd=cvJjDFRpaDFZVRMQXBSzZwOGF2a3d0


125 Unsupervised Domain Adaptation in the Absence of Source Data Roshni Sahoo, Divya Shamugam and John Guttag
https://mit.zoom.us/j/94674840454?pwd=L2NMaHBSaTRYZnE4TjZydmERClqUT09

126 A benchmark study on reliable molecular supervised learning via Bayesian learning Doyeong Hwang, Grace Lee, Hanseok Jo, Seyoul Yoon and Seongok Ryu https://meet.google.com/hyi-bbdm-qxo


128 Training Deep Neural Networks with Class-level Semantics for Explicable Classification Alberto Olmo, Sailik Sengupta and Subbarao Kambhampati
https://asu.zoom.us/j/93814557218?pwd=K1hjc3NqWXLIQm1p0N0HTk1rDUNJZz09

129 End-to-end Robustness for Sensing-Reasoning Machine Learning Pipelines Zhuolin Yang, Zhikuan Zhao, Hengzhi Pei, Boxin Wang, Bojan Karlas, Ji Liu, Heng Guo, Bo Li and Ce Zhang
https://illinois.zoom.us/j/2530610336?pwd=WXVZYlkeDHyuRNdfOq1kMURjUT09

130 Deep Robust Classification under Model Shift with Consistent Uncertainty Estimation Haoxuan Wang, Anqi Liu, Yisong Yue and Anima Anandkumar
https://caltech.zoom.us/j/92476182222?pwd=WS91UjhhY1p4WTHQjRRYUSLMnprJQT09

131 On Separability of Self-Supervised Representations Vikash Sehwag, Mung Chiang and Prateek Mittal
https://princeton.zoom.us/j/3871410688?pwd=RldBOUNwMnhUFz4ZGpRSVR2YWJzdnB9

132 Dropout Strikes Back: Improved Uncertainty Estimation via Diversity Sampling Evgenii Tsymbalov, Kirill Fedyanin and Maxim Panov
https://zoom.us/j/99398011263?pwd=RUJBYzUNN14ekdZwWEdZC0cz0NzJT09

133 Predictive Uncertainty for Probabilistic Novelty Detection in Text Classification Jordy Van Landeghem, Matthew Blaschko, Bertrand Anckaert and Marie-Francine Moens
https://us02web.zoom.us/j/86332883100?pwd=TVd3UytVTkF1R1RINpdTZqWlZ0

134 Maximizing the Representation Gap between In-domain & OOD examples Jay Nandy, Wynne Hsu and Mong Li Lee
https://nus-sg.zoom.us/j/99627439405?pwd=WFkNxzhdIo3TTU5YyyenduN0NwQT09

135 Regional Image Perturbation Reduces Lp Norms of Adversarial Perturbations While Maintaining Model-to-model Transferability Utku Ozbulak, Jonathan Peck, Wesley De Neve, Bart Goossens, Yvan Saets and Arnout Van Messem
https://zoom.us/j/7981481618?pwd=bFerdmJUTA3bRtTXN5U0hPa29YT09

136 Lookahead Adversarial Semantic Segmentation Hadi Jamali-Rad, Attila Szabo and Matteo Presutto meet.google.com/jgc-nfyp-nsn

137 Neural Network Calibration for Medical Imaging Classification Using Regularization Gongbo Liang, Yu Zhang and Nathan Jacobs
https://uky.zoom.us/j/93398011263?pwd=RUJBYzUNN14ekdZwWEdZC0cz0NzJT09

138 Deep k-NN Defense Against Clean-Label Data Poisoning Attacks Neehar Peri, Neal Gupta, Ronny Hung, Chen Zhu, Liam Fowl, Soheil Feizi, Tom Goldstein and John Dickerson
https://umd.zoom.us/j/93698080878?pwd=S2VjZjAzZH3RXISUIWtIRWh6K1N0dz09

139 Amortised Variational Inference for Hierarchical Mixture Models Javier Antoran, Jiayu Yao, Weiwei Pan, Finale Doshi-Velez and Jose Miguel Hernandez-Lobato
https://harvard.zoom.us/j/99584006384?pwd=NO9BS1RGNX5T1czNE5tQBgm9jdQ

140 Benchmarking Search Methods for Generating NLP Adversarial Examples Jin Yong Yoo, John X. Morris, Eli Lifland and Yanjun Qi
https://virginia.zoom.us/j/9796555398?pwd=aWd2ZW1EK211Y2wzL3ZMVGFOGW1rQjU9

Abstract 13: **Keyword #2 Finale Doshi-Velez in Uncertainty and Robustness in Deep Learning Workshop (UDL), Doshi-Velez 10:30 AM**
https://slideslive.com/38930573

Abstract 14: **Keyword #3 Percy Liang in Uncertainty and Robustness in Deep Learning Workshop (UDL), Liang 11:00 AM**
https://slideslive.com/38930574

Abstract 15: **Panel Discussion in Uncertainty and Robustness in Deep Learning Workshop (UDL), 11:30 AM**
https://us02web.zoom.us/s/85455699146?pwd=REw5QkhTNk5MV3cxUvUp5VnNzRWxh

Abstract 17: **Keyword #4 Raquel Urtasun in Uncertainty and Robustness in Deep Learning Workshop (UDL), Urtasun 01:30 PM**
https://slideslive.com/38930575

Abstract 18: **Contributed Talk 1: Confidence-Calibrated Adversarial Training: Generalizing to Unseen Attacks in Uncertainty and Robustness in Deep Learning Workshop (UDL), Stutz 02:00 PM**
https://slideslive.com/38930576

Abstract 19: **Contributed Talk 2: Improving robustness against common corruptions by covariate shift adaptation in Uncertainty and Robustness in Deep Learning Workshop (UDL), Schneider 02:10 PM**
https://slideslive.com/38930577
Beyond first order methods in machine learning systems

Albert S Berahas, Amir Gholaminejad, Tasos Kyrillidis, Michael Mahoney, Fred Roosta

Fri Jul 17, 08:00 AM

Optimization lies at the heart of many exciting developments in machine learning, statistics and signal processing. As models become more complex and datasets get larger, finding efficient, reliable and provable methods is one of the primary goals in these fields.

In the last few decades, much effort has been devoted to the development of first-order methods. These methods enjoy a low per-iteration cost and have optimal complexity, are easy to implement, and have proven to be effective for most machine learning applications. First-order methods, however, have significant limitations: (1) they require fine hyperparameter tuning, (2) they do not incorporate curvature information, and thus are sensitive to ill-conditioning, and (3) they are often unable to fully exploit the power of distributed computing architectures.

Higher-order methods, such as Newton, quasi-Newton and adaptive gradient descent methods, are extensively used in many scientific and engineering domains. At least in theory, these methods possess several nice features: they exploit local curvature information to mitigate the effects of ill-conditioning, they avoid or diminish the need for hyper-parameter tuning, and they have enough concurrency to take advantage of distributed computing environments. Researchers have even developed stochastic versions of higher-order methods, that feature speed and scalability by incorporating curvature information in an economical and judicious manner. However, often higher-order methods are undervalued.

This workshop will attempt to shed light on this statement. Topics of interest include, but are not limited to, second-order methods, adaptive gradient descent methods, regularization techniques, as well as techniques based on higher-order derivatives.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Introductory notes</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Talk by Peter Richtarik - Fast linear convergence of randomized BFGS</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Q&amp;A with Peter Richtarik</td>
</tr>
<tr>
<td>09:55 AM</td>
<td>Q&amp;A with Francis Bach</td>
</tr>
<tr>
<td>10:05 AM</td>
<td>Break until 10:30am (PDT)</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Spotlight talk 1 - A Second-Order Optimization Algorithm for Solving Problems Involving Group Sparse Regularization</td>
</tr>
<tr>
<td>10:40 AM</td>
<td>Spotlight talk 2 - Ridge Riding: Finding diverse solutions by following eigenvectors of the Hessian</td>
</tr>
<tr>
<td>10:50 AM</td>
<td>Spotlight talk 3 - PyHessian: Neural Networks Through the Lens of the Hessian</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Talk by Coralia Cartis - Dimensionality reduction techniques for large-scale optimization problems</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Q&amp;A with Coralia Cartis</td>
</tr>
<tr>
<td>11:55 AM</td>
<td>Break until 13:30pm (PDT)</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Spotlight talk 4 - MomentumRNN: Integrating Momentum into Recurrent Neural Networks</td>
</tr>
<tr>
<td>01:40 PM</td>
<td>Spotlight talk 5 - Step-size Adaptation Using Exponentiated Gradient Updates</td>
</tr>
</tbody>
</table>

This workshop will attempt to shed light on this statement. Topics of interest include, but are not limited to, second-order methods, adaptive gradient descent methods, regularization techniques, as well as techniques based on higher-order derivatives.

Beyond first order methods in machine learning systems

Albert S Berahas, Amir Gholaminejad, Tasos Kyrillidis, Michael Mahoney, Fred Roosta

Fri Jul 17, 08:00 AM

Optimization lies at the heart of many exciting developments in machine learning, statistics and signal processing. As models become more complex and datasets get larger, finding efficient, reliable and provable methods is one of the primary goals in these fields.

In the last few decades, much effort has been devoted to the development of first-order methods. These methods enjoy a low per-iteration cost and have optimal complexity, are easy to implement, and have proven to be effective for most machine learning applications. First-order methods, however, have significant limitations: (1) they require fine hyperparameter tuning, (2) they do not incorporate curvature information, and thus are sensitive to ill-conditioning, and (3) they are often unable to fully exploit the power of distributed computing architectures.

Higher-order methods, such as Newton, quasi-Newton and adaptive gradient descent methods, are extensively used in many scientific and engineering domains. At least in theory, these methods possess several nice features: they exploit local curvature information to mitigate the effects of ill-conditioning, they avoid or diminish the need for hyper-parameter tuning, and they have enough concurrency to take advantage of distributed computing environments. Researchers have even developed stochastic versions of higher-order methods, that feature speed and scalability by incorporating curvature information in an economical and judicious manner. However, often higher-order methods are undervalued.

This workshop will attempt to shed light on this statement. Topics of interest include, but are not limited to, second-order methods, adaptive gradient descent methods, regularization techniques, as well as techniques based on higher-order derivatives.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Introductory notes</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Talk by Peter Richtarik - Fast linear convergence of randomized BFGS</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Q&amp;A with Peter Richtarik</td>
</tr>
<tr>
<td>09:55 AM</td>
<td>Q&amp;A with Francis Bach</td>
</tr>
<tr>
<td>10:05 AM</td>
<td>Break until 10:30am (PDT)</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Spotlight talk 1 - A Second-Order Optimization Algorithm for Solving Problems Involving Group Sparse Regularization</td>
</tr>
<tr>
<td>10:40 AM</td>
<td>Spotlight talk 2 - Ridge Riding: Finding diverse solutions by following eigenvectors of the Hessian</td>
</tr>
<tr>
<td>10:50 AM</td>
<td>Spotlight talk 3 - PyHessian: Neural Networks Through the Lens of the Hessian</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Talk by Coralia Cartis - Dimensionality reduction techniques for large-scale optimization problems</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Q&amp;A with Coralia Cartis</td>
</tr>
<tr>
<td>11:55 AM</td>
<td>Break until 13:30pm (PDT)</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Spotlight talk 4 - MomentumRNN: Integrating Momentum into Recurrent Neural Networks</td>
</tr>
<tr>
<td>01:40 PM</td>
<td>Spotlight talk 5 - Step-size Adaptation Using Exponentiated Gradient Updates</td>
</tr>
</tbody>
</table>
### Abstract 2: Talk by Peter Richtarik - Fast linear convergence of randomized BFGS in Beyond first order methods in machine learning systems, Richtarik 08:15 AM

Since the late 1950's when quasi-Newton methods first appeared, they have become one of the most widely used and efficient algorithmic paradigms for unconstrained optimization. Despite their immense practical success, there is little theory that shows why these methods are so efficient. We provide a semi-local rate of convergence for the randomized BFGS method which can be significantly better than that of gradient descent, finally giving theoretical evidence supporting the superior empirical performance of the method.
We present PyHessian, a new scalable framework that enables fast computation of Hessian (i.e., second-order derivative) information for deep neural networks. PyHessian enables fast computation of the top Hessian eigenvalues, the Hessian trace, and the full Hessian eigenvalue/spectral density, and supports distributed-memory execution on cloud/supercomputer systems and available as open source. We show that this framework can be used to analyze neural network models, including the topology of the loss landscape (i.e., curvature information) to gain insight into the behavior of different models/optimizers. In particular, we analyze the effect of Batch Normalization layers on the trainability of NNs. We find that Batch Normalization does not necessarily make the loss landscape smoother, especially for shallow networks, as opposed to common belief.

Abstract 10: Talk by Coralia Cartis - Dimensionality reduction techniques for large-scale optimization problems in Beyond first order methods in machine learning systems, Cartis 11:00 AM

Known by many names, sketching techniques allow random projections of data from high to low dimensions while preserving pairwise distances. This talk explores ways to use sketching so as to improve the scalability of algorithms for diverse classes of optimization problems and applications, from linear to nonlinear, local to global, derivative-based to derivative-free. Regression problems and Gauss-Newton techniques will receive particular attention. Numerical illustrations on standard optimization test problems as well as on some machine learning set-ups will be presented. This work is joint with Jan Fiala (NAG Ltd), Jaroslav Fowkes (Oxford), Estelle Massart (Oxford and NPL), Adilet Otemissov (Oxford and Turing), Alex Puiu (Oxford), Lindon Roberts (Australian National University, Canberra), Zhen Shao (Oxford).

Abstract 13: Spotlight talk 4 - MomentumRNN: Integrating Momentum into Recurrent Neural Networks in Beyond first order methods in machine learning systems, Nguyen 01:30 PM

Designing deep neural networks is an art that often involves an expensive search over candidate architectures. To overcome this for recurrent neural nets (RNNs), we establish a connection between the hidden state dynamics in an RNN and gradient descent (GD). We then integrate momentum into this framework and propose a new family of RNNs, called \{tem MomentumRNNs\}. We theoretically prove and numerically demonstrate that MomentumRNNs alleviate the vanishing gradient issue in training RNNs. We study the momentum long/short term memory (MomentumLSTM) and verify its advantages in convergence speed and accuracy over its LSTM counterpart across a variety of benchmarks, with little compromise in computational or memory efficiency. We also demonstrate that MomentumRNN is applicable to many types of recurrent cells, including those in the state-of-the-art orthogonal RNNs. Finally, we show that other advanced momentum-based optimization methods, such as Adam and Nesterov accelerated gradients with a restart, can be easily incorporated into the MomentumRNN framework for designing new recurrent cells with even better performance.

Abstract 14: Spotlight talk 5 - Step-size Adaptation Using Exponentiated Gradient Updates in Beyond first order methods in machine learning systems, Armid 01:40 PM

Optimizers like Adam and AdaGrad have been very successful in training large-scale neural networks. Yet, the performance of these methods is heavily dependent on a carefully tuned learning rate schedule. We show that in many large-scale applications, augmenting a given optimizer with an adaptive tuning method of the step-size greatly improves the performance. More precisely, we maintain a global step-size scale for the update as well as a gain factor for each coordinate. We adjust the global scale based on the alignment of the average gradient and the current gradient vectors. A similar approach is used for updating the local gain factors. This type of step-size scale tuning has been done before with gradient descent updates. In this paper, we update the step-size scale and the gain variables with exponentiated gradient updates instead. Experimentally, we show that our approach can achieve compelling accuracy on standard models without using any specially tuned learning rate schedule. We also show the effectiveness of our approach for quickly adapting to distribution shifts in the data during training.

Abstract 15: Spotlight talk 6 - Competitive Mirror Descent in Beyond first order methods in machine learning systems, Schaefer 01:50 PM

Constrained competitive optimization involves multiple agents trying to minimize conflicting objectives, subject to constraints. This is a highly expressive modeling language that subsumes most of modern machine learning. In this work we propose competitive mirror descent (CMD): a general method for solving such problems based on first order information that can be obtained by automatic differentiation. First, by adding Lagrange multipliers, we obtain a simplified constraint set with an associated Bregman potential. At each iteration, we then solve for the Nash equilibrium of a regularized bilinear approximation of the full problem to obtain a direction of movement of the agents. Finally, we obtain the next iterate by following this direction according to the dual geometry induced by the Bregman potential. By using the dual geometry we obtain feasible iterates despite only solving a linear system at each iteration, eliminating the need for projection steps while still accounting for the global nonlinear structure of the constraint set. As a special case we obtain a novel competitive multiplicative weights algorithm for problems on the positive cone.

Abstract 16: Industry Panel - Talk by Boris Ginsburg - Large scale deep learning: new trends and optimization challenges in Beyond first order methods in machine learning systems, Ginsburg 02:00 PM

I will discuss two major trends in the deep learning. The first trend is an exponential growth in the size of models: from 340M (BERT-large) in 2018 to 175B (GPT3) in 2020. We need new, more memory efficient algorithms to train such huge models. The second trend is \#BERT-approachable\, when a model is first pre-trained in unsupervised or self-supervised manner on large unlabeled dataset, and then it is fine-tuned for another task using a. smaller labeled dataset. This trend sets new theoretical problems. Next, I will discuss a practical need in theoretical foundation for regularization methods used in the deep learning practice: data augmentation, dropout, label smoothing etc. Finally, I will describe an application-driven design of new optimization methods using NovoGrad as example.

Abstract 17: Industry Panel - Talk by Jonathan Hseu - ML Models in Production in Beyond first order methods in machine learning systems, Hseu 02:15 PM

We discuss the difficulties of training and improving ML models on large datasets in production. We also go into the process of an engineer working on ML models, and the challenges of trading off cost, model quality, and performance. Finally, we go into a wishlist of optimization improvements that could improve the workflow of engineers working on these models.
Abstract 18: Industry Panel - Talk by Andres Rodriguez - Shifting the DL industry to 2nd order methods in Beyond first order methods in machine learning systems, Rodriguez 02:30 PM

In this talk, we review the topology design process used by data scientists and explain why 1st order methods are computational expensive in the design process. We explore the benefits of 2nd order methods to reduce the topology design cost and highlight recent work that approximates the inverse Hessian. We conclude with recommendations to accelerate the adoption of these methods in the DL ecosystem.


Stochastic gradient descent (SGD) and its many variants serve as the workhorses of deep learning. One of the foremost pain points in using these methods in practice is hyperparameter tuning, especially the learning rate (step size). We propose a statistical adaptive procedure called SALSA to automatically schedule the learning rate for a broad family of stochastic gradient methods. SALSA first uses a smoothed line-search procedure to find a good initial learning rate, then automatically switches to a statistical method, which detects stationarity of the learning process under a fixed learning rate, and drops the learning rate by a constant factor whenever stationarity is detected. The combined procedure is highly robust and autonomous, and it matches the performance of the best hand-tuned methods in several popular deep learning tasks.

Abstract 21: Talk by Rachel Ward - Weighted Optimization: better generalization by smoother interpolation in Beyond first order methods in machine learning systems, Ward 03:30 PM

We provide a rigorous analysis of how implicit bias towards smooth interpolations leads to low generalization error in the overparameterized setting. We provide the first case study of this connection through a random Fourier series model and weighted least squares. We then argue through this model and numerical experiments that normalization methods in deep learning such as weight normalization improve generalization in overparameterized neural networks by implicitly encouraging smooth interpolants. This is work with Yuege (Gail) Xie, Holger Rauhut, and Hung-Hsu Chou.

Abstract 24: Talk by Rio Yokota - Degree of Approximation and Overhead of Computing Curvature, Information, and Noise Matrices in Beyond first order methods in machine learning systems, Yokota 05:00 PM

Hessian, Fisher, and Covariance matrices are not only used for preconditioning optimizers, but also in generalization metrics, predicting hyperparameters, and Bayesian inference. These matrices contain valuable information that can advance theory in statistical learning, but they are very expensive to compute exactly for modern deep neural networks with billions of parameters. We make use of a highly optimized implementation for computing these matrices with various degrees of approximation to close the gap between theory and practice. We are able to significantly reduce the overhead of computing these matrices through a hybrid data-parallel + model-parallel approach.
4th Lifelong Learning Workshop

Shagun Sodhani, Sarah Chandar, Balaraman Ravindran, Doina Precup

Sat Jul 18, 02:00 AM

One of the most significant and challenging open problems in Artificial Intelligence (AI) is the problem of Lifelong Learning. Lifelong Machine Learning considers systems that can continually learn many tasks (from one or more domains) over a lifetime. A lifelong learning system efficiently and effectively:

1. retains the knowledge it has learned from different tasks;
2. selectively transfers knowledge (from previously learned tasks) to facilitate the learning of new tasks;
3. ensures the effective and efficient interaction between (1) and (2).

Lifelong Learning introduces several fundamental challenges in training models that generally do not arise in a single task batch learning setting. This includes problems like catastrophic forgetting and capacity saturation. This workshop aims to explore solutions for these problems in both supervised learning and reinforcement learning settings.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>02:00 AM</td>
<td>Opening Comments</td>
<td>Chandar, Sodhani</td>
</tr>
<tr>
<td>02:15 AM</td>
<td>Challenges &amp; Opportunities in Lifelong Reinforcement Learning by</td>
<td>Hofmann, Antonova, Zintgraf</td>
</tr>
<tr>
<td></td>
<td>Katja Hoffman</td>
<td></td>
</tr>
<tr>
<td>02:45 AM</td>
<td>Q&amp;A with Katja Hoffman</td>
<td>Hofmann, Zintgraf, Antonova, Changar, Sodhani</td>
</tr>
<tr>
<td>03:00 AM</td>
<td>Never-ending Learning by Partha Pratim Talukdar</td>
<td>Talukdar</td>
</tr>
<tr>
<td>03:30 AM</td>
<td>Q&amp;A with Partha Pratim Talukdar</td>
<td>Talukdar, Sodhani, Changar</td>
</tr>
<tr>
<td>03:45 AM</td>
<td>Contributed Talk: Continual Deep Learning by Functional</td>
<td>Swaroop</td>
</tr>
<tr>
<td></td>
<td>Regularisation of Memorable Past</td>
<td></td>
</tr>
<tr>
<td>04:00 AM</td>
<td>Virtual Poster Session #1</td>
<td></td>
</tr>
<tr>
<td>06:00 AM</td>
<td>Q&amp;A with Jäürgen Schmidhuber</td>
<td>Schmidhuber, Sodhani, Changar</td>
</tr>
</tbody>
</table>

Abstracts (10):

Abstract 1: Opening Comments in 4th Lifelong Learning Workshop, Chandar, Sodhani 02:00 AM

Opening Remarks (Introduction and Overview)

Abstract 2: Challenges & Opportunities in Lifelong Reinforcement Learning by Katja Hoffman in 4th Lifelong Learning Workshop, Hofmann, Antonova, Zintgraf 02:15 AM

Lifelong reinforcement learning holds promise to enable autonomous decision making in applications ranging from household robotics to autonomous space exploration. In this talk we will discuss key challenges that need to be address to make progress towards this grand vision. First, we discuss the fundamental need for the community to invest in a shared understanding of problem formalizations and
evaluation paradigms and benchmarks. Second, we dive deeper into ongoing work towards tackling exploration and representation in lifelong reinforcement learning. Our aim is to spark debate and inspire research in this exciting space.

Abstract 4: Never-ending Learning by Partha Pratim Talukdar in 4th Lifelong Learning Workshop, Talukdar 03:00 AM

Humans learn many different functions and skills, from diverse experiences gained over many years, from a staged curriculum in which they first learn easier and later more difficult tasks, retain the learned knowledge and skills, which are used in subsequent learning to make it easier or more effective. Furthermore, humans self-reflect on their evolving skills, choose new learning tasks over time, teach one another, learn new representations, read books, discuss competing hypotheses, and more. In this talk, I shall share my thoughts on the question of how to design machine learning agents with similar capabilities.

Abstract 7: Virtual Poster Session #1 in 4th Lifelong Learning Workshop, 04:00 AM

* **[Sharing Less is More: Lifelong Learning in Deep Networks with Selective Layer Transfer](https://openreview.net/forum?id=G8NSCGPEJM)** || [Video](https://www.youtube.com/watch?v=Aw9wr8BeA50) || [Zoom Session](https://us02web.zoom.us/j/8173180936?pwd=UUsTYE3RDnNlWisSnh8cHh4Y2ZxQT09)

* **[Evaluating Local Generalization in Graph Neural Networks](https://openreview.net/forum?id=KjvCLOxMXU)** || [Video](https://www.youtube.com/watch?v=kEjAAc5mgg) || [Zoom Session](https://us02web.zoom.us/j/8695084805?pwd=SmxS9G5E2G5E5G6WxN0ZjZGV0ZjNvQT09)

* **[Efficient Adaptation for End-to-End Vision-Based Robotic Manipulation](https://openreview.net/forum?id=CNVsGzBFFFF)** || [Video](https://www.youtube.com/watch?v=psc9DzZu0W4) || [Zoom Session](https://us02web.zoom.us/j/8355409241?pwd=SEsTUIMVE1z9kFtXlVgQHHYwUJz09)

* **[Wandering Within a World: Online Contextualized Few-Shot Learning](https://openreview.net/forum?id=9gF8dXwugJ_L)** || [Video](https://www.youtube.com/plCvJwWmI2) || [Zoom Session](https://us02web.zoom.us/j/8725607019?pwd=TjVuJl0MeXVdXISEF2GhWJl09)

* **[Taming the Herd: Multi-Modal Meta-Learning with a Population of Agents](https://openreview.net/forum?id=SLcAnWO-vz)** || [Video](https://www.youtube.com/36yP-NNi) || [Zoom Session](https://us02web.zoom.us/j/8562192988?pwd=LzlsSGJVbJgUSU5FvQm5M4ZfTENhZ19)

* **[Lifelong Learning of Factored Policies via Policy Gradients](https://openreview.net/forum?id=vY2duSBw)** || [Video](https://www.youtube.com/watch?v=aZoRM8ByFw) || [Zoom Session](https://us02web.zoom.us/j/8866877816?pwd=MQNtbkDCd2g0aTg3U9a4yZGFzDF-19)

* **[Learning Intrinsically Motivated Options to Stimulate Policy Exploration](https://openreview.net/forum?id=VcHIDmBYJk)** || [Video](https://www.youtube.com/hY5FrXcYxj) || [Zoom Session](https://us02web.zoom.us/j/8568106826?pwd=ZXBnN9C9VMtSGF0aTi1bRDnWb4b09)

* **[Hierarchical Expert Networks for Meta-Learning](https://openreview.net/forum?id=vADVUH-NI)** || [Video](https://www.youtube.com/B0wHa4NEj) || [Zoom Session](https://us02web.zoom.us/j/81650691865?pwd=ZJFUQStiWVFtTnd2MKJmNE9)

* **[La-MAML: Look-ahead Meta Learning for Continual Learning](https://openreview.net/forum?id=G_N9PeXIC)** || [Video](https://youtu.be/HzewyVubLA) || [Zoom Session](https://us02web.zoom.us/j/8986934709?pwd=OFpZT2Cak9TeTFtSGFzF19)

* **[Covariate Distribution aware Meta-learning](https://openreview.net/forum?id=01vmjYESeO)** || [Video](https://www.youtube.com/watch?v=ERuB8jRkZuE) || [Zoom Session](https://us02web.zoom.us/j/8442358204?pwd=bWwnkRWuR2WhJHFTDz09)

* **[Multilayer Neuromodulated Architectures for Memory-Constrained Online Continual Learning](https://openreview.net/forum?id=Sy55E8TNwB)** || [Video](https://www.youtube.com/watch?v=Rt6JnG7mW) || [Zoom Session](https://us02web.zoom.us/j/8725498204?pwd=UGRAW2ZJZmSTUE9HmHyY)

* **[Tackling Non-forgetting and Forward Transfer with a Unified Lifelong Learning Approach](https://openreview.net/forum?id=OwoxV2K3z)** || [Video](https://www.youtube.com/watch?v=n4N7sL3k) || [Zoom Session](https://us02web.zoom.us/j/8472508245?pwd=MmdweVdOCyUupz3c3twCT09)

* **[Rewriting History with Inverse RL: Hindsight Inference for Policy Evaluation](https://openreview.net/forum?id=6PWQUYK_ips)** || [Video](https://www.youtube.com/watch?v=tnGp) || [Zoom Session](https://us02web.zoom.us/j/885105153?pwd=WFRDKiJcVjJdGABrW5Rm)

* **[Importance Weighting with a Adversarial Network for Large-Scale Sleep Staging](https://openreview.net/forum?id=rGhLSQvQy)** || [Video](https://www.youtube.com/watch?v=9qJq4nKnp) || [Zoom Session](https://us02web.zoom.us/j/8600712581?pwd=UlY9KJZIU0cU5R4uW1h)

* **[Unsupervised Progressive Learning and the STEM Architecture](https://openreview.net/forum?id=a4khlAFyx-)** || [Video](https://www.youtube.com/watch?v=3Z60eS94F8t) || [Zoom Session](https://us02web.zoom.us/j/8292426655?pwd=eGtrVEExM3hDbGUFySFXzT2)

* **[Active Online Domain Adaptation](https://openreview.net/forum?id=VLxyWdAsb)** || [Video](https://www.youtube.com/vmn509ck) || [Zoom Session](https://us02web.zoom.us/j/843567766?pwd=MU1uL0ZrY3V3yUUpL5w)

* **[Bringing Worlds in Reinforcement Learning with Model-Advantage](https://openreview.net/forum?id=xBRYXAc_fQ)** || [Video](https://www.youtube.com/a7MopkK8zg) || [Zoom Session](https://us02web.zoom.us/j/8954780083?pwd=CrnT2GFD-w8)

* **[Towards an Unsupervised Method for Model Selection in Few-Shot Learning](https://openreview.net/forum?id=wG02NgC-ua9)** || [Video](https://www.youtube.com/watch?v=nn8k) || [Zoom Session](https://us02web.zoom.us/j/89819767782?pwd=REhTAi5mVv0w)

* **[Generalisation Guarantees for Continual Learning with Orthogonal Gradient Descent](https://openreview.net/forum?id=mzgcEGD-yk4)** || [Video](https://www.youtube.com/watch?v=-P10xPd6fY) || [Zoom Session](https://us02web.zoom.us/j/86995088405?pwd=SnMW9G5E2G5E5G6WxN0ZjZGV0ZjNvQT09)
**[Attention-Critic]** ([https://openreview.net/forum?id=q69w5xjFfK](https://openreview.net/forum?id=q69w5xjFfK))

* **[Continual Learning from the Perspective of Compression]** ([https://openreview.net/forum?id=F0y6dPn57](https://openreview.net/forum?id=F0y6dPn57))

**[Attention]** ([https://openreview.net/forum?id=96wFs9pOFkI](https://openreview.net/forum?id=96wFs9pOFkI))

**[Efficient Diversity in Population Based Reinforcement Learning]** ([https://openreview.net/forum?id=ERiQ-MzV3d](https://openreview.net/forum?id=ERiQ-MzV3d))

**[Chaotic Continual Learning]** ([https://openreview.net/forum?id=rQvRqJf8L3](https://openreview.net/forum?id=rQvRqJf8L3))

**[Task Agnostic Continual Learning via Meta Learning]** ([https://openreview.net/forum?id=a2zVxdJedg](https://openreview.net/forum?id=a2zVxdJedg))

**[Logical Composition in Lifelong Reinforcement Learning]** ([https://openreview.net/forum?id=pB9FItPvFg](https://openreview.net/forum?id=pB9FItPvFg))

**[Off-Policy Meta-Reinforcement Learning Based on Feature Embedding Spaces]** ([https://openreview.net/forum?id=ib8AZrpf_dY](https://openreview.net/forum?id=ib8AZrpf_dY))

**[Off-Policy Adversarial Inverse Reinforcement Learning]** ([https://openreview.net/forum?id=9mp5d073IhX](https://openreview.net/forum?id=9mp5d073IhX))

**[Chaotic Continual Learning** ([https://openreview.net/forum?id=9mp5d073IhX](https://openreview.net/forum?id=9mp5d073IhX))

**[Efficient Diversity in Population Based Reinforcement Learning]** ([https://openreview.net/forum?id=--97Sv2Y02I](https://openreview.net/forum?id=--97Sv2Y02I))

**[Continuous Deep Learning by Functional Regularisation of Memorable Past]** ([https://openreview.net/forum?id=_b_vapuIazp](https://openreview.net/forum?id=_b_vapuIazp))

**[Logical Composition in Lifelong Reinforcement Learning** ([https://openreview.net/forum?id=pB9FItPvFg](https://openreview.net/forum?id=pB9FItPvFg))

**[Off-Policy Meta-Reinforcement Learning Based on Feature Embedding Spaces** ([https://openreview.net/forum?id=ib8AZrpf_dY](https://openreview.net/forum?id=ib8AZrpf_dY))

**[Deep Reinforcement Learning amidst Lifelong Non-Stationarity** ([https://openreview.net/forum?id=P1OwHAhDVbd](https://openreview.net/forum?id=P1OwHAhDVbd))


**[Exchangeable Models in Meta Reinforcement Learning** ([https://openreview.net/forum?id=SD7m4B3kGiQ](https://openreview.net/forum?id=SD7m4B3kGiQ))

**[ICML 2020 Workshop book** ([https://us02web.zoom.us/j/83155533983?pwd=eEx3bGhwzTjr0NoN2JUUnI](https://us02web.zoom.us/j/83155533983?pwd=eEx3bGhwzTjr0NoN2JUUnI))

**[Effective Diversity in Population Based Reinforcement Learning** ([https://openreview.net/forum?id=95TvzY20y](https://openreview.net/forum?id=95TvzY20y))


**[Task Agnostic Continual Learning via Meta Learning** ([https://openreview.net/forum?id=Okmqx6qXK](https://openreview.net/forum?id=Okmqx6qXK))

**[Off-Policy Adversarial Inverse Reinforcement Learning** ([https://openreview.net/forum?id=9mp5d073IhX](https://openreview.net/forum?id=9mp5d073IhX))

**[Chaotic Continual Learning** ([https://openreview.net/forum?id=rQvRqJf8L3](https://openreview.net/forum?id=rQvRqJf8L3))

**[Task Agnostic Continual Learning via Meta Learning** ([https://openreview.net/forum?id=Okmqx6qXK](https://openreview.net/forum?id=Okmqx6qXK))

**[Logical Composition in Lifelong Reinforcement Learning** ([https://openreview.net/forum?id=pB9FItPvFg](https://openreview.net/forum?id=pB9FItPvFg))

**[Off-Policy Meta-Reinforcement Learning Based on Feature Embedding Spaces** ([https://openreview.net/forum?id=ib8AZrpf_dY](https://openreview.net/forum?id=ib8AZrpf_dY))

**[Deep Reinforcement Learning amidst Lifelong Non-Stationarity** ([https://openreview.net/forum?id=P1OwHAhDVbd](https://openreview.net/forum?id=P1OwHAhDVbd))

**[Meta-Learning for Recalibration of EMG-Based Upper Limb Prostheses** ([https://openreview.net/forum?id=lr5JdL6w](https://openreview.net/forum?id=lr5JdL6w))

**[RATT: Recurrent Attention to Transient Tasks for Continual Image Captioning** ([https://openreview.net/forum?id=DiHyudBShm](https://openreview.net/forum?id=DiHyudBShm))

**[Effective Diversity in Population Based Reinforcement Learning** ([https://openreview.net/forum?id=95TvzY20y](https://openreview.net/forum?id=95TvzY20y))

ICML 2020 Workshop book

Abstract 8: Credit assignment and meta-learning in a single lifelong trial by JÅ‘rgen Schmidhuber in 4th Lifelong Learning Workshop, Schmidhuber 06:00 AM

Most current artificial reinforcement learning (RL) agents are trained under the assumption of repeatable trials, and are reset at the beginning of each trial. Humans, however, are never reset. Instead, they are allowed to discover computable patterns across trials, e.g. in every third trial, go left to obtain reward, otherwise go right. General RL (sometimes called AGI) must assume a single lifelong trial which may or may not include identifiable sub-trials. General RL must also explicitly take into account that policy changes in early life may affect properties of later sub-trials and policy changes. In particular, General RL must take into account recursively that early meta-meta-learning is setting the stage for further meta-learning which is setting the stage for later learning etc. Most popular RL mechanisms, however, ignore such lifelong credit assignment chains. Exceptions are the success story algorithm (1990s), AIXI (2000s), and the mathematically optimal Godel Machine (2003).

Related work:
Reinforcement Learning: A Review and Perspectives.


Abstract 14: Virtual Poster Session #2 in 4th Lifelong Learning Workshop. 08:00 AM

**[Sharing Less is More: Lifelong Learning in Deep Networks with Selective Layer Transfer](https://openreview.net/forum?id=G8NSCGPE-JM)** || Video(https://www.youtube.com/watch?v=aZwrr8BqAS0) || [Zoom Session](https://us02web.zoom.us/j/81731809364?pwd=UUtJteY1E3RddnNWl5SHhPhQ5FQ1JlQT09)

**[Evaluating Logical Generalization in Graph Neural Networks](https://openreview.net/forum?id=KjvCLoxcMXU)** || Video(https://www.youtube.com/watch?v=KTejA4m4jg&feature=youtu.be) || [Zoom Session](https://us02web.zoom.us/j/8695088405?pwd=SnMxSHh2OG5zEwGmQ0d3Q2s4bXFnQT09)

**[Efficient Adaptation for End-to-End Vision-Based Robotic Manipulation](https://openreview.net/forum?id=CVNScZBFfG) || Video(https://youtu.be/dj8oCgZbwN4) || [Zoom Session](https://us02web.zoom.us/j/8355409215?pwd=SEIsTUIQVE1z9KexibGg5ZVwZycyZbQ09)

**[Wandering Within a World: Online Contextualized Few-Shot Learning](https://openreview.net/forum?id=99gdBuwq_Q) || Video(https://youtu.be/PLcWwWmmU2I) || [Zoom Session](https://us02web.zoom.us/j/87265700196?pwd=TJuallM0eXVJYTRISOF2SEjBvZBTCg09)

**[Taming the Herd: Multi-Modal Meta-Learning with a Population of Agents](https://openreview.net/forum?id=SLcfoWO-2T) || Video(https://youtu.be/36byP-NNI1o) || [Zoom Session](https://us02web.zoom.us/j/85821929987?pwd=LzlsTGJveXVJaHJ0UHJ0a3d0Z1I3QT09)

**[Lifelong Learning of Factored Policies via Policy Gradient](https://openreview.net/forum?id=v2duy_SbwA) || Video(https://www.youtube.com/watch?v=a2oRM4v8Fw) || [Zoom Session](https://us02web.zoom.us/j/86666778164?pwd=MUNTBtbkcd2q0aTg3U945Zm5dE1J6dGc09)

**[Learning Intrinsically Motivated Options to Stimulate Policy Exploration](https://openreview.net/forum?id=Vcf1DmBYjK) || Video(https://youtu.be/hyY5FrzXCyj) || [Zoom Session](https://us02web.zoom.us/j/85681063264?pwd=ZXBBDni9CVzSVGF0aTI1bGE4Wj9yQT09)

**[Hierarchical Expert Networks for Meta-Learning](https://openreview.net/forum?id=vADVUh-N11) || Video(https://www.youtube.com/watch?v=BoWaH44NNeig) || [Zoom Session](https://us02web.zoom.us/j/8165069185?pwd=ZFUQSSWVFtNgZ2MkMmY5EGnqZjV5OQ09)

**[La-MAML; Look-ahead Meta Learning for Continual Learning](https://www.openreview.net/forum?id=G_N9PBeX-8) || Video(https://youtu.be/HzeyvyU8L4Y) || [Zoom Session](https://us02web.zoom.us/j/89699470974?pwd=OPfPZT2ICank90FTlISGhM5bEj5bQ1JlQT09)

**[Covariate Distribution aware Meta-learning](https://openreview.net/forum?id=0v1jY6SeO0) || Video(https://youtu.be/0BrCTESEuY) || [Zoom Session](https://us02web.zoom.us/j/84418366038?pwd=bWNNRWlURwXMjHFDDTDAzLS0tOT09)

**[Multilayer Neuromodulated Architectures for Memory-Constrained Online Continual Learning](https://openreview.net/forum?id=SySi5E8TnWB) || Video(https://youtu.be/RIf6cJnG7MW) || [Zoom Session](https://us02web.zoom.us/j/84725482047?pwd=MmdwVo44OcJuyUUpwZ3cTQT09)

**[Tackling Non-forgetting and Forward Transfer with a Unified Lifelong Learning Approach](https://openreview.net/forum?id=OwWxRk698) || Video(https://www.youtube.com/watch?v=n4N7s9l_bk) || [Zoom Session](https://us02web.zoom.us/j/84725482047?pwd=MmdwVo44OcJuyUUpwZ3cTQT09)

**[Rewriting History with Inverse RL: Hindsight Inference for Policy Improvement](https://openreview.net/forum?id=6PwQVRKQ_I) || Video(https://youtu.be/36yP-NNlB1O) || [Zoom Session](https://us02web.zoom.us/j/82453690116?pwd=OXB6aG81aUrA5VNNmTm6bJc09)

**[Off-Dynamics Reinforcement Learning: Training for Transfer with Domain Classifiers](https://openreview.net/forum?id=JPaVK6cu3w) || Video(https://youtu.be/6SU0PGCZqxs) || [Zoom Session](https://us02web.zoom.us/j/88510158153?pwd=WFRD4kcvJYa2DFabWl5RmZ09)

**[Importance Weighting with a Adversarial Network for Large-Scale Sleep Staging](https://openreview.net/forum?id=GhsL5qVny) || Video(https://www.youtube.com/watch?v=5koxK9Kp1Q) || [Zoom Session](https://us02web.zoom.us/j/8600772581?pwd=Uy9KSFJZU0U5O0U5U0U5U0U5L1h9)

**[Unsupervised Progressive Learning and the STAM Architecture](https://openreview.net/forum?id=a4kkwANFy-) || Video(https://youtu.be/lcPEmXhD5bb) || [Zoom Session](https://us02web.zoom.us/j/82924265659?pwd=eGrVEExM3h0DGy5SFZ7nT2O)

**[Active Online Domain Adaptation](https://openreview.net/forum?id=vLxY5dAb) || Video(https://www.youtube.com/vrvoX56Q7c) || [Zoom Session](https://us02web.zoom.us/j/84356077666?pwd=MU1uL0ZCyV4TW0UIpL5w)

**[Bridging Worlds in Reinforcement Learning with Model-Advantage](https://openreview.net/forum?id=XBRYX4c_XFQ) || Video(https://youtu.be/AlMop6K58go) || [Zoom Session](https://us02web.zoom.us/j/89547800837?pwd=dEhLM3ZFeHFCFwJp6nNhXUO9)

**[Towards an Unsupervised Method for Model Selection in Few-Shot Learning](https://openreview.net/forum?id=UGO2NgC-ua9) || Video(https://www.youtube.com/watch?v=RAoHrTB4Ctw) || [Zoom Session](https://us02web.zoom.us/j/89819767782?pwd=UyIuL0ZCyV4TW0UJpL5w)

**[Generalisation Guarantees for Continual Learning with Orthogonal Gradient Descent](https://openreview.net/forum?id=mzgcEGD-yk4) || Video(https://www.youtube.com/watch?v=p1rOXpD6Y) || [Zoom Session](https://us02web.zoom.us/j/8133699281?pwd=amJUdDlFS1J6OWNQM0pxQnJLZmFLZmJ9)

**[Explore then Execute: Adapting without Rewards via Factorized Meta-Adaptation](https://openreview.net/forum?id=OwoxbRKze6i) || Video(https://youtu.be/EiC00Rk2-s) || [Zoom Session](https://us02web.zoom.us/j/8167000053?pwd=QjM5UWRLzV2YnNhdMs2ZjZ9)

**[Federated Continual Learning with Weighted Inter-client Transfer](https://openreview.net/forum?id=XLr6qOCJU3m) || Video(https://youtu.be/6SU0PGCZqxs) || [Zoom Session](https://us02web.zoom.us/j/82453690116?pwd=OXB6aG81aUrA5VNNmTm6bJc09)
A life-long learning agent should learn not only to solve problems, but also to pose new problems for itself. In reinforcement learning, the starting problems are maximizing reward and predicting value, and the natural new problems are achieving subgoals and predicting what will happen next. There has been a lot of work that provides a language for learning new problems (e.g., on auxiliary tasks and general value functions), but precious little that actually learns them (e.g., McGovern on learning subgoal states). In this talk I present a general strategy for learning new problems and, moreover, for learning an endless cycle of problems and solutions, each leading to the other. I call this cycle the FOAK cycle, because it is based on Features, Options, And Knowledge, and refers to an agent's option-conditional model of the transition dynamics of the world. The new problems in the FOAK cycle are 1) to find options that attain state features and 2) to model the consequences of those options. As these problems are solved and the models are used in planning, more abstract features are formed and made the basis for new options and models, continuing the cycle. The FOAK cycle is intended to produce a model-based reinforcement learning agent with successively more abstract representations and knowledge of its world, in other words, a life-long learning agent.

Abstract 20: Panel Discussion in 4th Lifelong Learning Workshop, Eaton, White, Precup, Rish, van Seijen 10:30 AM

**You can submit questions for the panel at https://app.sli.do/event/x5illav/live/questions**

Abstract 22: Accepted Papers in 4th Lifelong Learning Workshop, N/A

**[PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees](https://openreview.net/forum?id=a5rlUm5zZR)**
**[La-MAML: Look-ahead Meta Learning for Continual Learning](https://openreview.net/forum?id=6IqTG69Lkky)**
**[Hierarchical Expert Networks for Hierarchical Reinforcement Learning](https://openreview.net/forum?id=vADVUh-NI1)**
**[Hierarchical reinforcement learning for efficient exploration and transfer](https://openreview.net/forum?id=slq7G99Lklyy)**
**([Video](https://youtu.be/hY5FzrXCyIk))**

**[Tasklets](https://openreview.net/forum?id=Wh27ubUkkz)**
**[Video](https://www.youtube.com/watch?v=TKEjaA4m4jg&feature=youtu.be)**

**[Continual Learning of Object Instances](https://openreview.net/forum?id=o8uwQONJ22)**
**[Meta-Learning GNN Initializations for Low-Resource Molecular Transfer](https://openreview.net/forum?id=MQ_t7LRvsW)**
**([Video](https://youtu.be/F6uIGzCwFCk))**

**[Gradient Based Memory Editing for Task-Free Continual Learning](https://openreview.net/forum?id=6MEKW3WzRFy)**
**([Video](https://www.youtube.com/watch?v=JaCVn5ZBFFFF)**

**[Storing Encoded Episodes as Concepts for Continual Learning](https://openreview.net/forum?id=7rVoGCQRmw)**
**([Video](https://youtu.be/aaoSMcD_drTg))**

**[Selective Layer Gradients](https://openreview.net/forum?id=vY2du_5Bwa)**
**[Hierarchical abstraction for selective adaptation](https://openreview.net/forum?id=SLcfDWo-ztp)**
**([Video](https://youtu.be/B0wHa4NNejg))**

**[Wandering Within a World: Online Contextualized Few-Shot Transfer](https://openreview.net/forum?id=36yP-NNlB1o)**
**([Video](https://www.youtube.com/watch?v=TCrGsd66oM8))**

**[Taming the Herd: Multi-Modal Meta-Learning with a Population of Agents](https://openreview.net/forum?id=367YjP6P91)**
**([Video](https://youtu.be/klLJy2lC7Vw))**

**[Efficient Adaptation for End-to-End Vision-Based Robotic Manipulation](https://openreview.net/forum?id=LRq5z00QfYY)**
**([Video](https://www.youtube.com/watch?v=tKJEaA4m4jg&feature=youtu.be))**

**[Hierarchical Learning of Factored Policies via Policy Gradients](https://openreview.net/forum?id= مر2du_5Bwa)**
**([Video](https://www.youtube.com/watch?v=aZoRM48pVFw))**

**[Efficient Intrinsically Motivated Options to Stimulate Policy Exploration](https://openreview.net/forum?id=vFdfDMbYJk)**
**([Video](https://youtu.be/hY5FzrXCyIk))**

**[Hierarchical Expert Networks for Meta-Learning](https://openreview.net/forum?id=vADVUh-N1I)**
**([Video](https://youtu.be/B0wHa4NNej))**

**[La-MAML: Look-ahead Meta Learning for Continual Learning](https://openreview.net/forum?id=D9NePeXIC-9)**
**([Video](https://youtu.be/HzewyVuLay))**

**[Covariate Distribution aware Meta-learning](https://openreview.net/forum?id=01mv/YEbSeO)**

Abstract 17: The FOAK Cycle for Model-based Life-long Learning by Rich Sutton in 4th Lifelong Learning Workshop, Sutton 09:30 AM
ICML 2020 Workshop

Generated Tue Sep 29, 2020

* [Pseudo-Rehearsal for Continual Learning with Normalizing Flows](https://openreview.net/forum?id=vJYj8aFofy)** || [Video](https://youtu.be/aVEs7z72aA0) || [Zoom]

* [Curriculum Learning with Diversity for Supervised Computer Vision Tasks](https://openreview.net/forum?id=WH77BdU7kzj)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]


* [Generalizing Curricula for Reinforcement Learning](https://openreview.net/forum?id=TYCysL_070N)** || [Video](https://youtu.be/MyPm5ZwqA9y) || [Zoom]

* [A Brief Look at Generalization in Visual Meta-Reinforcement Learning](https://openreview.net/forum?id=WcFlzV0Evm)** || [Video](https://youtu.be/wjCv7yPpssQ) || [Zoom]

* [PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees](https://openreview.net/forum?id=a5rImUm5rZR)** || [Video](https://youtu.be/CIDQT96k3w) || [Zoom]

* [Task Agnostic Continual Learning via Meta Learning](https://openreview.net/forum?id=AelzVzXjgeb)** || [Video](https://youtu.be/nmKk3f31z1g) || [Zoom]

* [Continual Learning of Object Recognition](https://openreview.net/forum?id=9e5YEuJ5rZm)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]

* [Continual Learning from the Perspective of Compression](https://openreview.net/forum?id=F0yk8u5pNS7)** || [Video](https://youtu.be/2aRNpZw0n) || [Zoom]

* [Continual Learning with Diversity](https://openreview.net/forum?id=6M6K3W3zRr)** || [Video](https://youtu.be/aSOMdFtrG) || [Zoom]

* [End-to-end learning of reusable skills through intrinsic motivation](https://openreview.net/forum?id=v_aUdGzzLV)** || [Video](https://youtu.be/TkKJCfKDK1o) || [Zoom]

* [Gradient Based Memory Editing for Task-Free Continual Learning](https://openreview.net/forum?id=6meK3W3zRr)** || [Video](https://youtu.be/rHvJGZCv) || [Zoom]

* [Task Agnostic Continual Learning via Meta Learning](https://openreview.net/forum?id=AelzVzXjgeb)** || [Video](https://youtu.be/nmKk3f31z1g) || [Zoom]

* [Hierarchical reinforcement learning for efficient exploration and transfer](https://openreview.net/forum?id=6lqTQ69Lklk)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]

* [Off-Policy Meta-Reinforcement Learning Based on Feature Embedding Spaces](https://openreview.net/forum?id=ib8AZrpf_dY)** || [Video](https://youtu.be/BA3ymn7v0) || [Zoom]

* [Meta-Learning GNN Initializations for Low-Resource Molecular Property Prediction](https://openreview.net/forum?id=MQ_7LRvssW)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]

* [Exchangeable Models in Meta Reinforcement Learning](https://openreview.net/forum?id=tFzSejFPT)** || [Video](https://youtu.be/tvSNp_zGo4) || [Zoom]

* [Effective Diversity in Population Based Reinforcement Learning](https://openreview.net/forum?id=97Syv2YqO)** || [Video](https://www.youtube.com/watch?v=2k26xPw3CA&feature=youtu.be) || [Zoom]

* [Pseudo-Rehearsal for Continual Learning with Normalizing Flows](https://openreview.net/forum?id=vJYj8aFofy)** || [Video](https://youtu.be/aVEs7z72aA0) || [Zoom]

* [A Policy Gradient Method for Task-Agnostic Exploration](https://openreview.net/forum?id=d9j_RNHtQEO)** || [Video](https://youtu.be/l0t7dWtJMuA) || [Zoom]

* [Optimizing for the Future in Non-Stationary MDPs](https://openreview.net/forum?id=H2kOzrpzblb2)** || [Video](https://icml.cc/virtual/2020/paper/6316) || [Zoom]


* [Generalizing Curricula for Reinforcement Learning](https://openreview.net/forum?id=TYCysL_070N)** || [Video](https://youtu.be/MyPm5ZwqA9y) || [Zoom]

* [A Brief Look at Generalization in Visual Meta-Reinforcement Learning](https://openreview.net/forum?id=WcFlzV0Evm)** || [Video](https://youtu.be/wjCv7yPpssQ) || [Zoom]

* [PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees](https://openreview.net/forum?id=a5rImUm5rZR)** || [Video](https://youtu.be/CIDQT96k3w) || [Zoom]

* [Task Agnostic Continual Learning via Meta Learning](https://openreview.net/forum?id=AelzVzXjgeb)** || [Video](https://youtu.be/nmKk3f31z1g) || [Zoom]

* [Hierarchical reinforcement learning for efficient exploration and transfer](https://openreview.net/forum?id=6lqTQ69Lklk)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]

* [Off-Policy Meta-Reinforcement Learning Based on Feature Embedding Spaces](https://openreview.net/forum?id=ib8AZrpf_dY)** || [Video](https://youtu.be/BA3ymn7v0) || [Zoom]

* [Meta-Learning GNN Initializations for Low-Resource Molecular Property Prediction](https://openreview.net/forum?id=MQ_7LRvssW)** || [Video](https://youtu.be/83eOZe6yj8o) || [Zoom]

* [Exchangeable Models in Meta Reinforcement Learning](https://openreview.net/forum?id=tFzSejFPT)** || [Video](https://youtu.be/tvSNp_zGo4) || [Zoom]

* [Effective Diversity in Population Based Reinforcement Learning](https://openreview.net/forum?id=97Syv2YqO)** || [Video](https://www.youtube.com/watch?v=2k26xPw3CA&feature=youtu.be) || [Zoom]

* [Pseudo-Rehearsal for Continual Learning with Normalizing Flows](https://openreview.net/forum?id=vJYj8aFofy)** || [Video](https://youtu.be/aVEs7z72aA0) || [Zoom]

* [A Policy Gradient Method for Task-Agnostic Exploration](https://openreview.net/forum?id=d9j_RNHtQEO)** || [Video](https://youtu.be/l0t7dWtJMuA) || [Zoom]

* [Optimizing for the Future in Non-Stationary MDPs](https://openreview.net/forum?id=H2kOzrpzblb2)** || [Video](https://icml.cc/virtual/2020/paper/6316) || [Zoom]

INNF-: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models

Chin-Wei Huang, David Krueger, Rianne Van den Berg, George Papamakarios, Chris Cremer, Ricky T. Q. Chen, Danilo J. Rezende

Sat Jul 18, 02:25 AM

Neural networks are explicit likelihood models using invertible neural networks to construct flexible probability distributions of high-dimensional data. Compared to other generative models, the main advantage of normalizing flows is that they can offer exact and efficient likelihood computation and data generation. Since their recent introduction, normalizing flows have seen a significant resurgence of interest in the machine learning community. As a result, powerful flow-based models have been developed, with successes in density estimation, variational inference, and generative modeling of images, audio and video.

This workshop is the 2nd iteration of the ICML 2019 workshop on Neural Networks for Probabilistic Modeling: Flows, Generative Models, and Explicit Likelihoods.
Invertible Neural Networks and Normalizing Flows. While the main goal of last year’s workshop was to make flow-based models more accessible to the general machine learning community, as the field is moving forward, we believe there is now a need to consolidate recent progress and connect ideas from related fields. In light of the interpretation of latent variable models and autoregressive models as flows, this year we expand the scope of the workshop and consider likelihood-based models more broadly, including flow-based models, latent variable models and autoregressive models. We encourage the researchers to use these models in conjunction to exploit their benefits at once, and to work together to resolve some common issues of likelihood-based methods, such as mis-calibration of out-of-distribution uncertainty.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>02:25 AM</td>
<td>Opening remarks</td>
</tr>
<tr>
<td>02:30 AM</td>
<td>Invited talk 1: Unifying VAEs and Flows</td>
</tr>
<tr>
<td>03:05 AM</td>
<td>Q&amp;A with Max Welling</td>
</tr>
<tr>
<td>03:10 AM</td>
<td>Spotlight talk: Neural Manifold Ordinary Differential Equations</td>
</tr>
<tr>
<td>03:15 AM</td>
<td>Spotlight talk: The Convolution Exponential</td>
</tr>
<tr>
<td>03:20 AM</td>
<td>Spotlight talk: WaveNODE: A Continuous Normalizing Flow for Speech Synthesis</td>
</tr>
<tr>
<td>03:25 AM</td>
<td>Spotlight talk: Neural Ordinary Differential Equations on Manifolds</td>
</tr>
<tr>
<td>03:30 AM</td>
<td>Poster session 1</td>
</tr>
<tr>
<td>04:10 AM</td>
<td>Invited talk 2: Detecting Distribution Shift with Deep Generative Models</td>
</tr>
<tr>
<td>04:35 AM</td>
<td>Q&amp;A with Eric Nalisnick</td>
</tr>
<tr>
<td>04:40 AM</td>
<td>Invited talk 3: Representational limitations of invertible models</td>
</tr>
<tr>
<td>05:05 AM</td>
<td>Q&amp;A with Emilien Dupont</td>
</tr>
<tr>
<td>05:10 AM</td>
<td>Spotlight talk: You say Normalizing Flows I see Bayesian Networks</td>
</tr>
<tr>
<td>05:15 AM</td>
<td>Spotlight talk: Variational Inference with Continuously-Indexed Normalizing Flows</td>
</tr>
<tr>
<td>05:20 AM</td>
<td>Spotlight talk: NOTAGAN: Flows for the data manifold</td>
</tr>
<tr>
<td>05:25 AM</td>
<td>Spotlight talk: Ordering Dimensions with Nested Dropout Normalizing Flows</td>
</tr>
<tr>
<td>05:30 AM</td>
<td>Spotlight talk: The Lipschitz Constant of Self-Attention</td>
</tr>
<tr>
<td>05:35 AM</td>
<td>Spotlight talk: Autoregressive flow-based causal discovery and inference</td>
</tr>
<tr>
<td>05:40 AM</td>
<td>Lunch break</td>
</tr>
<tr>
<td>07:00 AM</td>
<td>Invited talk 4: Divergence Measures in Variational Inference and How to Choose Them</td>
</tr>
<tr>
<td>07:25 AM</td>
<td>Q&amp;A with Cheng Zhang</td>
</tr>
<tr>
<td>07:30 AM</td>
<td>Invited talk 5: Adversarial Learning of Prescribed Generative Models</td>
</tr>
<tr>
<td>07:55 AM</td>
<td>Q&amp;A with Adji Bousso Dieng</td>
</tr>
<tr>
<td>08:00 AM</td>
<td>Contributed talk: Modeling Continuous Stochastic Processes with Dynamic Normalizing Flows</td>
</tr>
<tr>
<td>08:25 AM</td>
<td>Q&amp;A with authors of contributed talk</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Invited talk 6: Likelihood Models for Science</td>
</tr>
<tr>
<td>08:55 AM</td>
<td>Q&amp;A with Kyle Cranmer</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited talk 7: Flows in Probabilistic Modeling &amp; Inference</td>
</tr>
<tr>
<td>09:25 AM</td>
<td>Q&amp;A with Martin Jankowiak</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Contributed talk: Learning normalizing flows from Entropy-Kantorovich potentials</td>
</tr>
<tr>
<td>09:55 AM</td>
<td>Q&amp;A with authors of contributed talk</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Poster session 2</td>
</tr>
<tr>
<td>N/A</td>
<td>Poster presentation: Faster Orthogonal Parameterization with Householder Matrices</td>
</tr>
<tr>
<td>N/A</td>
<td>Poster presentation: Consistency Regularization for Variational Auto-encoders</td>
</tr>
</tbody>
</table>
Poster presentation: Stochastic Normalizing Flows

Poster presentation: Time Series Decomposition with Slow Flows

Poster presentation: Robust model training and generalisation with Studentising flows

Poster presentation: Deep Generative Video Compression with Temporal Autoregressive Transforms

Poster presentation: Normalizing Flows Across Dimensions

Poster presentation: Differentially Private Normalizing Flows for Privacy-Preserving Density Estimation

Poster presentation: Quasi-Autoregressive Residual (QuAR) Flows

Poster presentation: MoFlow: An Invertible Flow Model for Molecular Graph Generation


Poster presentation: Flow-based SVDD for anomaly detection

Poster presentation: TraDE: Transformers for Density Estimation

Poster presentation: Improving Sample Quality by Training and Sampling from Latent Energy

Poster presentation: Black-box Adversarial Example Generation with Normalizing Flows

Poster presentation: Sequential Autoregressive Flow-Based Policies

Poster presentation: A Fourier State Space Model for Bayesian ODE Filters

Poster presentation: The Power Spherical distribution

Poster presentation: Conditional Normalizing Flows for Low-Dose Computed Tomography Image Reconstruction

Poster presentation: Relative gradient optimization of the Jacobian term in unsupervised deep learning

Poster presentation: Why Normalizing Flows Fail to Detect Out-of-Distribution Data

Poster presentation: Normalizing Flows with Multi-Scale Autoregressive Priors

Poster presentation: Model-Agnostic Searches for New Physics with Normalizing Flows

Poster presentation: Scaling RBMs to High Dimensional Data with Invertible Neural Networks

Poster presentation: WeakFlow: Iterative Invertible Distribution Transformations via Weak Destructive Flows

Poster presentation: Metropolized Flow: from Invertible Flow to MCMC

Poster presentation: Woodbury Transformations for Deep Generative Flows

Link: Slack

Poster presentation: Density Deconvolution with Normalizing Flows
Joint work with Didrik Nielsen, Priyank Jaini, Emiel Hoogeboom.

This suggests a third possibility that bridges the gap between the two: a surjective map which is deterministic and surjective in one direction, and topologically preserving and can therefore not map between spaces with different topologies. This has both theoretical and numerical consequences. In the context of normalizing flows for example, the source and target density often have different topologies leading to numerically ill-posed models and training. On top of reviewing the theoretical and practical aspects of this, the talk will also cover several recent models, methods and ideas for alleviating some of these limitations.


This talk will review recent work on the representational limitations of invertible models both in the context of neural ODEs and normalizing flows. In particular, it has been shown that invertible neural networks are numerically ill-posed models and training. On top of reviewing the theoretical and practical aspects of this, the talk will also cover several recent models, methods and ideas for alleviating some of these limitations.

Abstract 13: Spotlight talk: You say Normalizing Flows I see Bayesian Networks in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, INNF 05:10 AM

[[ video ]](https://slideslive.com/38931456)


[[ video ]](https://slideslive.com/38931442)


[[ video ]](https://slideslive.com/38931480)


[[ video ]](https://slideslive.com/38931460)

Abstract 17: Spotlight talk: The Lipschitz Constant of Self-Attention in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, INNF 05:30 AM

[[ video ]](https://slideslive.com/38931447)
Abstract 18: Spotlight talk: Autoregressive flow-based causal discovery and inference in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, INNF 05:35 AM
[[ video ]](https://slideslive.com/38931450)

Volatility of GANs.


Variational inference (VI) plays an essential role in approximate Bayesian inference due to its computational efficiency and broad applicability. Crucial to the performance of VI is the selection of the associated divergence measure, as VI approximates the intractable distribution by minimizing this divergence. In this talk, I will discuss variational inference with different divergence measures first. Then, I will present a new meta-learning algorithm to learn the divergence metric suited for the task of interest, automating the design of VI methods.


Parameterizing latent variable models with deep neural networks has become a major approach to probabilistic modeling. The usual way of fitting these deep latent variable models is to use maximum likelihood. This gives rise to variational autoencoders (VAEs). They jointly learn an approximate posterior distribution over the latent variables and the model parameters by maximizing a lower bound to the log-marginal likelihood of the data. In this talk, I will present an alternative approach to fitting parameters of deep latent-variable models. The idea is to marry adversarial learning and entropy regularization. The family of models fit with this procedure is called Prescribed Generative Adversarial Networks (PresGANs). I will describe PresGANs and discuss how they generate samples with high perceptual quality while avoiding the ubiquitous mode collapse issue of GANs.

Abstract 24: Contributed talk: Modeling Continuous Stochastic Processes with Dynamic Normalizing Flows in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, INNF 08:00 AM
[[ video ]](https://slideslive.com/38930931)


Statistical inference is at the heart of the scientific method, and the likelihood function is at the heart of statistical inference. However, many scientific theories are formulated as mechanistic models that do not admit a tractable likelihood. While traditional approaches to confronting this problem may seem somewhat naive, they reveal numerous other considerations in the scientific workflow beyond the approximation error of the likelihood. I will highlight how normalizing flows and other techniques from machine learning are impacting scientific practice, discuss current challenges for state-of-the-art methods, and identify promising new directions in this line of research.


I give an overview of the many uses of flows in probabilistic modeling and inference. I focus on settings in which flows are used to speed up or otherwise improve inference (i.e. settings in which flows are not part of the model specification), including applications to Optimal Experimental Design, Hamiltonian Monte Carlo, and Likelihood-Free Inference. I conclude with a brief discussion of how flows enter into probabilistic programming language (PPL) systems and suggest research directions that are important for improved PPL integration.

Abstract 30: Contributed talk: Learning normalizing flows from Entropy-Kantorovich potentials in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, INNF 09:30 AM
[[ video ]](https://slideslive.com/38930934)

Abstract 58: Link: Poster presentations and zoom links in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, N/A

Author's availability and zoom link
https://docs.google.com/spreadsheets/d/1HJa8F0bMSIM2qQWNCJ3LO26hVmnhrFzL/whoami

Abstract 61: Link: Slack in INNF+: Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, N/A

to join the slack, please use the link
https://join.slack.com/t/innf2020/shared_invite/zt-tp5gs7l-1XAZrKGL1xtIP03Fsd5ZQQ

---

**Inductive Biases, Invariances and Generalization in Reinforcement Learning**

**Anirudh Goyal, Rosemary Nan Ke, Jane Wang, Theo Weber, Fabio Viola, Bernhard Schölkopf, Stefan Bauer**

Sat Jul 18, 03:00 AM

One proposed solution towards the goal of designing machines that can extrapolate experience across environments and tasks, are inductive biases. Providing and starting algorithms with inductive biases might help to learn invariances e.g. a causal graph structure, which in turn will allow the agent to generalize across environments and tasks. While some inductive biases are already available and correspond to common knowledge, one key requirement to learn inductive biases from data seems to be the possibility to perform and learn from interventions. This assumption is partially motivated by the accepted hypothesis in psychology about the need to experiment in order to discover causal relationships. This corresponds to an reinforcement learning environment, where the agent can discover causal factors through interventions and observing their effects.

We believe that one reason which has hampered progress on building intelligent agents is the limited availability of good inductive biases. Learning inductive biases from data is difficult since this corresponds to an interactive learning setting, which compared to classical regression or classification frameworks is far less understood e.g. even formal definitions of generalization in RL have not been developed. While Reinforcement Learning has already achieved impressive results, the sample complexity required to achieve consistently good performance is...
often prohibitively high. This has limited most RL to either games or settings where an accurate simulator is available. Another issue is that RL agents are often brittle in the face of even tiny changes to the environment (either visual or mechanistic changes) unseen in the training phase.

To build intuition for the scope of the generalization problem in RL, consider the task of training a robotic car mechanic that can diagnose and repair any problem with a car. Current methods are all insufficient in some respect -- on-policy policy gradient algorithms need to cycle through all possible broken cars on every single iteration, off-policy algorithms end up with a mess of instability due to perception and highly diverse data, and model-based methods may struggle to fully estimate a complex web of causality.

In our workshop we hope to explore research and new ideas on topics related to inductive biases, invariances and generalization, including:

- What are efficient ways to learn inductive biases from data?
- Which inductive biases are most suitable to achieve generalization?
- Can we make the problem of generalization in particular for RL more concrete and figure out standard terms for discussing the problem?
- Causality and generalization especially in RL
- Model-based RL and generalization.
- Sample Complexity in reinforcement learning.
- Can we create models that are robust visual environments, assuming all the underlying mechanics are the same. Should this count as generalization or transfer learning?
- Robustness to changes in the mechanics of the environment, such as scaling of rewards.
- Can we create a theoretical understanding of generalization in RL, and understand how it is related to the well developed ideas from statistical learning theory.
- in RL, the training data is collected by the agent and it is affected by the agent’s policy.

Therefore, the training distribution is not a fixed distribution. How does this affect how we should think about generalization?

The question of generalization in reinforcement learning is essential to the field’s future both in theory and in practice. However there are still open questions about the right way to think about generalization in RL, the right way to formalize the problem, and the most important tasks. This workshop would help to address this issue by bringing together researchers from different backgrounds to discuss these challenges.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>03:00 AM</td>
<td>Opening remarks</td>
<td></td>
</tr>
<tr>
<td>03:15 AM</td>
<td>Poster Session 1</td>
<td></td>
</tr>
<tr>
<td>04:30 AM</td>
<td>Invited talk 1 Silver</td>
<td>Silver</td>
</tr>
<tr>
<td>05:00 AM</td>
<td>QA for invited talk 1 Silver</td>
<td>Silver</td>
</tr>
<tr>
<td>05:10 AM</td>
<td>Invited talk 2 Uhler</td>
<td>Uhler</td>
</tr>
<tr>
<td>05:40 AM</td>
<td>QA for invited talk 2 Uhler</td>
<td>Uhler</td>
</tr>
<tr>
<td>05:50 AM</td>
<td>Automatic Data Augmentation for Generalization in Reinforcement Learning</td>
<td>Raileanu</td>
</tr>
</tbody>
</table>

06:15 AM Poster Session 2
07:30 AM Invited talk 3 Yang Yang
08:10 AM Invited talk 4 Bengio Bengio
08:40 AM QA for invited talk 4 Bengio
08:50 AM Off-Dynamics Reinforcement Learning: Training for Transfer with Domain Classifiers Asawa, Eysenbach
09:05 AM QA for invited talk 3 Yang Yang
09:15 AM Invited talk 5 White White
09:45 AM QA for invited talk 5 White White
10:00 AM Invited talk 6 Heess Heess
10:30 AM QA for invited talk 6 Heess Heess
10:40 AM Poster Session 3
11:55 AM Invited talk 7 Wang Wang
12:25 PM QA for invited talk 7 Wang Wang
12:35 PM Invited talk 8 Kakade Kakade
01:05 PM QA for invited talk 8 Kakade Kakade
01:15 PM Panel Discussion Session

N/A Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels Kostrikov
N/A Learning to Learn from Failures Using Replay Chen
N/A One Solution is Not All You Need: Few-Shot Extrapolation via Structured MaxEnt RL Kumar, Kumar
N/A Meta Attention Networks: Meta Learning Attention To Modulate Information Between Sparsely Interacting Recurrent Modules Madan
N/A Structure Mapping for Transferability of Causal Models Pruthi
N/A Efficient Adaptation for End-to-End Vision-Based Robotic Manipulation Sukhatme
N/A HAT: Hierarchical Alternative Training for Long Range Policy Transfer Sun, Tseng
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditioning of Reinforcement Learning Agents and its Policy Regularization Application</td>
<td>Asadulaev</td>
</tr>
<tr>
<td>Multi-Task Reinforcement Learning as a Hidden-Parameter Block MDP</td>
<td>Zhang</td>
</tr>
<tr>
<td>If MaxEnt RL is the Answer, What is the Question?</td>
<td>Eysenbach</td>
</tr>
<tr>
<td>Towards Self-Paced Context Evaluation for Contextual Reinforcement Learning</td>
<td>Eimer</td>
</tr>
<tr>
<td>Learning Invariant Representations for Reinforcement Learning without Reconstruction</td>
<td>Zhang</td>
</tr>
<tr>
<td>Learning Action Priors for Visuomotor transfer</td>
<td>Ajay</td>
</tr>
<tr>
<td>Learning Off-Policy with Online Planning</td>
<td>Sikchi</td>
</tr>
<tr>
<td>Probing Dynamic Environments with Informed Policy Regularization</td>
<td>Kamienny</td>
</tr>
<tr>
<td>Discrete Planning with End-to-end Trained Neuro-algorithmic Policies</td>
<td>Vlastelica Pogancic</td>
</tr>
<tr>
<td>Meta-Reinforcement Learning for Robotic Industrial Insertion Tasks</td>
<td>Schoettler</td>
</tr>
<tr>
<td>Nesterov Momentum Adversarial Perturbations in the Deep Reinforcement Learning Domain</td>
<td>Korkmaz</td>
</tr>
<tr>
<td>Planning to Explore via Self-Supervised World Models</td>
<td>Sekar</td>
</tr>
<tr>
<td>Spatially Structured Recurrent Modules</td>
<td>Rahaman</td>
</tr>
<tr>
<td>Evaluating Agents without Rewards</td>
<td>Hafner</td>
</tr>
<tr>
<td>Bridging Worlds in Reinforcement Learning with Model-Advantage</td>
<td>Kalyan, Modhe</td>
</tr>
<tr>
<td>Reinforcement Learning Generalization with Surprise Minimization</td>
<td>Chen</td>
</tr>
<tr>
<td>Counterfactual Data Augmentation using Locally Factored Dynamics</td>
<td>Pitis</td>
</tr>
<tr>
<td>Meta-Reinforcement Learning Robust to Distributional Shift via Model Identification and Experience Relabeling</td>
<td>Mendonca</td>
</tr>
<tr>
<td>On the Equivalence of Bi-Level Optimization and Game-Theoretic Formulations of Invariant Risk Minimization</td>
<td>Ahuja</td>
</tr>
<tr>
<td>Learning Robust Representations with Score Invariant Learning</td>
<td>Idnani</td>
</tr>
<tr>
<td>Maximum Entropy Model Rollouts: Fast Model Based Policy Optimization without Compounding Errors</td>
<td>Zhang</td>
</tr>
<tr>
<td>Exact (Then Approximate) Dynamic Programming for Deep Reinforcement Learning</td>
<td>Marklund</td>
</tr>
<tr>
<td>DisCor: Corrective Feedback in Reinforcement Learning via Distribution Correction</td>
<td>Kumar</td>
</tr>
<tr>
<td>Efficient Imitation Learning with Local Trajectory Optimization</td>
<td>Song</td>
</tr>
<tr>
<td>Learning Long-term Dependencies Using Cognitive Inductive Biases in Self-attention RNNs</td>
<td>Kerg</td>
</tr>
<tr>
<td>MOPO: Model-based Offline Policy Optimization</td>
<td>Yu</td>
</tr>
<tr>
<td>Neural Dynamic Policies for End-to-End Sensorimotor Learning</td>
<td>Gupta</td>
</tr>
<tr>
<td>Model-based Adversarial Meta-Reinforcement Learning</td>
<td>Ma, Lin</td>
</tr>
<tr>
<td>Attention Option-Critic</td>
<td>Chunduru</td>
</tr>
<tr>
<td>Fighting Copycat Agents in Behavioral Cloning From Multiple Observations</td>
<td>Wen</td>
</tr>
<tr>
<td>Towards TempoRL: Learning When to Act</td>
<td>Biedenkapp</td>
</tr>
</tbody>
</table>
N/A A Differentiable Newton Euler Algorithm for Multi-body Model Learning Lutter

N/A Long-Horizon Visual Planning with Goal-Conditioned Hierarchical Predictors Pertsch

N/A Group Equivariant Deep Reinforcement Learning Mondal

N/A Counterfactual Transfer via Inductive Bias in Clinical Settings Killian

N/A Watch your Weight Reinforcement Learning Müller

N/A PAC Imitation and Model-based Batch Learning of Contextual MDPs Nair

N/A Robust Reinforcement Learning using Adversarial Populations Vinitsky

Abstracts (55):

Abstract 3: Invited talk 1 Silver in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Silver 04:30 AM

Meta Gradient Reinforcement Learning

Abstract 5: Invited talk 2 Uhler in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Uhler 05:10 AM

Abstract 7: Automatic Data Augmentation for Generalization in Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Raileanu 05:50 AM

http://slideslive.com/38931360

Abstract 9: Invited talk 3 Yang in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Yang 07:30 AM

Abstract 10: Invited talk 4 Bengio in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Bengio 08:10 AM

Augmenting data to improve robustness â€“ a blessing or a curse?

Abstract 14: Invited talk 5 White in Inductive Biases, Invariances and Generalization in Reinforcement Learning, White 09:15 AM

A New RNN algorithm using the computational inductive bias of span independence

Abstract 16: Invited talk 6 Heess in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Heess 10:00 AM

Abstract 18: Invited talk 7 Wang in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Wang 11:55 AM

Abstract 19: Invited talk 8 Rakade in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Rakade 12:35 PM

Abstract 20: Invited talk 9 Bengio in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Bengio 13:15 AM

On the Theory of Policy Gradient Methods: Optimality, Generalization and Distribution Shift

Abstract 21: Invited talk 10 Rakade in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Rakade 12:35 PM

Abstract 22: Invited talk 11 Rakade in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Rakade 12:35 PM

Abstract 23: Invited talk 12 Rakade in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Rakade 12:35 PM

Abstract 24: Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Kostrikov N/A

http://slideslive.com/38931361

Abstract 25: Learning to Learn from Failures Using Replay in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Chen N/A

http://slideslive.com/38931328

Abstract 26: One Solution is Not All You Need: Few-Shot Extrapolation via Structured MaxEnt RL in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Kumar, Kumar N/A

http://slideslive.com/38931345

Abstract 27: Meta Attention Networks: Meta Learning Attention To Modulate Information Between Sparsely Interacting Recurrent Modules in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Madan N/A

http://slideslive.com/38931327

Abstract 28: Structure Mapping for Transferability of Causal Models in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Pruthi N/A

http://slideslive.com/38931337

Abstract 29: Efficient Adaptation for End-to-End Vision-Based Robotic Manipulation in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Sukhatme N/A

http://slideslive.com/38931340

Abstract 30: HAT: Hierarchical Alternative Training for Long Range Policy Transfer in Inductive Biases, Invariances and Generalization

http://slideslive.com/38931338
Abstract 31: Conditioning of Reinforcement Learning Agents and its Policy Regularization Application in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Asadulaev N/A
http://slideslive.com/38931436

Abstract 32: Multi-Task Reinforcement Learning as a Hidden-Parameter Block MDP in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Zhang N/A
http://slideslive.com/38931356

Abstract 33: If MaxEnt RL is the Answer, What is the Question? in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Eysenbach N/A
http://slideslive.com/38931440

Abstract 34: Towards Self-Paced Context Evaluation for Contextual Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Eimer N/A
http://slideslive.com/38931344

Abstract 35: Learning Invariant Representations for Reinforcement Learning without Reconstruction in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Zhang N/A
http://slideslive.com/38931349

Abstract 36: Learning Action Priors for Visuomotor transfer in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Ajay N/A
http://slideslive.com/38931326

Abstract 37: Learning Off-Policy with Online Planning in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Sikchi N/A
http://slideslive.com/38931439

Abstract 38: Probing Dynamic Environments with Informed Policy Regularization in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Kamienny N/A
http://slideslive.com/38931348

Abstract 39: Discrete Planning with End-to-end Trained Neuro-algorithmic Policies in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Vlastelica Pogancic N/A
http://slideslive.com/38931357

Abstract 40: Meta-Reinforcement Learning for Robotic Industrial Insertion Tasks in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Schoettler N/A
http://slideslive.com/38931362

Abstract 41: Nesterov Momentum Adversarial Perturbations in the Deep Reinforcement Learning Domain in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Korkmaz N/A
http://slideslive.com/38931437

Abstract 42: Planning to Explore via Self-Supervised World Models in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Sekar N/A
http://slideslive.com/38931334

Abstract 43: Spatially Structured Recurrent Modules in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Rahaman N/A
http://slideslive.com/38931303

Abstract 44: Evaluating Agents without Rewards in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Hafner N/A
http://slideslive.com/38931301

Abstract 45: Bridging Worlds in Reinforcement Learning with Model-Advantage in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Kalyan, Modhe N/A
http://slideslive.com/38931333

Abstract 46: Reinforcement Learning Generalization with Surprise Minimization in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Chen N/A
http://slideslive.com/38931304

Abstract 47: Counterfactual Data Augmentation using Locally Factored Dynamics in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Pitis N/A
http://slideslive.com/38931359

Abstract 48: Meta-Reinforcement Learning Robust to Distributional Shift via Model Identification and Experience Relabeling in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Mendonca N/A
http://slideslive.com/38931350

Abstract 49: On the Equivalence of Bi-Level Optimization and Game-Theoretic Formulations of Invariant Risk Minimization in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Ahuja N/A
http://slideslive.com/38931353

Abstract 50: Learning Robust Representations with Score Invariant Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning, Idnani N/A
http://slideslive.com/38931352
Abstract 51: Maximum Entropy Model Rollouts: Fast Model Based Policy Optimization without Compounding Errors in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Zhang N/A

http://slideslive.com/38931358

Abstract 52: Exact (Then Approximate) Dynamic Programming for Deep Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Marklund N/A

http://slideslive.com/38931341

Abstract 53: DisCor: Corrective Feedback in Reinforcement Learning via Distribution Correction in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Kumar N/A

http://slideslive.com/38931343

Abstract 54: Efficient Imitation Learning with Local Trajectory Optimization in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Song N/A

http://slideslive.com/38931339

Abstract 55: Learning Long-term Dependencies Using Cognitive Inductive Biases in Self-attention RNNs in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Kerg N/A

http://slideslive.com/38931335

Abstract 56: MOPO: Model-based Offline Policy Optimization in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Yu N/A

http://slideslive.com/38931302

Abstract 57: Neural Dynamic Policies for End-to-End Sensorimotor Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Gupta N/A

http://slideslive.com/38931329

Abstract 58: Model-based Adversarial Meta-Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Ma, Lin N/A

http://slideslive.com/38931355

Abstract 59: Attention Option-Critic in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Chunduru N/A

http://slideslive.com/38931346

Abstract 60: Fighting Copycat Agents in Behavioral Cloning From Multiple Observations in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Wen N/A

http://slideslive.com/38931530

Abstract 61: Towards TempoRL: Learning When to Act in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Biedenkapp N/A

http://slideslive.com/38931336

Abstract 62: A Differentiable Newton Euler Algorithm for Multi-body Model Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Lutter N/A

http://slideslive.com/38931342

Abstract 63: Long-Horizon Visual Planning with Goal-Conditioned Hierarchical Predictors in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Pertsch N/A

http://slideslive.com/38931332

Abstract 64: Group Equivariant Deep Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Mondal N/A

http://slideslive.com/38931347

Abstract 65: Counterfactual Transfer via Inductive Bias in Clinical Settings in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Killian N/A

http://slideslive.com/38931346

Abstract 66: Watch your Weight Reinforcement Learning in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Müller N/A

http://slideslive.com/38931330

Abstract 67: PAC Imitation and Model-based Batch Learning of Contextual MDPs in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Nair N/A

http://slideslive.com/38931354

Abstract 68: Robust Reinforcement Learning using Adversarial Populations in Inductive Biases, Invariances and Generalization in Reinforcement Learning. Vinitsky N/A

http://slideslive.com/38931336

Negative Dependence and Submodularity: Theory and Applications in Machine Learning

Zelda Mariet, Michal Derezinski, Mike Gartrell

Sat Jul 18, 05:00 AM

Models of negative dependence and submodularity are increasingly important in machine learning. Whether selecting training data, finding an optimal experimental design, exploring in reinforcement learning and Bayesian optimization, or designing recommender systems, selecting high-quality yet diverse items has become a core challenge. This workshop aims to bring together researchers who, using theoretical or applied techniques, leverage negative dependence and submodularity in their work. Expanding upon last year's workshop, we will highlight recent developments in the rich mathematical theory of negative dependence, cover novel critical applications, and discuss the most promising directions for future research.
### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:00 AM</td>
<td>Opening remarks</td>
<td>Mariet, Gartrell, Derezinski</td>
</tr>
<tr>
<td>05:15 AM</td>
<td>Diversity in reinforcement learning</td>
<td>Osogami</td>
</tr>
<tr>
<td>05:45 AM</td>
<td>Diversity in reinforcement learning</td>
<td></td>
</tr>
<tr>
<td>05:50 AM</td>
<td>From random matrices to kernel quadrature: how repulsiveness can speed up Monte Carlo integration</td>
<td>Bardenet</td>
</tr>
<tr>
<td>06:20 AM</td>
<td>From random matrices to kernel quadrature: how repulsiveness can speed up Monte Carlo integration</td>
<td></td>
</tr>
<tr>
<td>06:25 AM</td>
<td>Ensemble Kernel Methods, Implicit Regularization and Determinantal Point Processes</td>
<td>Schreurs, Fanuel, Suykens</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Scaling DPP MAP Inference</td>
<td>Gillenwater</td>
</tr>
<tr>
<td>07:40 AM</td>
<td>Scaling DPP MAP Inference Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>07:45 AM</td>
<td>Negative Dependence and Sampling</td>
<td>Jegelka</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Negative Dependence and Sampling</td>
<td></td>
</tr>
<tr>
<td>08:20 AM</td>
<td>Mode Finding for SLC Distributions via Regularized Submodular Maximization</td>
<td>Kazemi, Karbasi, Feldman</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Poster session</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Exponentially Faster Algorithms for Machine Learning</td>
<td>Singer</td>
</tr>
<tr>
<td>10:45 AM</td>
<td>Exponentially Faster Algorithms for Machine Learning</td>
<td></td>
</tr>
<tr>
<td>10:50 AM</td>
<td>Searching for Diverse Biological Sequences</td>
<td>Colwell</td>
</tr>
<tr>
<td>11:20 AM</td>
<td>Searching for Diverse Biological Sequences</td>
<td></td>
</tr>
<tr>
<td>11:25 AM</td>
<td>Constrained Maximization of Lattice Submodular Functions</td>
<td>Sahin, Buhmann, Krause</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Determinantal Point Processes in Randomized Numerical Linear Algebra</td>
<td>Mahoney</td>
</tr>
<tr>
<td>12:40 PM</td>
<td>Determinantal Point Processes in Randomized Numerical Linear Algebra</td>
<td></td>
</tr>
</tbody>
</table>

### Abstracts (12):

**Abstract 2: Diversity in reinforcement learning in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Osogami 05:15 AM**

Reinforcement learning has seen major success in games and other artificial environments, but its applications in industries and real life are still limited. This limited applicability is partly due to the requirement of the large amount of the training data that needs to be collected through trial and error as well as the difficulty in effectively dealing with multiple or many agents. Diversity and negative dependence are a promising approach to resolve some of the major challenges in today’s reinforcement learning and have gained increasing attention in recent years. In this talk, we will briefly review some of the approaches to introducing diversity in reinforcement learning with a focus on the use of determinantal point processes for effective multi-agent reinforcement learning.

**Abstract 4: From random matrices to kernel quadrature: how repulsiveness can speed up Monte Carlo integration in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Bardenet 05:50 AM**

Eigenvalues of many models of random matrices tend to repel each other. They sometimes repel so much that sample averages over these eigenvalues converge much faster than if the eigenvalues were i.i.d. I will first show how to transform this kind of result into a generic importance sampler with mean square error decreasing as $SN^{\frac{1}{d}}$. This result crucially depends on a repulsive point process for the integration nodes, called an orthogonal polynomial ensemble, itself a particular case of determinantal point process (DPP). With more assumptions on the integrand and more general DPPs, I will then show how to obtain faster Monte Carlo rates. Further generalizing to mixtures of DPPs, I will finally show how to obtain tight integration rates for integrands in a large class of reproducing kernel Hilbert spaces. This last result involves a continuous equivalent to volume sampling, a discrete point process of recent interest in numerical linear algebra. This talk is intended to connect to a few other talks of the day, and is based on the following papers:


**Abstract 6: Ensemble Kernel Methods, Implicit Regularization and Determinantal Point Processes in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Schreurs, Fanuel, Suykens 06:25 AM**

By using the framework of Determinantal Point Processes (DPPs), some theoretical results concerning the interplay between diversity and
regularization can be obtained. In this paper we show that sampling subsets with kDPPs results in implicit regularization in the context of ridgeless Kernel Regression. Furthermore, we leverage the common setup of state-of-the-art DPP algorithms to sample multiple small subsets and use them in an ensemble of ridgeless regressions. Our first empirical results indicate that ensemble of ridgeless regressors can be interesting to use for datasets including redundant information.

Abstract 7: Scaling DPP MAP Inference in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Gillenwater 07:10 AM

DPP MAP inference, the problem of finding the highest probability set under the distribution defined by a DPP, is of practical importance for a variety of areas such as recommender systems, active learning, and data compression. Unfortunately, finding the exact MAP solution is NP-hard. Often though, the standard greedy submodular maximization algorithm works well in practice for approximating the solution. In this talk, we discuss ways to speed up this simple greedy algorithm, as well as slower, but more accurate alternatives to it. We also discuss how to scale greedy for customized DPPs, where we want to solve the MAP problem multiple times with different weightings of item features. We conclude with a brief note on the complexity of MAP for nonsymmetric DPPs, where we show that greedy scales fairly well if we assume a particular kernel decomposition.

Abstract 9: Negative Dependence and Sampling in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Jegelka 07:45 AM

Probability distributions with strong notions of negative dependence arise in various forms in machine learning. Examples include diversity-inducing probabilistic models, interpretability, exploration and active learning, and randomized algorithms. While, perhaps surprisingly, being more delicate than its positive counterpart, negative dependence enjoys rich mathematical connections and properties that offer a promising toolbox for machine learning. In this talk, I will summarize some recently important notions of negative dependence, and their implications for sampling algorithms. These results exploit connections to the geometry of polynomials, log concavity, and submodular optimization. We will conclude with an example application of sampling minibatches for optimization.

Abstract 11: Mode Finding for SLC Distributions via Regularized Submodular Maximization in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Kazemi, Karbasi, Feldman 08:20 AM

In this paper, we propose scalable methods for maximizing a regularized submodular function $f = g - \ell$ expressed as the difference between a monotone submodular function $g$ and a modular function $\ell$. Indeed, submodularity is inherently related to the notions of diversity, coverage, and representativeness. In particular, finding the mode (i.e., the most likely configuration) of many popular probabilistic models of diversity, such as determinantal point processes, submodular probabilistic models, and strongly log-concave distributions, involves maximization of (regularized) submodular functions. Since a regularized function $f$ can potentially take on negative values, the classic theory of submodular maximization, which heavily relies on the non-negativity assumption of submodular functions, may not be applicable. To circumvent this challenge, we develop the first streaming and distributed algorithm for maximizing regularized submodular functions. Furthermore, we show that how can we find the mode of a strongly log-concave (SLC) distribution by regularized submodular maximization.

Abstract 12: Poster session in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, 09:30 AM

### Submodular maximization

- Online Algorithms for Budget-Constrained DR-Submodular Maximization, "Omid Sadeghi, Reza Eghbali, Maryam Fazel"
- Constrained Maximization of Lattice Submodular Functions, "Aytunc Sahin, Joachim Buhmann, Andreas Krause"
- Mode Finding for SLC Distributions via Regularized Submodular Maximization, "Ehsan Kazemi, Shervin Minaee, Moran Feldman, Amin Karbasi"

### Determinantal point processes

- MO-PaDGAN: Generating Diverse Designs with Multivariate Performance Enhancement, "Wei Chen, Faez Ahmed"
- On the Relationship Between Probabilistic Circuits and Determinantal Point Processes, "Honghua Zhang, Steven J Holtzen, Guy Van den Broeck"
- Ensemble Kernel Methods, Implicit Regularization and Determinantal Point Processes, "Joachim Schreurs, Michaël Fanuel, Johan Suykens"

### Negative dependence for inference and bipartite matching

- DisARM: An Antithetic Gradient Estimator for Binary Latent Variables, "Zhe Dong, Andriy Mnih, George Tucker"
- On Diverse Bipartite b-Matching, "Saba Ahmadi, Faez Ahmed, John P Dickerson, Mark Fuge, Samir Khuller"
- Negative Dependence Tightens Variational Bounds, "Pierre-Alexandre Mattei, Jes Freilisen"


In this talk I'll describe a novel approach that yields algorithms whose parallel running time is exponentially faster than any algorithm previously known for a broad range of machine learning applications. The algorithms are designed for submodular function maximization which is the algorithmic engine behind applications such as clustering, network analysis, feature selection, Bayesian inference, ranking, speech and document summarization, recommendation systems, hyperparameter tuning, and many others. Since applications of submodular functions are ubiquitous across machine learning and data sets become larger, there is consistent demand for accelerating submodular optimization. The approach we describe yields simple algorithms whose parallel runtime is logarithmic in the size of the data rather than linear. I'll introduce the frameworks we recently developed and present experimental results from various application domains.
Abstract 15: Searching for Diverse Biological Sequences in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Colwell 10:50 AM

A central challenge in biotechnology is to be able to predict functional properties of a protein from its sequence, and thus (i) discover new proteins with specific functionality and (ii) better understand the functional effect of genomic mutations. Experimental breakthroughs in our ability to read and write DNA allows data on the relationship between sequence and function to be rapidly acquired. This data can be used to train and validate machine learning models that predict protein function from sequence. However, the cost and latency of wet-lab experiments requires methods that find good sequences in few experimental rounds, where each round contains large batches of sequence designs. In this setting, model-based optimization allows us to take advantage of sample inefficient methods to find diverse optimal sequence candidates to be tested in the wet-lab. These requirements are illustrated by a collaboration that involves the design and experimental validation of AAV capsid protein variants that assemble integral capsids and package their genome, for use in gene therapy applications.

Abstract 17: Constrained Maximization of Lattice Submodular Functions in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Sahin, Buhmann, Krause 11:25 AM

Submodular optimization over the integer lattice has many applications in machine learning. Although the constrained maximization of submodular functions with coordinate-wise concavity (also called \{em DR-Submodular\} functions) is well studied, the maximization of \{em general\} lattice submodular functions is considerably more challenging. In this work, we first show that we can optimize lattice submodular functions subject to a discrete (integer) polymatroid constraint using a recently proposed extension, called the Generalized Multilinear Extension. Then, we establish a bound on the rounding error for the discrete polymatroid constraint, which depends on the \mbox{\textit{a distance\textit{}}} between the lattice submodular function to a DR-Submodular function. Lastly, we demonstrate the effectiveness of our algorithm on a Bayesian experimental design problem with repetition and a concave cost.

Abstract 18: Determinantal Point Processes in Randomized Numerical Linear Algebra in Negative Dependence and Submodularity: Theory and Applications in Machine Learning, Mahoney 12:10 PM

Randomized Numerical Linear Algebra (RandNLA) is an area which uses randomness, most notably random sampling and random projection methods, to develop improved algorithms for ubiquitous matrix problems, such as those that arise in scientific computing, data science, machine learning, etc. A seemingly different topic, but one which has a long history in pure and applied mathematics, is that of Determinantal Point Processes (DPPs), which are stochastic point processes, the probability distribution of which is characterized by sub-determinants of some matrix. Recent work has uncovered deep and fruitful connections between DPPs and RandNLA. For example, random sampling with a DPP leads to new kinds of unbiased estimators for classical RandNLA tasks, enabling more refined statistical and inferential understanding of RandNLA algorithms; a DPP is, in some sense, an optimal randomized method for many RandNLA problems; and a standard RandNLA technique, called leverage score sampling, can be derived as the marginal distribution of a DPP. This work will be reviewed, as will recent algorithmic developments, illustrating that, while not quite as efficient as simply applying a random projection, these DPP-based algorithms are only moderately more expensive.


Scaling probabilistic models to large realistic problems and datasets is a key challenge in machine learning. Central to this effort is the development of tractable probabilistic models (TPMs): models whose structure guarantees efficient probabilistic inference algorithms. The current landscape of TPMs is fragmented: there exist various kinds of TPMs with different strengths and weaknesses. Two of the most prominent classes of TPMs are determinantal point processes (DPPs) and probabilistic circuits (PCs). This paper provides the first systematic study of their relationship. We propose a unified analysis and shared language for discussing DPPs and PCs. Then we establish theoretical barriers for the unification of these two families, and prove that there are cases where DPPs have no compact representation as a class of PCs. We close with a perspective on the central problem of unifying these models.

Federated Learning for User Privacy and Data Confidentiality

**Nathalie Baracaldo, Olivia Choudhury, Gauri Joshi, Ramesh Raskar, Shiqiang Wang, Han Yu**

**Sat Jul 18, 05:45 AM**

Training machine learning models in a centralized fashion often faces significant challenges due to regulatory and privacy concerns in real-world use cases. These include distributed training data, computational resources to create and maintain a central data repository, and regulatory guidelines (GDPR, HIPAA) that restrict sharing sensitive data. Federated learning (FL) is a new paradigm in machine learning that can mitigate these challenges by training a global model using distributed data, without the need for data sharing. The extensive application of machine learning to analyze and draw insight from real-world, distributed, and sensitive data necessitates familiarization with and adoption of this relevant and timely topic among the scientific community. Despite the advantages of federated learning, and its successful application in certain industry-based cases, this field is still in its infancy due to new challenges that are imposed by limited visibility of the training data, potential lack of trust among participants training a single model, potential privacy inferences, and in some cases, limited or unreliable connectivity. The goal of this workshop is to bring together researchers and practitioners interested in FL. This day-long event will facilitate interaction among students, scholars, and industry professionals from around the world to understand the topic, identify technical challenges, and discuss potential solutions. This will lead to an overall advancement of FL and its impact in the community.

For a detailed workshop schedule, please visit: http://federated-learning.org/fi-icml-2020/

**Workshop date: July 18, 2020 (Saturday)**
Starting at 9 am in US Eastern Daylight Time, https://time.is/EDT

**Schedule**
05:45 AM Arrival (Presenters should connect and test the system)

06:00 AM Opening remarks Baracaldo, Choudhury, Joshi, Raskar, Wang, Yu

Keynote Session 1: Balancing Efficiency and Security in Federated Learning, by Qiang Yang (WeBank)

06:10 AM Keynote Session 1: Balancing Efficiency and Security in Federated Learning, by Qiang Yang (WeBank)

06:35 AM Technical Talks Session 1 Singh, Rieger, Høegh, Lu, Jeong

06:40 AM Keynote Session 2: Federated Learning in Enterprise Settings, by Rania Khalaf (IBM Research)

07:25 AM Break (Presenters should connect and test the system)

07:40 AM Keynote Session 2: Federated Learning in Enterprise Settings, by Rania Khalaf (IBM Research)

08:05 AM Lightning Talks Session 1 Yang, Navia-Vázquez, Li, Ono, Liu, Sun, Aspodeh, Hwang, Menuet

08:35 AM Poster Session 1

09:05 AM Lunch (Presenters should connect and test the system 15 minutes before the next session starts)

Keynote Session 3: Federated Learning Applications in Alexa, by Shiv Vitaladevuni (Amazon Alexa)

10:20 AM Keynote Session 3: Federated Learning Applications in Alexa, by Shiv Vitaladevuni (Amazon Alexa)

10:45 AM Technical Talks Session 2 So, Liu, Yuan, Pillutla, Barnes, Yousefpour, Kadhe

12:10 PM Break (Presenters should connect and test the system)

Keynote Session 4: The Shuffle Model and Federated Learning, by Ilya Mironov (Facebook)

12:25 PM Keynote Session 4: The Shuffle Model and Federated Learning, by Ilya Mironov (Facebook)

12:50 PM Lightning Talks Session 2 Chung, Prakash, Khodak, Rahman, Mugunthan, zhang, Hosseini

01:15 PM Poster Session 2

01:45 PM Break (Presenters should connect and test the system)

02:00 PM Keynote Session 5: Advances and Open Problems in Federated Learning, by Brendan McMahan (Google)

02:25 PM Closing remarks Baracaldo, Choudhury, Joshi, Raskar, Wang, Yu

Abstracts (16):

Abstract 1: Arrival (Presenters should connect and test the system) in Federated Learning for User Privacy and Data Confidentiality, 05:45 AM

Presenters of Keynote Session 1 and Technical Talks Session 1 please connect to the main Zoom room of this workshop, to make sure that everything works well.

Abstract 3: Keynote Session 1: Balancing Efficiency and Security in Federated Learning, by Qiang Yang (WeBank) in Federated Learning for User Privacy and Data Confidentiality, Yang 06:10 AM

Abstract: Federated learning systems need to balance the efficiency and security of machine learning algorithms while maintaining model accuracy. In this talk we discuss this trade-off in two settings. One is when two collaborating organisations wish to transfer the knowledge from one to another via a federated learning framework. We present a federated transfer learning algorithm to both improve the security and the performance while preserving privacy. Another case is when one exploits differential privacy in a federated learning framework to ensure efficiency, but this may cause security degradation. To solve the problem, we employ a dual-headed network architecture that guarantees training data privacy by exerting secret gradient perturbations to original gradients, while maintaining high performance of the global shared model. We find that the combination of secret-public networks provides a preferable alternative to DP-based mechanisms in federated learning applications.

Biography: Qiang Yang is Chief Artificial Intelligence Officer of WeBank and Chair Professor of CSE Department of Hong Kong Univ. of Sci. and Tech. He is the Conference Chair of AAAI-21, President of Hong Kong Society of Artificial Intelligence and Robotics(HKSAIR) and a former President of IJCAI (2017-2019). He is a fellow of AAAI, ACM, IEEE and AAAS. His research interests include transfer learning and federated learning. He is the founding EiC of two journals: IEEE Transactions on Big Data and ACM Transactions on Intelligent Systems and Technology.

Abstract 4: Technical Talks Session 1 in Federated Learning for User Privacy and Data Confidentiality, Singh, Rieger, Høegh, Lu, Jeong 06:35 AM

1. Wonyong Jeong, Jaehong Yoon, Eunho Yang and Sung Ju Hwang. Federated Semi-Supervised Learning with Inter-Client Consistency
2. Ishika Singh, Haoyi Zhou, Kunlin Yang, Meng Ding, Bill Lin and Pengtao Xie. Differentially-private Federated Neural Architecture Search
3. Laura Rieger, Rasmus Malik Thaarup Høegh and Lars Kai Hansen. Client Adaptation improves Federated Learning with Simulated Non-IID Clients
4. Hanlin Lu, Changchang Liu, Ting He, Shiqiang Wang and Kevin S. Chan. Sharing Models or Coresets: A Study based on Membership
Inference Attack

Abstract 5: Break (Presenters should connect and test the system) in Federated Learning for User Privacy and Data Confidentiality, 07:25 AM

Presenters of Keynote Session 2 and Lightning Talks Session 1 please connect to the main Zoom room of this workshop, to make sure that everything works well.

Abstract 6: Keynote Session 2: Federated Learning in Enterprise Settings, by Rania Khalaf (IBM Research) in Federated Learning for User Privacy and Data Confidentiality, Khalaf 07:40 AM

Abstract: Federated learning in consumer scenarios has garnered a lot of interest. However, its application in large enterprises brings to bear additional needs and guarantees. In this talk, I will highlight key drivers for federated learning in enterprises, illustrate representative use cases, and summarize the requirements for a platform that can support it. I will then present the newly released IBM Federated Learning framework (git, white paper) and show how it can be used and extended by researchers. Finally, I will highlight recent advances in federated learning and privacy from IBM Research.

Biography: Rania Khalaf is the Director of AI Platforms and Runtimes at IBM Research where she leads teams pushing the envelope in AI platforms to make creating AI models and applications easy, fast, and safe for data scientists and developers. Her multi-disciplinary teams tackle key problems at the intersection of core AI, distributed systems, human computer interaction and cloud computing. Prior to this role, Rania was Director of Cloud Platform, Programming Models and Runtimes. Rania serves as a Judge for the MIT Solve AI for Humanity Prize, on the Leadership Challenge Group for MIT Solve’s Learning for Girls and Women Challenge and on the Advisory Board of the Hariri Institute for Computing at Boston University. She has received several Outstanding Technical Innovation awards for major impact to the field of computer science and was a finalist for the 2019 MassTLC CTO of the Year award.

Abstract 7: Lightning Talks Session 1 in Federated Learning for User Privacy and Data Confidentiality, Yang, Navia-Vázquez, Li, Ono, Liu, Sun, Asoodeh, Hwang, Menuet 08:05 AM

1. Zhaoxui Yang, Mingzhe Chen, Walid Saad, Choong Seon Hong, Mohammad Shikh-Bahaei, H. Vincent Poor and Shuguang Cui. Delay Minimization for Federated Learning Over Wireless Communication Networks
4. Hajeon Ono and Tsubasa Takahashi. Locally Private Distributed Reinforcement Learning
5. Yang Liu, Zhihao Yi and Tianjian Chen. Defending backdoor attacks in feature-partitioned collaborative learning
6. Tianyi Chen, Xiao Jin, Yuejiao Sun and Wotao Yin, VAFL: a Method of Vertical Asynchronous Federated Learning
7. Shahab Asoodeh and Flavio Calmon. Differentially Private Federated Learning: An Information-Theoretic Perspective
9. Myungjae Shin, Chihoon Hwang, Joongheon Kim, Jihong Park, Mehdi Bennis and Seong-Lyun Kim. XOR Mixup: Privacy-Preserving Data Augmentation for One-Shot Federated Learning

Abstract 8: Poster Session 1 in Federated Learning for User Privacy and Data Confidentiality, 08:35 AM

Poster session with presenters of Lightning Talks Session 1. Individual Zoom links will be provided separately.

Abstract 9: Lunch (Presenters should connect and test the system 15 minutes before the next session starts) in Federated Learning for User Privacy and Data Confidentiality, 09:05 AM

Presenters of Keynote Session 3 and Technical Talks Session 2 please connect to the main Zoom room of this workshop at 1:05 pm EDT, to make sure that everything works well.

Abstract 10: Keynote Session 3: Federated Learning Applications in Alexa, by Shiv Vitaladevuni (Amazon Alexa) in Federated Learning for User Privacy and Data Confidentiality, Vitaladevuni 10:20 AM

Abstract: Alexa is a virtual assistant AI technology launched by Amazon in 2014. One of key enabling technologies is wkward, which allows users to interact with Alexa devices hands-free via voice. We present some of the unique ML challenges posed in wkward, and how Federated Learning can be used to address them. We also present some considerations when bringing Federated Learning to consumer grade, embedded applications.

Biography: Shiv Vitaladevuni is a Senior Manager in Machine Learning at Amazon Alexa, focusing on R&D for Alexa family of devices such as Echo, Dot, FireTV, etc. At Amazon, Shiv leads a team of scientists and engineers inventing embedded speechand ML products used by millions of Alexa customers across all Alexa devices, around the globe. His team conducts research in areas such as Federated ML, Large scale semi/unsupervised learning, User diversity and fairness in ML, Speaker adaptation and personalization,memory efficient deep learning models, etc. Prior to Amazon, Shiv worked on video and text document analysis at Raytheon BBN Technologies, and bio-medical image analysis at Howard Hughes Medical Institute.

Abstract 11: Technical Talks Session 2 in Federated Learning for User Privacy and Data Confidentiality, So, Liu, Yuan, Pillutla, Barnes, Yousefpour, Kadhe 10:45 AM

3. Honglin Yuan and Tengyu Ma. Federated Accelerated Stochastic Gradient Descent
4. Krishna Pillutla, Sham Kakade and Zaid Harchaoui. Robust Aggregation for Federated Learning
5. Leighton Pate Barnes, Huseyn A. Inan, Berivan Isik and Ayfer Oszur. rTop-k: A Statistical Estimation Approach to Distributed SGD
6. Ashkan Yousefpour, Brian Nguyen, Siddartha Devic, Guanhua Wang, Abdul Rahman Kreidieh, Hans Lobel, Alexandre Bayen and Jason Jue. ResilNet: Failure-Resilient Inference in Distributed Neural Networks
Poster session with presenters of Lightning Talks Session 2. Individual Zoom links will be provided separately.

Abstract 16: **Break (Presenters should connect and test the system) in Federated Learning for User Privacy and Data Confidentiality**, 01:45 PM

Presenter of Keynote Session 5 please connect to the main Zoom room of this workshop, to make sure that everything works well.

Abstract 17: **Keynote Session 5: Advances and Open Problems in Federated Learning, by Brendan McMahan (Google) in Federated Learning for User Privacy and Data Confidentiality**, McMahan 02:00 PM

Abstract: Motivated by the explosive growth in federated learning research, 22 Google researchers and 36 academics from 24 institutions collaborated on a paper titled *Advances and Open Problems in Federated Learning*. In this talk, I will survey some of the main themes from the paper, particularly the defining characteristics and challenges of different FL settings. I will then briefly discuss some of the ways FL increasingly powers Google products, and also highlight several exciting FL research results from Google.

**Biography:** Brendan McMahan is a research scientist at Google, where he leads efforts on decentralized and privacy-preserving machine learning. His team pioneered the concept of federated learning, and continues to push the boundaries of what is possible when working with decentralized data using privacy-preserving machine learning. Previous experiences include working on Google’s key product features, including search ranking, and on Google’s artificial intelligence (AI) research. McMahan has published over 50 papers, 12 patents granted, and a book. He has been an NIH expert consultant and a Sloan Research Fellow. McMahan received his Ph.D. in computer science from Carnegie Mellon University.

---

**Machine Learning for Global Health**

**Danielle Belgrave, Stephanie Hyland, Charles Onu, Nicholas Furnham, Ernest Mwebaze, Neil Lawrence**

**Sat Jul 18, 05:45 AM**

Machine learning is increasingly being applied to problems in the healthcare domain. However, there is a risk that the development of machine learning models for improving health remain focused within areas and diseases which are more economically incentivised and resourced. This presents the risk that as research and technological entities aim to develop machine-learning-assisted consumer healthcare devices, or bespoke algorithms for their populations within a certain geographical region, that the challenges of healthcare in resource-constrained settings will be overlooked. The predominant research focus of machine learning for healthcare in the economically advantaged world means that there is a skew in our current knowledge of how machine learning can be used to improve health on a more global scale for everyone. This workshop aims to draw attention to the ways that machine learning can be used for problems in global health, and to promote research on problems outside high-resource environments.

**Schedule**

---

**05:45 AM Opening Remarks**
**Intended Use: A human-centered approach to developing ML applications for clinical practice**

**06:00 AM**
- Panel

**06:30 AM**
- AI-augmented genomic pathogen surveillance - promises and pitfalls

**07:00 AM**
- Coffee Break

**07:35 AM**
- An Unsupervised Learning Approach to Mitigate the Risk of Polio Recurrence in India

**07:45 AM**
- Anonymous Survey System and Methodology to Enable COVID-19 Surveillance

**07:55 AM**
- Prediction of neonatal mortality in Sub-Saharan African countries using data-level linkage of multiple surveys

**08:05 AM**
- Contributed Talks Q&A + Panel

**08:20 AM**
- Posters

**09:20 AM**
- Lunch

**10:30 AM**
- Machine Learning and Epidemiology

**11:00 AM**
- The Million Death Study and Systems for the Early Detection and Prevention of Infant Mortality in India

**11:30 AM**
- Panel Afternoon

**11:50 AM**
- Poster Session 2

**12:50 PM**
- Using Machine Learning to Analyze and Provide Real-Time Access to all Published Clinical Trial Reports

**01:00 PM**
- Automatic semantic segmentation for prediction of tuberculosis using lens-free microscopy images

**01:10 PM**
- Kernel-based antimicrobial resistance prediction from MALDI-TOF mass spectra

**01:20 PM**
- Contributed Talks Q&A + Panel

**01:40 PM**
- Breakout Session
Abstract 12: Machine Learning and Epidemiology in Machine Learning for Global Health, Nsoesie 10:30 AM

* A scoping review on mental health and machine learning in low-resource settings. [Join Zoom](https://us02web.zoom.us/j/86540452270?pwd=R0ljWm9rUGYy29rbXrQmSmJ1d012dFFnZz09)

* Synthesizing Wearable ECG Data for Future Deep Space Missions. [Join Zoom](https://us02web.zoom.us/j/89785420495?pwd=ZmxoSGtvPVC92MzhrR1d2ZmF6SnZAZz09)

* Using Capsule Neural Network to detect Tuberculosis in microscopic images. [Join Zoom](https://us02web.zoom.us/j/82635681248?pwd=Z2xGV104RnoyL3MWhnRlSFBNbGV0b19)

* Design Considerations for High Impact, Automated Echocardiogram Analysis. [Join Zoom](https://us02web.zoom.us/j/83881388045?pwd=ZnpVSDhzZMkFHMTQyWWh0dVd3V3gwZz09)

* A novel approach for predicting epidemiological forecasting parameters based on real-time signals and Data Assimilation. [Join Zoom](https://us02web.zoom.us/j/85061906660?pwd=RG1qUkgyU3dROXVmdldlSE95Rl03Ny9)

* Gene Expression Imputation with Generative Adversarial Imputation Nets. [Join Zoom](https://us02web.zoom.us/j/83373254666?pwd=WGhqdf0FrdFMaBPUeJKANCNdK)

* Simulation-Based Inference for Global Health Decisions. [Join Zoom](https://us02web.zoom.us/j/84273319614?pwd=WT16dFVRUc1RXZWEWE9D29xXVYUT09)

* Towards uncertainty representations for decision support system patient referrals in healthcare contexts. [Join Zoom](https://us02web.zoom.us/j/83570787923?pwd=YTZESVpjNVRNaTRHVXU9mQjZzFudz09)

* SIRNet: Understanding Social Distancing Measures with Hybrid Neural Network Model for COVID-19 Infectious Spread. [Join Zoom](https://us02web.zoom.us/j/8183003221?pwd=U1dcFdZTko0umxENHnY2k3Zz09)

* Non-Pharmaceutical Intervention Discovery with Topic Modeling. [Join Zoom](https://us02web.zoom.us/j/83774245353?pwd=YTMneGgyS3hieG03MERTVWhTb3pCUT09)

* Sequential Decision Making in Resource Constrained Global Health Settings. [Join Zoom](https://us02web.zoom.us/j/88371066221?pwd=MS9QjFJMVV3dn0Z5l5USUFmZWRzaW10)

* Measuring the Role of Income Shocks on the Health of Low-Income Families. [Join Zoom](https://us02web.zoom.us/j/86807125362?pwd=aTFjeR6bUw5RGVVBTRU1c2lDd3lnZz09)

* PET-guided Attention Network for Segmentation of Lung Tumors from PET/CT images that accounts for missing PET images. [Join Zoom](https://us02web.zoom.us/j/84878071921?pwd=by82OWNkbktTDVBPUw14Ym5cZz09)

* The Causal Effect of Stay-At-Home Orders: a synthetic control study in the San Francisco Bay Area. [Join Zoom](https://us02web.zoom.us/j/87555800005?pwd=NIpUem5RMzJsUVVPdGd4b3Vz09)

* An Unsupervised Learning Approach to Mitigate the Risk of Polio Recurrence in India. [Join Zoom](https://us02web.zoom.us/j/88950642737?pwd=dElSV0pEbFN5d1Rstdkw3bBtZVlDUT09)

Abstract 13: The Million Death Study and Systems for the Early Detection and Prevention of Infant Mortality in India in Machine Learning for Global Health, Jha 11:00 AM

* Watch talk here:

Abstract 14: Panel Afternoon in Machine Learning for Global Health, 11:30 AM

* Watch talk here:

Abstract 15: Poster Session 2 in Machine Learning for Global Health, 11:50 AM

* Watch talk here:

Abstract 16: Using Machine Learning to Analyze and Provide Real-Time Access to all Published Clinical Trial Reports in Machine Learning for Global Health, Marshall 12:50 PM
Bridge Between Perception and Reasoning: Graph Neural Networks & Beyond

Jian Tang, Le Song, Jure Leskovec, Renjie Liao, Yujia Li, Sanja Fidler, Richard Zemel, Russ Salakhutdinov

Sat Jul 18, 05:50 AM

Deep learning has achieved great success in a variety of tasks such as recognizing objects in images, predicting the sentiment of sentences, or image/speech synthesis by training on a large-amount of data. However, most existing success are mainly focusing on perceptual tasks, which is also known as System I intelligence. In real world, many complicated tasks, such as autonomous driving, public policy decision making, and multi-hop question answering, require understanding the relationship between high-level variables in the data to perform logical reasoning, which is known as System II intelligence. Integrating system I and II intelligence lies in the core of artificial intelligence and machine learning.

Graph is an important structure for System II intelligence, with the universal representation ability to capture the relationship between different variables, and support interpretability, causality, and transferability / inductive generalization. Traditional logic and symbolic reasoning over graphs has relied on methods and tools which are very different from deep learning models, such Prolog language, SMT solvers, constrained optimization and discrete algorithms. Is such a methodology separation between System I and System II intelligence necessary? How to build a flexible, effective and efficient bridge to smoothly connect these two systems, and create higher order artificial intelligence?

Graph neural networks, have emerged as the tool of choice for graph representation learning, which has led to impressive progress in many classification and regression problems such as chemical synthesis, 3D-vision, recommender systems and social network analysis. However, prediction and classification tasks can be very different from logic/symbolic reasoning.

Bits and pieces of evidence can be gleaned from recent literature, suggesting graph neural networks may be a general tool to make such a connection. For example, \cite{silver2019few,alet2019graph} combined logic reasoning with reinforcement learning. How do these alternative methods compare with graph neural networks for interacting systems, 
\cite{kipf2018neural} used graph neural networks for reasoning on knowledge graphs with graph neural networks, and 
\cite{kipf2018relational,barcelo2019logical} viewed graph neural networks as tools to incorporate explicitly logic reasoning bias. 
\cite{kipf2018neural} used graph neural network to reason about interacting systems,
\cite{yoon2018inference,zhang2020efficient} used neural networks for logic and probabilistic inference, \cite{hu2019language} used graph neural networks for reasoning on scene graphs for visual question reasoning, \cite{qu2019probabilistic} studied reasoning on knowledge graphs with graph neural networks, and 
\cite{khalil2017learning,xu2018powerful,velickovic2019neural,sato2019approximation} used graph neural networks for discrete graph algorithms. However, there can still be a long way to go for a satisfactory and definite answers on the ability of graph neural networks for automatically discovering logic rules, and conducting long-range multi-step complex reasoning in combination with perception inputs such as language, vision, spatial and temporal variation.

Can graph neural networks be the key bridge to connect System I and System II intelligence? Are there other more flexible, effective and efficient alternatives? For instance, \cite{wang2019satnet} combined max satisfiability solver with deep learning,
\cite{manhaeve2018deepproblog} combined directed graphical and Problog with deep learning, \cite{skryagin2020splog} combined sum product network with deep learning, 
\cite{silver2019few,alet2019graph} combined logic reasoning with reinforcement learning. How do these alternative methods compare with graph neural networks for being a bridge?

The goal of this workshop is to bring researchers from previously separate fields, such as deep learning, logic/symbolic reasoning, statistical relational learning, and graph algorithms, into a common roof to discuss this potential interface and integration between System I and System intelligence. By providing a venue for the confluence of new advances in theoretical foundations, models and algorithms, as well as empirical discoveries, new benchmarks and impactful applications,
### ICML 2020 Workshop book

#### Invited Talk 5: Ferran Alet (Q&A)

**01:30 PM**  
Ferran Alet

#### Invited Talk 6: Kristian Kersting

**01:40 PM**  
Kristian Kersting

**02:10 PM**  
Kristian Kersting (Q&A)

#### Concluding Remarks

**02:20 PM**

---

### Abstracts (2):

**Abstract 10:** **Morning Poster Session in Bridge Between Perception and Reasoning: Graph Neural Networks & Beyond, 07:40 AM**

Jinghan Shi et al. Heterogeneous Graph Neural Network for Recommendation.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 1)

Yangyang Hu et al. Enhancing Neural Mathematical Reasoning by Abductive Combination with Symbolic Library  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 2)

Binbin Hu et al. KGNN: Distributed Framework for Graph Neural Knowledge Representation.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 3)

Hao Tang et al. Towards Scale-Invariant Graph-related Problem Solving by Iterative Homogeneous Graph Neural Networks.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 4)

Giuseppe Futia et al. Modeling the semantics of data sources with graph neural networks  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 5)

Laetitia Teodorescu et al. SpatialSim: Recognizing Spatial Configurations of Objects with Graph Neural Networks.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 6)

Binghong Chen et al. Learning Retrosynthetic Planning with Chemical Reasoning  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 7)

Maxwell Crouse et al. Neural Analogical Matching.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 8)

Yuta Kawachi et al. End-to-end permutation learning with Hungarian algorithm.  
https://us02web.zoom.us/j/82916453037?pwd=SGlwN3hnM0RKaGJUZ0lZSHA1Kzh5QT09 (Breakout Room 9)
Economics of privacy and data labor

Nikolaos Vasiloglou, Rachel Cummings, Glen Weyl, Paris Koutris, Meg Young, Ruoxi Jia, David Dao, Bo Waggoner

Sat Jul 18, 06:00 AM

Although data is considered to be the "new oil," it is very hard to be priced. Raw use of data has been invaluable in several sectors such as advertising, healthcare, etc., but often in violation of people’s privacy. Labeled data has also been extremely valuable for the training of machine learning models (driverless car industry). This is also indicated by the growth of annotation companies such as Figure8 and Scale.AI, especially in the image space. Yet, it is not clear what is the right pricing for data workers who annotate the data or the individuals who contribute their personal data while using digital services. In the latter case, it is very unclear how the value of the services offered is compared to the private data exchanged. While the first data marketplaces have appeared, such as AWS, Narrative.io, nitrogen.ai, etc., they suffer from a lack of good pricing models. They also fail to maintain the right of the data owners to define how their own data will be used. There have been numerous suggestions for sharing data while maintaining privacy, such as training generative models that preserve original data statistics.
Abstract 1: Designing Differentially Private Estimators in High Dimensions by Aditya Dhar in Economics of privacy and data labor, 07:00 AM

We study differentially private mean estimation in a high-dimensional setting. Existing differential privacy techniques applied to large dimensions lead to computationally intractable problems or estimators with excessive privacy loss. Recent work in high-dimensional robust statistics has identified computationally tractable mean estimation algorithms with asymptotic dimension-independent error guarantees. We incorporate these results to develop a strict bound on the global sensitivity of the robust mean estimator. This yields a computationally tractable algorithm for differentially private mean estimation in high dimensions with dimension-independent privacy loss. Finally, we show on synthetic data that our algorithm significantly outperforms classic differential privacy methods, overcoming barriers to high-dimensional differential privacy.

Abstract 2: Really Useful Synthetic Data â€“ A Framework to Evaluate the Quality of Differentially Private Synthetic Data by Christian Arnold in Economics of privacy and data labor, 07:15 AM

Recent advances in generating synthetic data that allow to add principled ways of protecting privacy -- such as Differential Privacy -- are a crucial step in sharing statistical information in a privacy preserving way. But while the focus has been on privacy guarantees, the resulting private synthetic data is only useful if it still carries statistical information from the original data. To further optimise the inherent trade-off between data privacy and data quality, it is necessary to think closely about the latter. What is it that data analysts want? Acknowledging that data quality is a subjective concept, we develop a framework to evaluate the quality of differentially private synthetic data from an applied researcher's perspective. Data quality can be measured along two dimensions. First, quality of synthetic data can be evaluated against training data or against an underlying population. Second, the quality of synthetic data depends on general similarity of distributions or on performance for specific tasks such as inference or prediction. It is clear that accommodating all goals at once is a formidable challenge. We invite the academic community to jointly advance the privacy-quality frontier.

Abstract 3: Generating Privacy-Preserving Synthetic Tabular Data Using Oblivious Variational Autoencoders by L Vivek Harsha in Economics of privacy and data labor, 07:30 AM

Recent advances in generating synthetic data that allow to add principled ways of protecting privacy -- such as Differential Privacy -- are a crucial step in sharing statistical information in a privacy preserving way. But while the focus has been on privacy guarantees, the resulting private synthetic data is only useful if it still carries statistical information from the original data. To further optimise the inherent trade-off between data privacy and data quality, it is necessary to think closely about the latter. What is it that data analysts want? Acknowledging that data quality is a subjective concept, we develop a framework to evaluate the quality of differentially private synthetic data from an applied researcher's perspective. Data quality can be measured along two dimensions. First, quality of synthetic data can be evaluated against training data or against an underlying population. Second, the quality of synthetic data depends on general similarity of distributions or on performance for specific tasks such as inference or prediction. It is clear that accommodating all goals at once is a formidable challenge. We invite the academic community to jointly advance the privacy-quality frontier.

Abstract 4: European Privacy Law and Global Markets for Data by Christian Peukert in Economics of privacy and data labor, 07:45 AM

To Call or not to Call? Using ML Prediction APIs more Accurately and Economically by Lingxiao Chen in Economics of privacy and data labor, 08:00 AM

Abstract 5: Do Markets Make Sense for Personal Data? by Aileen Nielsen in Economics of privacy and data labor, 08:15 AM

Abstract 6: Optimal Query Complexity of Secure Stochastic Convex Optimization by Wei Tang in Economics of privacy and data labor, 08:30 AM

We study the \textit{secure} stochastic convex optimization problem: a learner aims to learn the optimal point of a convex function through sequentially querying a \textit{(stochastic)} gradient oracle, in the meantime, there exists an adversary who aims to free-ride and infer the learning outcome of the learner from observing the learner's queries. The adversary observes only the points of the queries but not the feedback from the oracle. The goal of the learner is to optimize the accuracy, i.e., obtaining an accurate estimate of the optimal point, while securing her privacy, i.e.,
making it difficult for the adversary to infer the optimal point. We formally quantify this tradeoff between learner’s accuracy and privacy and characterize the lower and upper bounds on the learner’s query complexity as a function of desired levels of accuracy and privacy. For the analysis of lower bounds, we provide a general template based on information theoretical analysis and then tailor the template to several families of problems, including stochastic convex optimization and (noisy) binary search. We also present a generic secure learning protocol that achieves the matching upper bound up to logarithmic factors.

Abstract 7: On Detecting Data Pollution Attacks On Recommender Systems Using Sequential GANs by Behzad Shahrasb in Economics of privacy and data labor, 09:15 AM

Recommender systems are an essential part of any e-commerce platform. Recommendations are typically generated by aggregating large amounts of user data. A malicious actor may be motivated to sway the output of such recommender systems by injecting malicious datapoints to leverage the system for financial gain. In this work, we propose a semi-supervised attack detection algorithm to identify the malicious datapoints. We do this by leveraging a portion of the dataset that has a lower chance of being polluted to learn the distribution of genuine datapoints. Our proposed approach modifies the Generative Adversarial Network architecture to take into account the contextual information from user activity. This allows the model to distinguish legitimate datapoints from the injected ones.

Abstract 8: Efficient Privacy-Preserving Stochastic Nonconvex Optimization by Lingxiao Wang in Economics of privacy and data labor, 09:30 AM

While many solutions for privacy-preserving convex empirical risk minimization (ERM) have been developed, privacy-preserving nonconvex ERM remains a challenge. We study nonconvex ERM, which takes the form of minimizing a finite-sum of nonconvex loss functions over a training set. We propose a new differentially private stochastic gradient descent algorithm for nonconvex ERM that achieves strong privacy guarantees efficiently, and provide a tight analysis of its privacy and utility guarantees, as well as its gradient complexity. Our algorithm substantially reduces gradient complexity while matching the best previous utility guarantee given by Wang et al. (NeurIPS 2017). Our experiments on benchmark nonconvex ERM problems demonstrate superior performance in terms of both training cost and utility gains compared with previous differentially private methods using the same privacy budgets.

Abstract 10: European Privacy Law and Global Markets for Data by Christian Peukert in Economics of privacy and data labor, 10:30 AM

We demonstrate how privacy law interacts with competition and trade policy in the context of the European General Data Protection Regulation (GDPR). We follow more than 110,000 websites for 18 months to show that websites reduced their connections to web technology providers after GDPR became effective, especially regarding requests involving personal data. This also holds for websites catering to non-EU audiences and therefore not bound by GDPR. We further document an increase in market concentration in web technology services after the introduction of GDPR. While most firms lose market share, the leading firm, Google, significantly increases market share.

Abstract 11: To Call or not to Call? Using ML Prediction APIs more Accurately and Economically by Lingjiao Chen in Economics of privacy and data labor, 10:45 AM

Prediction APIs offered for a fee are a fast growing industry and an important part of machine learning as a service. While many such services are available, the heterogeneity in their price and performance makes it challenging for users to decide which API or combination of APIs to use for their own data and budget. In this paper, we take a first step towards addressing this challenge by proposing FrugalML, a principled framework that jointly learns the strength and weakness of each API on different data, and performs an efficient optimization to automatically identify the best sequential strategy to adaptively use the available APIs within a budget constraint. Preliminary experiments using ML APIs from Google, Microsoft and Face++ for a facial emotion recognition task show that FrugalML typically leads to more than 50% cost reduction while matching the accuracy of the best single API.

Abstract 12: Do Markets Make Sense for Personal Data? by Aileen Nielsen in Economics of privacy and data labor, 11:00 AM

Collection and sale of personal data is a common and economically rewarding activity. However, the contractual model of notice and consent that governs this activity under U.S. law relies on an assumption that personal data can and does function as a market good. This paper presents experimental evidence of a conflict between the market nature of personal data assumed by many legal frameworks and the conceptual categorization of personal data transactions by the ordinary people putatively protected by notice and consent legal frameworks. I present two online vignette studies that repurpose designs from the taboo trade-offs literature and suggest that protection of personal data rises to the level of a sacred value.

Abstract 14: Intersectional Social Data by Glen Weyl in Economics of privacy and data labor, 11:30 AM

Data are interpersonal relational rather than atomistically personal or universally objective. Yet the group of people to which a datum pertains is different across all data pertaining to any person, and thus every person sits at the intersection of a diversity of data collectives. A data structure representing this as well as interpersonal relationships of trust has the potential to add a trust layer to internet-type structures, allow the verification at higher levels of trust of a far wider range of data than current data structures and eventually to enable political economies far more sophisticated than even those currently considered innovative (such as current advocated by organizations like RadicalxChange).

7th ICML Workshop on Automated Machine Learning (AutoML 2020)

Frank Hutter, Joaquin Vanschoren, Marius Lindauer, Charles Weil, Katharina Eggensperger, Matthias Feurer

Sat Jul 18, 06:00 AM

Machine learning has achieved considerable successes in recent years, but this success often relies on human experts, who construct appropriate features, design learning architectures, set their hyperparameters, and develop new learning algorithms. Driven by the demand for off-the-shelf machine learning methods from an ever-growing
community, the research area of AutoML targets the progressive automation of machine learning aiming to make effective methods available to everyone. Hence, the workshop targets a broad audience ranging from core machine learning researchers in different fields of ML connected to AutoML, such as neural architecture search, hyperparameter optimization, meta-learning, and learning to learn, to domain experts aiming to apply machine learning to new types of problems.

The schedule is wrt CEST (i.e., the time zone of Vienna)

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:00 AM</td>
<td>Welcome</td>
<td>Hutter</td>
</tr>
<tr>
<td>06:05 AM</td>
<td>&quot;Open Challenges for Automated Machine Learning: Solving Intellectual Debt with Auto AI&quot; by Neil Lawrence</td>
<td></td>
</tr>
<tr>
<td>06:05 AM</td>
<td>1min Intro</td>
<td></td>
</tr>
<tr>
<td>06:30 AM</td>
<td>Keynote Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>06:45 AM</td>
<td>Contributed Talk 1: Provably Efficient Online Hyperparameter Optimization with Population-Based Bandits</td>
<td>Parker-Holder, Nguyen, Roberts</td>
</tr>
<tr>
<td>07:00 AM</td>
<td>Contributed Talk 1 Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>07:10 AM</td>
<td>1.4 Multi-Source Unsupervised Hyperparameter Optimization</td>
<td>Nomura</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>1.1 MTL2L: A Context Aware Neural Optimiser</td>
<td>Kuo</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>1.3 Cost-Aware Bayesian Optimization</td>
<td>Lee</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>1.5 Regression Networks for Meta-Learning Few-Shot Classification</td>
<td>Devos, Grossglauser</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>1.2 AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data</td>
<td>Mueller</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>1.9 Bayesian optimization for Iterative Learning</td>
<td>Nguyen</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>1.10 Weighted Meta-Learning</td>
<td>Cai</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>1.6 Mining Documentation to Extract Hyperparameter Schemas</td>
<td>Baudart, Kirchner, Hirzel, Kate</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>1.7 Tiny Video Networks: Architecture Search for Efficient Video Models</td>
<td>Piergiovanni</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>1.8 Solving Heterogeneous AutoML Problems with AutoGOAL</td>
<td>Estévez Velarde</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>1.15 Analysis of Imbalance Strategies</td>
<td>Costa</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>1.11 Stabilizing Bi-Level Hyperparameter Optimization</td>
<td>Dhar</td>
</tr>
<tr>
<td>07:20 AM</td>
<td>1.13 A Study on Encodings for Neural Architecture Search</td>
<td>White</td>
</tr>
<tr>
<td>07:25 AM</td>
<td>1.14 Multi-fidelity zero-shot Classification</td>
<td>Ram, Liu</td>
</tr>
<tr>
<td>07:25 AM</td>
<td>1.19 Self-Supervised Prototypical Transfer Learning for Few-Shot Classification</td>
<td>Medina, Devos, Grossglauser</td>
</tr>
<tr>
<td>07:25 AM</td>
<td>1.18 Toward Synergism in Macro Action Ensembles</td>
<td>Chang, Hong, Lee</td>
</tr>
<tr>
<td>07:30 AM</td>
<td>Poster Session 1</td>
<td></td>
</tr>
<tr>
<td>08:40 AM</td>
<td>1 min Intro</td>
<td></td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Contributed Talk 2: Bayesian Optimization with Fairness Constraints</td>
<td>Perrone</td>
</tr>
<tr>
<td>08:55 AM</td>
<td>Contributed Talk 2 Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:50 AM</td>
<td>&quot;Automated ML and its transformative impact on medicine and healthcare&quot; by Mihaela van der Schaar</td>
<td>van der Schaar</td>
</tr>
<tr>
<td>10:50 AM</td>
<td>1min Intro</td>
<td></td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Keynote Talk Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Contributed Talk 3: How far are we from true AutoML: reflection from winning solutions and results of AutoDL challenge</td>
<td>Liu</td>
</tr>
</tbody>
</table>
11:45 AM Contributed Talk 3 Q&A

11:55 AM 2.2 Towards Algorithm-Agnostic Uncertainty Estimation: Predicting Classification Error in an Automated Machine Learning Setting

11:55 AM 2.1 Federated Meta-Learning: Democratizing Algorithm Selection Across Disciplines and Software Libraries

11:55 AM 2.5 Geometric Dataset Distances via Optimal Transport

11:55 AM 2.3 Collecting Empirical Data About Hyperparameters for Data Driven AutoML

11:55 AM 2.4 Local Search is State of the Art for Neural Architecture Search Benchmarks

12:00 PM 2.7 A Simple Setting for Understanding Neural Architecture Search with Weight-Sharing

12:00 PM 2.10 Uncertainty aware Search framework for Multi-Objective Bayesian Optimization with Constraints

12:00 PM 2.6 Meta-Learning for Recalibration of EMG-Based Upper Limb Prostheses

12:00 PM 2.8 Meta-SAC: Auto-tune the Entropy Temperature of Soft Actor-Critic via Metail gradient

12:00 PM 2.9 aâ€™Algorithm-Performance Personasâ€™ for Siamese Meta-Learning and Automated Algorithm Selection

12:00 PM 2.11 RicciNets: Curvature-guided Pruning of High-performance Neural Networks Using Ricci Flow

12:00 PM 2.14 On Evaluation of AutoML Systems

Abstracts (41):

Abstract 1: Welcome in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Hutter 06:00 AM

[Video](https://slideslive.com/38930620)

Abstract 2: "Open Challenges for Automated Machine Learning: Solving Intellectual Debt with Auto AI" by Neil Lawrence in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Lawrence 06:05 AM

Machine learning models are deployed as part of wider systems where outputs of one model are consumed by other models. This composite structure for machine learning systems is the dominant approach for deploying artificial intelligence. Such deployed systems can be complex to understand, they bring with them intellectual debt. In this talk we'll argue that the next frontier for automated machine learning is to move to automation of the systems design, going from AutoML to AutoAI.

[Video](https://slideslive.com/38930621)

Abstract 5: Contributed Talk 1: Provably Efficient Online Hyperparameter Optimization with Population-Based Bandits in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Parker-Holder, Nguyen, Roberts 06:45 AM

[Video](https://slideslive.com/38930622)

Abstract 7: 1.4 Multi-Source Unsupervised Hyperparameter Optimization in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Nomura 07:10 AM

[Poster presentation (Zoom meeting)](https://us02web.zoom.us/j/85167826018?pwd=S1czSzRYeXRiNDZxS0xptTtXb2xTRj09)

Abstract 8: 1.1 MTL2L: A Context Aware Neural Optimiser in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Kuo 07:10 AM

Abstract 9: 1.3 Cost-Aware Bayesian Optimization in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Lee 07:10 AM

Abstract 10: 1.5 Regression Networks for Meta-Learning Few-Shot Classification in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Devos, Grossglauser 07:10 AM

Abstract 11: 1.2 AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Mueller 07:10 AM

Abstract 12: 1.9 Bayesian optimization for Iterative Learning in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Nguyen 07:15 AM

Abstract 13: 1.10 Weighted Meta-Learning in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Cai 07:15 AM

Abstract 14: 1.6 Mining Documentation to Extract Hyperparameter Schemas in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Baudart, Kirchner, Hirzel, Kate 07:15 AM


Abstract 16: 1.8 Solving Heterogeneous AutoML Problems with AutoGOAL in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Estévez Velarde 07:15 AM

Abstract 17: 1.11 Stabilizing Bi-Level Hyperparameter Optimization using Moreau-Yosida Regularization in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Dhar 07:20 AM

ICML 2020 Workshop book

Abstract 19: 1.13 A Study on Encodings for Neural Architecture Search in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), White 07:20 AM
[Video](https://slideslive.com/38930633)

[Video](https://slideslive.com/38930633)

[Video](https://slideslive.com/38930633)

[Video](https://slideslive.com/38930663)

Abstract 23: 1.18 Toward Synergism in Macro Action Ensembles in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Chang, Hong, Lee 07:25 AM
[Video](https://slideslive.com/38930663)

Abstract 24: 1.17 W-EDGE: Weight Updating in Directed Graph Ensembles to improve Classification in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Fontes 07:25 AM
[Video](https://slideslive.com/38930633)

Contributed Talk 1: How far are we from true AutoML: a keynote session, I will explain the unique characteristics of healthcare that make it a challenging but extremely promising domain in which to apply AutoML. I will give an overview of several novel approaches we have developed to tackle problems as complex and diverse as AutoML for survival analysis, causal inference, and dynamic forecasting from time-series data. I will also highlight medical AutoML frameworks used in real-world contexts, including predictive tools deployed in response to the COVID-19 pandemic.

Contributed Talk 2: Bayesian Optimization with Fairness Constraints in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Perrone 08:40 AM
[Video](https://slideslive.com/38930633)

Contributed Talk 3: How far are we from true AutoML: reflection from winning solutions and results of AutoDL challenge in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Liu 11:30 AM
[Video](https://slideslive.com/38930633)

Abstract 25: 1.16 Learning to Prune Deep Neural Networks via Reinforcement Learning in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Gupta 07:25 AM
[Video](https://slideslive.com/38930633)

Abstract 26: 1.15 Solving Constrained CASH Problems with ADMM in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Ram, Liu 07:20 AM
[Video](https://slideslive.com/38930633)

[Video](https://slideslive.com/38930633)

Abstract 28: Contributed Talk 2: Bayesian Optimization with Fairness Constraints in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Perrone 08:40 AM
[Video](https://slideslive.com/38930633)

Abstract 29: Contributed Talk 3: How far are we from true AutoML: reflection from winning solutions and results of AutoDL challenge in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Liu 11:30 AM
[Video](https://slideslive.com/38930633)

Abstract 30: Contributed Talk 2: Bayesian Optimization with Fairness Constraints in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Perrone 08:40 AM
[Video](https://slideslive.com/38930633)

Abstract 31: "Automated ML and its transformative impact on medicine and healthcare" by Mihaela van der Schaar in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), van der Schaar 11:55 AM
[Video](https://slideslive.com/38930633)

Abstract 32: 1.13 A Study on Encodings for Neural Architecture Search in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), White 07:20 AM
[Video](https://slideslive.com/38930633)

[Video](https://slideslive.com/38930651)


[Video](https://slideslive.com/38930654)

Abstract 39: 2.3 Collecting Empirical Data About Hyperparameters for Data Driven AutoML in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Binder 11:55 AM

[Video](https://slideslive.com/38930666)

Abstract 40: 2.4 Local Search is State of the Art for Neural Architecture Search Benchmarks in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), White 11:55 AM

[Video](https://slideslive.com/38930655)


[Video](https://slideslive.com/38930658)

Abstract 42: 2.10 Uncertainty aware Search framework for Multi-Objective Bayesian Optimization with Constraints in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Proroković 12:00 PM

[Video](https://slideslive.com/38930659)

Abstract 43: 2.6 Meta-Learning for Recalibration of EMG-Based Upper Limb Prostheses in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Proroković 12:00 PM

[Video](https://slideslive.com/38930660)

Abstract 44: 2.8 Meta-SAC: Auto-tune the Entropy Temperature of Soft Actor-Critic via Metagradient in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Wang, Ni 12:00 PM

[Video](https://slideslive.com/38930661)

Abstract 45: 2.9 Algorithm-Performance Personas™ for Universal Meta-Learning and Automated Algorithm Selection in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Belakaria 12:00 PM

[Video](https://slideslive.com/38930662)

Abstract 46: 2.9 'Algorithm-Performance Personas' for Universal Meta-Learning and Automated Algorithm Selection in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Proroković 12:00 PM

[Video](https://slideslive.com/38930663)

[Video](https://slideslive.com/38930661)

[Poster presentation (Zoom meeting)](https://us02web.zoom.us/j/84136152376?pwd=ajFuT3ArT0VodThUWXlialc2NTR4Zz09)


[Video](https://slideslive.com/38930664)

[Poster presentation (Zoom meeting)](https://us02web.zoom.us/j/88635013048?pwd=c3hQU0orZ3NxcxhDM3RidnTIQXFUz09)


Abstract 49: 2.13 Bayesian Optimization for real-time, automatic design of face stimuli in human-centred research in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), F da Costa 12:05 PM

[Video](https://slideslive.com/38930665)

[Poster presentation (Zoom meeting)](https://us02web.zoom.us/j/88019874576?pwd=QlNvK1U5czZBNVYoTmFwZGxPOT09)


[Video](https://slideslive.com/38930662)

[Poster presentation (Zoom meeting)](https://us02web.zoom.us/j/89600172124?pwd=V1BrZTlEUnFLYzhQVBMVTRsYzVQUT09)


Abstract 52: “AutoGluon and Distillation” by Alex Smola in 7th ICML Workshop on Automated Machine Learning (AutoML 2020), Smola 01:15 PM

[Video](https://slideslive.com/38930673)

Real World Experiment Design and Active Learning

Ilja Bogunovic, Willie Neiswanger, Yisong Yue

Sat Jul 18, 07:00 AM

This workshop aims to bring together researchers from academia and industry to discuss major challenges, outline recent advances, and highlight future directions pertaining to novel and existing large-scale real-world experiment design and active learning problems. We aim to highlight new and emerging research opportunities for the machine learning community that arise from the evolving needs to make experiment design and active learning procedures that are theoretically and practically relevant for realistic applications.
In situations where a task can be cleanly formulated and data is plentiful, modern machine learning (ML) techniques have achieved impressive (and often super-human) results. Here, "plentiful data" can mean labels from humans, access to a simulator and well designed reward function, or other forms of interaction and supervision.

On the other hand, in situations where tasks cannot be cleanly formulated and plentifully supervised, ML has not yet shown the same progress. We still seem far from flexible agents that can learn without human engineers carefully designing or collating their supervision. This is problematic in many settings where machine learning is or will be applied in real world settings, where these agents have to interact with human users and may be used in settings that go beyond any initial clean training data used during system development. A key open question is how to make machine learning effective and robust enough to operate in real world open domains.

Artificial ("open") worlds are ideal laboratories for studying how to extend the successes of ML to build such agents. Open worlds are characterized by:

- Large (or perhaps infinite) collections of tasks, often not specified till test time; or lack of well defined tasks altogether (despite there being lots to do).
- "unbounded" environments, long "episodes"; or no episodes at all.
- Many interacting agents; more generally, emergent behavior from interactions with the environment.

On one hand, they retain many of the challenging features of the real world with respect to studying learning agents. On the other hand, they allow cheap collection of environment interaction data. Furthermore, because many artificial worlds of interest are games that people enjoy playing, they could allow interaction with humans at scale.

A particularly promising direction is that open world games can bridge: the closed domains and benchmarks that have traditionally driven research progress and open ended real world applications in which resulting technology is deployed.

We propose a workshop designed to catalyze research towards addressing these challenges posed by machine learning in open worlds. Our goal is to bring together researchers with a wide range of perspectives whose work focuses on, or is enabled by, open worlds. This would be the very first workshop focused on this topic, and we anticipate that it would play a key role in sharing experience, brainstorming ideas, and catalyzing novel directions for research.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07:00 AM</td>
<td>Opening remarks</td>
<td>Hofmann</td>
</tr>
<tr>
<td>07:15 AM</td>
<td>The NetHack Learning Environment</td>
<td>Rocktäschel</td>
</tr>
<tr>
<td>07:35 AM</td>
<td>The NetHack Learning Environment Q&amp;A</td>
<td>Rocktäschel, Hofmann</td>
</tr>
<tr>
<td>07:45 AM</td>
<td>Open-ended environments for advancing RL</td>
<td>Jaderberg</td>
</tr>
<tr>
<td>08:05 AM</td>
<td>Open-ended environments for advancing RL Q&amp;A</td>
<td>Jaderberg, Hofmann</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Coffee break</td>
<td>Srinet</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Collaborative Construction and Communication in Minecraft</td>
<td>Hockenmaier</td>
</tr>
<tr>
<td>08:50 AM</td>
<td>Collaborative Construction and Communication in Minecraft Q&amp;A</td>
<td>Hockenmaier, Szlam</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Collaboration in Situated Instruction Following</td>
<td>Artzi</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Collaboration in Situated Instruction Following Q&amp;A</td>
<td>Artzi, Szlam</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Lunch and networking</td>
<td>Srinet</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>What an Agent Knows: Evaluation in Open Worlds</td>
<td>Keamey</td>
</tr>
<tr>
<td>10:20 AM</td>
<td>What an Agent Knows: Evaluation in Open Worlds Q&amp;A</td>
<td>Keamey, Guss</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Endless Frontiers?</td>
<td>Togelius</td>
</tr>
<tr>
<td>10:50 AM</td>
<td>Endless Frontiers? Q&amp;A</td>
<td>Togelius, Guss</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Coffee break</td>
<td>Srinet</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Poster session</td>
<td>Kuno, Srinet, Guss, Houghton</td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Panel discussion</td>
<td>Srinet, Hofmann, Artzi, Keamey, Hockenmaier</td>
</tr>
</tbody>
</table>

Abstracts (14):
Abstract 2: The NetHack Learning Environment in Workshop on Learning in Artificial Open Worlds, Rocktäschel 07:15 AM

Progress in Reinforcement Learning (RL) algorithms goes hand-in-hand with the development of challenging environments that test the limits of current methods. While existing RL environments are either sufficiently complex or based on fast simulation, they are rarely both these things. In this talk, I present the NetHack Learning Environment (NLE), a scalable, procedurally generated, stochastic, rich, and challenging environment for RL research based on the popular single-player terminal-based rogue like game, NetHack. We argue that NetHack is sufficiently complex to drive long-term research on problems such as exploration, planning, skill acquisition, and language-conditioned RL, while dramatically reducing the computational resources required to gather a large amount of experience. We compare NLE and its task suite to existing alternatives, and discuss why it is an ideal medium for testing the robustness and systematic generalization of RL agents. We demonstrate empirical success for early stages of the game using a distributed deep RL baseline and Random Network Distillation exploration, alongside qualitative analysis of various agents trained in the environment. NLE is open-source at https://github.com/facebookresearch/nle .

Abstract 3: The NetHack Learning Environment Q&A in Workshop on Learning in Artificial Open Worlds, Rocktäschel, Hofmann 07:35 AM

Please join using the Zoom link and post your questions on Rocketchat.

Abstract 4: Open-ended environments for advancing RL in Workshop on Learning in Artificial Open Worlds, Jaderberg 07:45 AM

The field of reinforcement learning is pushed forwards by the presence of challenging environments. Over the years, the complexity of these environments has continued to increase, but the question is how can we continue to push the complexity of environments with respect to the optimal policy complexity in a scalable manner. Here I will discuss using multi-agent environments to create more open-ended environments, and discuss examples of our work to move in this direction with Capture the Flag and Starcraft 2. Finally I will discuss some future directions for generating even more open-ended environments to further push our RL algorithms.

Abstract 5: Open-ended environments for advancing RL Q&A in Workshop on Learning in Artificial Open Worlds, Jaderberg, Hofmann 08:05 AM

Please join using the Zoom link and post your questions on Rocketchat.

Abstract 7: Collaborative Construction and Communication in Minecraft in Workshop on Learning in Artificial Open Worlds, Hockenmaier 08:30 AM

I will present work done by my group on defining a collaborative construction task that allows us to use the Minecraft platform to study situated natural language generation and understanding. In this task, one player (the Architect) needs to instruct another (the Builder) via a chat interface to construct a given target structure that only the Architect is shown. I will discuss what makes this task interesting and challenging. I will also describe models that we have developed for the Architect and the Builder role, and discuss what remains to be done to create agents that can solve this task.

Abstract 8: Collaborative Construction and Communication in Minecraft Q&A in Workshop on Learning in Artificial Open Worlds, Hockenmaier, Szlam 08:50 AM

Please join using the Zoom link and post your questions on Rocketchat.

Abstract 9: Collaboration in Situated Instruction Following in Workshop on Learning in Artificial Open Worlds, Artzi 09:00 AM

I will focus on the problem of executing natural language instructions in a collaborative environment. I will propose the task of learning to follow sequences of instructions in a collaborative scenario, where two agents, a leader and a follower, execute actions in the environment and the leader controls the follower using natural language. To study this problem, we build CerealBar, a multi-player 3D game where a leader instructs a follower, and both act in the environment together to accomplish complex goals. I will focus on learning an autonomous follower that executes the instructions of a human leader. I will briefly describe a model to address this problem, and a learning method that relies on static recorded human-human interactions, while still learning to recover from cascading errors between instructions.

Abstract 10: Collaboration in Situated Instruction Following Q&A in Workshop on Learning in Artificial Open Worlds, Artzi, Szlam 09:20 AM

Please join using the Zoom link and post your questions on Rocketchat.

Abstract 12: What an Agent Knows: Evaluation in Open Worlds in Workshop on Learning in Artificial Open Worlds, Kearney 10:00 AM

Agents tackling complex problems in open environments often benefit from the ability to construct knowledge. Learning to independently solve sub-tasks and form models of the world can help agents progress in solving challenging problems. In this talk, we draw attention to challenges that arise when evaluating an agent’s knowledge, specifically focusing on methods that express an agent’s knowledge as predictions. Using the General Value Function framework we highlight the distinction between useful knowledge and strict measures of accuracy. Having identified challenges in assessing an agent’s knowledge, we propose a possible evaluation approach that is compatible with large and open worlds.


Please join using the Zoom link and post your questions on Rocketchat.

Abstract 14: Endless Frontiers? in Workshop on Learning in Artificial Open Worlds, Togelius 10:30 AM

The research community is gradually coming to a realization that policies trained arcade-like video games are very limited. They overfit badly and are not going to take us far along the way to some sort of general intelligence. This should perhaps not be surprising, given that such games generally have tightly defined tasks, fixed perspectives, and generally static worlds. More and more attention is therefore given to games that are in some sense open-ended or feature open worlds. Could such games be the solution to our problems, allowing the development of more general artificial intelligence? Perhaps, but basing competitions or benchmarks on open-ended games is not going to be
easy, as the very features which make for a good benchmark are the same that lead to brittle policies. Shoe-horning open-world games into a standard RL framework is unlikely to be the best option for going forward. Many of the most interesting opportunities for developing intelligent behavior are likely to come from agents constructing their own challenges and environments. The boundary between playing a game and constructing a world is not well-defined: I will give examples from where the same RL setup was used to play SimCity and to develop game levels. I will also briefly introduce the Generative Design in Minecraft Competition, which focuses on building believable settlements.

Abstract 15: Endless Frontiers? Q&A in Workshop on Learning in Artificial Open Worlds, Togelius, Guss 10:50 AM

Please join using the Zoom link and post your questions on Rocketchat.

Abstract 17: Poster session in Workshop on Learning in Artificial Open Worlds, Kuno, Srinet, Guss, Houghton 11:30 AM

Hey all!
Our poster session is hosted in a virtual town.
To go to the poster session, please go to:
https://gather.town/bWIYAhNyvQPRiCE/iCML2020-LAOW

and use the password: flatfish

Abstract 18: Panel discussion in Workshop on Learning in Artificial Open Worlds, Srinet, Holtmann, Artzi, Kearney, Hockenmaier 01:00 PM

Join us for a panel discussion with our invited speakers

1st Workshop on Language in Reinforcement Learning (LaReL)

Nantas Nardelli, Jelena Luketina, Nantas Nardelli, Jakob Foerster, Victor Zhong, Jacob Andreas, Edward Grefenstette, Tim Rocktäschel

Sat Jul 18, 07:00 AM

Language is one of the most impressive human accomplishments and is believed to be the core to our ability to learn, teach, reason and interact with others. Yet, current state-of-the-art reinforcement learning agents are unable to use or understand human language at all. The ability to integrate and learn from language, in addition to rewards and demonstrations, has the potential to improve the generalization, scope and sample efficiency of agents. Furthermore, many real-world tasks, including personal assistants and general household robots, require agents to process language by design, whether to enable interaction with humans, or simply use existing interfaces. The aim of our workshop is to advance this emerging field of research by bringing together researchers from several diverse communities to discuss recent developments in relevant research areas such as instruction following and embodied language learning, and identify the most important challenges and promising research avenues.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>07:00 AM</td>
<td>Welcome Remarks</td>
</tr>
<tr>
<td>07:10 AM</td>
<td>Invited Talk: Angeliki Lazaridou</td>
</tr>
<tr>
<td>07:40 AM</td>
<td>Invited Talk: Arthur Szlam</td>
</tr>
<tr>
<td>08:10 AM</td>
<td>Coffee break 1</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>Poster session 1</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Invited Talk: Felix Hill</td>
</tr>
<tr>
<td>10:45 AM</td>
<td>Invited Talk: Karthik Narasimhan</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Invited Talk: Yoav Artzi</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Coffee break 2</td>
</tr>
<tr>
<td>12:05 PM</td>
<td>Poster session 2</td>
</tr>
<tr>
<td>12:50 PM</td>
<td>Short Break</td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Invited Talk: Marc-Alexandre CÃ³tÅ©</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Invited Talk: Alison Gopnik</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Closing Remarks</td>
</tr>
</tbody>
</table>

Abstracts (9):

Abstract 2: Invited Talk: Angeliki Lazaridou in 1st Workshop on Language in Reinforcement Learning (LaReL), Lazaridou 07:10 AM

The ability to cooperate through language is a defining feature of humans. As the perceptual, motory and planning capabilities of deep artificial networks increase, researchers are studying whether they also can develop a shared language to interact. In this talk, I will highlight recent advances in this field but also common headaches (or perhaps limitations) with respect to experimental setup and evaluation of emergent communication. Towards making multi-agent communication a building block of human-centric AI, and by drawing from my own recent work, I will discuss approaches on making emergent communication relevant for human-agent communication in natural language.

Abstract 3: Invited Talk: Arthur Szlam in 1st Workshop on Language in Reinforcement Learning (LaReL), Nardelli 07:40 AM

I will discuss our progress on a research program aimed at building a Minecraft assistant. I will cover the tools and platform we have built allowing players to interact with the agents and to record those interactions, and the data we have collected. I will also cover the design of our current agent, from which we (and hopefully others) can iterate.

Abstract 5: Poster session 1 in 1st Workshop on Language in Reinforcement Learning (LaReL), Nardelli 08:30 AM

Check out the papers and their short presentations here: https://larel-ws.github.io/accepted-papers/

Meet the authors in LaReL’s Gather Town:
https://tinyurl.com/gather-larel
Abstract 7: Invited Talk: Felix Hill in 1st Workshop on Language in Reinforcement Learning (LaReL), Hill 10:15 AM

Models like BERT or GPT-2 can do amazing things with language, and this raises the interesting question of whether such text-based models could ever really "understand" it. One clear difference between BERT-understanding and human understanding is that BERT doesn’t learn to connect language to its actions or its perception of the world it inhabits. I’ll discuss an alternative approach to language understanding in which a neural-network-based agent is trained to associate words and phrases with things that it learns to see and do. First, I’ll provide some evidence for the promise of this approach by showing that the interactive, first-person perspective of an agent affords it with a particular inductive bias that helps it to extend its training experience to generalize to out-of-distribution settings in ways that seem natural or 'systematic'. Second, I’ll show the amount of ‘propositional’ (i.e. linguistic) knowledge that emerges in the internal states of the agent as it interacts with the world can be increased significantly by it learning to make predictions about observations multiple timesteps into the future. This underlines some important common ground between the agent-based and BERT-style approaches: both attest to the power of prediction and the importance of context in acquiring semantic representations. Finally, I’ll connect BERT and agent-based learning in a more literal way, by showing how an agent endowed with BERT representations can achieve substantial (zero-shot) transfer from template-based language to noisy natural instructions given by humans with access to the agent’s world.

Abstract 8: Invited Talk: Karthik Narasimhan in 1st Workshop on Language in Reinforcement Learning (LaReL), Narasimhan 10:45 AM

In recent years, reinforcement learning (RL) has been used with considerable success in games and robotics as well as language understanding applications like dialog systems. However, the question of what language can provide for RL remains relatively under-explored. In this talk, I make the case that leveraging language will be essential to developing general-purpose interactive agents that can perform more than a single task and operate in scenarios beyond the ones they are trained on. Natural language allows us to incorporate more semantic structure into the RL framework while also making it easier to obtain guidance from humans. Specifically, I will show how several parts of the traditional RL setup (e.g. transitions, rewards, actions, goals) can be expressed in language to build agents that can handle combinatorially large spaces as well as generalize to unseen subspaces in each of these aspects.

Abstract 9: Invited Talk: Yoav Artzi in 1st Workshop on Language in Reinforcement Learning (LaReL), Artzi 11:15 AM

I will discuss the task of executing natural language instructions with a physical robotic agent. In contrast to existing work, we do not engineer formal representations of language meaning or the robot environment. Instead, we learn to directly map raw observations and language to low-level continuous control of a quadcopter drone. We use an interpretable neural network model that mixes learned representations with differentiable geometric operations. For training, we introduce Supervised and Reinforcement Asynchronous Learning (SuReAL), a learning algorithm that utilizes supervised and reinforcement learning processes that constantly interact to learn robust reasoning with limited data. Our learning algorithm uses demonstrations and a plan-following intrinsic reward signal. While we do not require any real-world autonomous flight during learning, our model works effectively both in simulation and the real environment.

Abstract 11: Poster session 2 in 1st Workshop on Language in Reinforcement Learning (LaReL), Nardelli 12:05 PM

Check out the papers and their short presentations here: https://larel-ws.github.io/accepted-papers/

Meet the authors in LaReL’s Gather Town: https://tinyurl.com/gather-larel

Abstract 13: Invited Talk: Marc-Alexandre CÃ¨tÃ© in 1st Workshop on Language in Reinforcement Learning (LaReL), CÃ¨tÃ© 01:00 PM

Text-based games are complex, interactive simulations in which text describes the game state and players make progress by entering text commands. They are fertile ground for language-focused machine learning research. In addition to language understanding, successful play requires skills like long-term memory and planning, exploration (trial and error), and common sense. The talk will introduce TextWorld, a sandbox learning environment for the training and evaluation of RL agents on text-based games. Its generative mechanisms give precise control over the difficulty, scope, and language of constructed games, and can be used to study generalization and transfer learning. This talk will also give an overview of the recent attempts to solve text-based games either using reinforcement learning or more handcrafted approaches.

Abstract 14: Invited Talk: Alison Gopnik in 1st Workshop on Language in Reinforcement Learning (LaReL), Nardelli 01:30 PM

Understanding, learning and reasoning with abstract relations, like same and different or bigger and smaller, is challenging. We show that in an RL like causal learning task, very young children, 18-30 month olds, can learn both same and different relations and the functions becoming bigger and becoming smaller, generalize those relations to brand new and perceptually different objects, and use them to solve novel tasks. We suggest that both abstract causal representations, similar to causal graphical models, and early language may support this knowledge and learning.

Incentives in Machine Learning

Boi Faltings, Yang Liu, David Parkes, Goran Radanovic, Dawn Song

Sat Jul 18, 08:00 AM

Artificial Intelligence (AI), and Machine Learning systems in particular, often depend on the information provided by multiple agents. The most well-known example is federated learning, but also sensor data, crowdsourced human computation, or human trajectory inputs for inverse reinforcement learning. However, eliciting accurate data can be costly, either due to the effort invested in obtaining it, as in crowdsourcing, or due to the need to maintain automated systems, as in distributed sensor systems. Low-quality data not only degrades the performance of AI systems, but may also pose safety concerns. Thus, it becomes important to verify the correctness of data and be smart in how data is aggregated, and to provide incentives to promote effort and high-quality data. During the recent workshop on Federated Learning at NeurIPS 2019, 4 of 6 panel members mentioned incentives as the most important open issue.
This workshop is proposed to understand this aspect of Machine Learning, both theoretically and empirically. We particularly encourage contributions on the following aspects:

- How to collect high quality and credible data for machine learning systems from self-interested and possibly malicious agents, considering the game-theoretical properties of the problem?
- How to evaluate the quality of data supplied by self-interested and possibly malicious agents and how to optimally aggregate it?
- How to make use of machine learning in game-theoretic mechanisms that will facilitate the collection of high-quality data?

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Plenary session</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Bridging Truthfulness and Corruption-Robustness in Multi-Armed Bandit Mechanisms</td>
</tr>
<tr>
<td>N/A</td>
<td>Invited Talk: Incentive-Compatible Forecasting Competitions</td>
</tr>
<tr>
<td>N/A</td>
<td>Invited Talk: Strategic Considerations in Statistical Estimation and Learning</td>
</tr>
<tr>
<td>N/A</td>
<td>Invited Talk: Dominantly Truthful Multi-task Peer Prediction with a Constant Number of Tasks</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Incentivizing and Rewarding High-Quality Data via Influence Functions</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Debiasing Evaluations That are Biased by Evaluations</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Classification with Strategically Withheld Data</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Catch Me if I Can: Detecting Strategic Behaviour in Peer Assessment</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: Linear Models are Robust Optimal Under Strategic Behavior</td>
</tr>
<tr>
<td>N/A</td>
<td>Contributed Talk: From Predictions to Decisions: Using Lookahead Regularization</td>
</tr>
</tbody>
</table>

Please refer to the detailed schedule on the following link:
Abstract 2: Contributed Talk: Bridging Truthfulness and Corruption-Robustness in Multi-Armed Bandit Mechanisms in Incentives in Machine Learning, Abernethy, Kumar, Lykouris, Xu N/A

The talk is available on the following link:

Abstract 3: Invited Talk: Incentive-Compatible Forecasting Competitions in Incentives in Machine Learning, Wilkowski N/A

We consider the design of forecasting competitions in which multiple forecasters make predictions about one or more independent events and compete for a single prize. We have two objectives: (1) to award the prize to the most accurate forecaster, and (2) to incentivize forecasters to report truthfully, so that forecasts are informative and forecasters need not spend any cognitive effort strategizing about reports. Proper scoring rules incentivize truthful reporting if all forecasters are paid according to their scores. However, incentives become distorted if only the best-scoring forecaster wins a prize, since forecasters can often increase their probability of having the highest score by reporting extreme beliefs. In this paper, we introduce a truthful forecaster selection mechanism. We lower-bound the probability that our mechanism selects the most accurate forecaster, and give rates for how quickly this bound approaches 1 as the number of events grows. Our techniques can be generalized to the related problems of outputting a ranking over forecasters and hiring a forecaster with high accuracy on future events.

Abstract 4: Invited Talk: Strategic Considerations in Statistical Estimation and Learning in Incentives in Machine Learning, Chen N/A

Learning and estimation techniques draw insights about unknown underlying relations based on statistical properties of the observed data. Factors that can change the statistical properties of the observed data thus affect the conclusions drawn by these techniques. One such factor is the strategic nature of data holders. The strategic behavior of data holders can alter the observed data even when the underlying relations are unchanged, hence leading to inaccurate conclusions. The cause of the strategic behavior varies from cost of providing data to vested interests in the learning outcomes. In this talk, I will discuss a few attempts to account for the strategic behavior of data holders in statistical estimation and learning. It calls for a more integrated approach for thinking of data and learning.

The talk is available on the following link:

Abstract 5: Invited Talk: Dominantly Truthful Multi-task Peer Prediction with a Constant Number of Tasks in Incentives in Machine Learning, Kong N/A

In the setting where participants are asked multiple similar possibly subjective multi-choice questions (e.g. Do you like Panda Express? Y/N; do you like Chick-fil-A? Y/N), a series of peer prediction mechanisms are designed to incentivize honest reports and some of them achieve dominantly truthfulness: truth-telling is a dominant strategy and strictly dominate other "non-permutation strategy" with some mild conditions. However, a major issue hinders the practical usage of those mechanisms: they require the participants to perform an infinite number of tasks. When the participants perform a finite number of tasks, these mechanisms only achieve approximated dominant truthfulness. The existence of a dominantly truthful multi-task peer prediction mechanism that only requires a finite number of tasks remains to be an open question that may have a negative result, even with full prior knowledge.

This work answers this open question by proposing a new mechanism, Determinant based Mutual Information Mechanism (DMI-Mechanism), that is dominantly truthful when the number of tasks is at least 2C. C is the number of choices for each question (C=2 for binary-choice questions). DMI-Mechanism also pays truth-telling higher than any strategy profile and strictly higher than uninformative strategy profiles (informed truthfulness). In addition to the truthfulness properties, DMI-Mechanism is also easy to implement since it does not require any prior knowledge (detail-free) and only requires at least two participants. The core of DMI-Mechanism is a novel information measure, Determinant based Mutual Information (DMI). DMI generalizes Shannon’s mutual information and the square of DMI has a simple unbiased estimator. In addition to incentivizing honest reports, DMI-Mechanism can also be transferred into an information evaluation rule that identifies high-quality information without verification when there are at least three participants.

Abstract 13: Invited Talk: Follow the money, not the majority: Incentivizing and aggregating expert opinions with Bayesian markets in Incentives in Machine Learning, Baillon N/A

For some questions, such as whether extraterrestrial life exists, it is uncertain if and when the answer will be known. Asking experts for their opinion yields two practical problems. First, how can truth-telling be incentivized if the correct answer is unknowable? Second, if experts disagree, who should be trusted? This paper solves both problems simultaneously. Experts decide whether to endorse a statement and trade an asset whose value depends on the endorsement rate. The respective payoffs of buyers and sellers indicate whom to trust. We demonstrate theoretically and illustrate empirically that "following the money" outperforms selecting the majority opinion.

Abstract 19: Contributed Talk: Incentivizing Bandit Exploration:Recommendations as Instruments in Incentives in Machine Learning, Ngo, Stapleton, Syrgkanis, Wu N/A

The talk is available on the following link:

Abstract 22: Invited Talk: Thwarting Dr. Deceit’s Malicious Activities in Conference Peer Review in Incentives in Machine Learning, Shah N/A

Peer review is an essential part of scientific research, and has a considerable influence on careers of researchers. Hence enter Dr. Deceit, who by various dishonest means, tries to game the peer review system (yes, this does happen in reality). Our goal is to thwart Dr. Deceit’s malicious activities.

Dr. Deceit: As a reviewer, I will manipulate the scores or rankings of the papers that I review in order to increase the chances of my own paper getting accepted. Ha ha ha!

Us: We will use an impartial mechanism, e.g., via a partition-based method, which guarantees a reviewer cannot influence their own paper’s outcome. We show via an analysis on ICLR data that such a mechanism is feasible in conference peer review, despite the complexity and constraints of the conference peer-review process.

Dr. Deceit: But using such a mechanism reduces the efficiency of the process. So if there is no deceitful reviewer like me in the conference, the mechanism will hurt the efficiency of the peer review. Would you...
Us: We can help make that decision -- we design statistical tests to detect the existence of such strategic behavior in peer assessment.

Dr. Deceit: Ok so you will stop me from manipulating my reviews to help my own paper. But I will strike a quid pro quo deal with another potential reviewer for my paper: the reviewer will try to get to review my paper and give a positive review, and in exchange I'll do the same for them in another conference. Your impartial mechanisms can't do anything about this.

Us: We also design randomized reviewer-assignment algorithms which optimally mitigate such arbitrary reviewer-author collusions. Evaluations on data from four conferences show their promise for use in practice.

Dr. Deceit: Fine. I will recruit not one, but multiple such reviewers.

Us: Hmm...then we get into computational-hardness-land. But there is probably some structure on your colluders (e.g., colluding reviewers are at the same institution). Then we have optimal mitigating strategies computable in polynomial time. Keep trying in vain, Dr. Deceit!

Throughout the talk, Dr. Deceit will also throw some more important challenges at us whose solutions are yet unknown.

Abstract 23: Invited Talk: What is my data worth? Towards a Principled and Practical Approach for Data Valuation in Incentives in Machine Learning, Jia N/A

People give massive amounts of their personal data to companies every day and these data are used to generate tremendous business values. Some economists, politicians, and activists argue that people should be paid for their contributions but the million-dollar question is: by how much? In this talk, I will present some recent work on data valuation. I will start by introducing a principled notion for data value and then present a suite of algorithms that we developed to efficiently compute the data value. I will also discuss the applications of our data valuation techniques to the tasks beyond data pricing, such as detecting bad training data.

The talk is available on the following link:

Machine Learning for Media Discovery

Erik Schmidt, Oriol Nieto, Fabien Gouyon, Yves Raimond, Katherine Kinnaird, Gert Lanckriet

Sat Jul 18, 09:10 AM

The ever-increasing size and accessibility of vast media libraries has created a demand more than ever for AI-based systems that are capable of organizing, recommending, and understanding such complex data.

While this topic has received only limited attention within the core machine learning community, it has been an area of intense focus within the applied communities such as the Recommender Systems (RecSys), Music Information Retrieval (MIR), and Computer Vision communities. At the same time, these domains have surfaced nebulous problem spaces and rich datasets that are of tremendous potential value to machine learning and the AI communities at large.

This year’s Machine Learning for Media Discovery (ML4MD) aims to build upon the five previous Machine Learning for Music Discovery editions at ICML, broadening the topic area from music discovery to media discovery. The added topic diversity is aimed towards having a broader conversation with the machine learning community and to offer cross-pollination across the various media domains.

One of the largest areas of focus in the media discovery space is on the side of content understanding. The recommender systems community has made great advances in terms of collaborative feedback recommenders, but these approaches suffer strongly from the cold-start problem. As such, recommendation techniques often fall back on content-based machine learning systems, but defining the similarity of media items is extremely challenging as myriad features all play some role (e.g., cultural, emotional, or content features, etc.). While significant progress has been made, these problems remain far from solved.

In addition, these complex data present many challenges beyond the development of machine learning systems to model and understand them. One of the largest challenges is scale. One example is commercial music libraries, which span into the tens of millions. However, user-generated content platforms such as YouTube and Pinterest have libraries stretching into the billions—a scale at which many of the traditional approaches discussed in the literature simply cannot perform.

On the other side of this problem sits the recent explosion of work in the area of Creative AI. Relevant examples include Google Magenta, Amazon’s DeepComposer, who seek to develop algorithms capable of composing and performing completely original (and compelling) works of music. The same also happens in the world of visual media creation (e.g., DeepDream, Deep Fakes). Certain work in this area adds an interesting dimension to the conversation as understanding how content is created is a prerequisite to generating.

This workshop proposal is timely in that it will bridge these separate pockets of otherwise very related research. In addition to making progress on the challenges above, we hope to engage the wide AI and machine learning community with our rich problem space, and connect them with the many available datasets the community has to offer.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Welcome Remarks</td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Graph Neural Networks for Reasoning over Multimodal Content</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Novel Audio Embeddings for Personalized Recommendations on Newly Released Tracks</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Musical Word Embedding: Bridging the Gap between Listening Contexts and Music</td>
</tr>
<tr>
<td>10:20 AM</td>
<td>Poster Session #1</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Graphs for music analysis</td>
</tr>
<tr>
<td>Time</td>
<td>Session</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Deep Active Learning Toward Crisis-related Tweets Classification</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>The Unsung Heroes of Music Recommendation: an Essay</td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Beyond Being Accurate: Solving Real-World Recommendation Problems with Neural Modeling</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Character-focused Video Thumbnail Retrieval</td>
</tr>
<tr>
<td>01:50 PM</td>
<td>HitPredict: Using Spotify Data to Predict Billboard Hits</td>
</tr>
<tr>
<td>02:10 PM</td>
<td>Poster Session #2</td>
</tr>
<tr>
<td>02:50 PM</td>
<td>Hit Song Prediction</td>
</tr>
<tr>
<td>03:20 PM</td>
<td>I know why you like this movie: Interpretable Efficient Multimodal Recommender</td>
</tr>
<tr>
<td>03:40 PM</td>
<td>Content-based Music Similarity with Siamese Networks</td>
</tr>
</tbody>
</table>

Abstracts (2):

Abstract 5: **Poster Session #1 in Machine Learning for Media Discovery**, 10:20 AM

Both poster sessions will feature the full poster program.

**Bach or Mock? A Grading Function for Chorales in the Style of J.S. Bach**

Alexander Fang, Alisa Liu, Prem Seetharaman and Bryan Pardo

Topic: ML4MD Poster: Fang et al.

https://netflix.zoom.us/j/99893045607?pwd=RHRzeUZueGhYMVo1Rzk0V3NiSHV11F09

Meeting ID: 998 9304 5607

Password: ML4MD

**Incorporating Music Knowledge in Continual Dataset Augmentation for Music Generation**

Alisa Liu, Alexander Fang, GaÅ-Etian Hadjeres, Prem Seetharaman and Bryan Pardo

Topic: ML4MD Poster: Liu et al.

https://netflix.zoom.us/j/96351571989?pwd=UmtrUlSpSTkt1UEhHUVR1V1NzL0xPQT09

Meeting ID: 963 5157 1989

Password: ML4MD

**Generating Modelling for Controllable Audio Synthesis of Expressive Piano Performance**

Hao Hao Tan, Yin-Jyun Luo and Dorien Herremans

Topic: ML4MD Poster: Fang et al.

https://netflix.zoom.us/j/94487939901?pwd=TXBFNU1Zd0FCZY9sQlzZktMUTUSQQT09

Meeting ID: 944 8793 9901

Password: ML4MD

**Web Interface for Exploration of Latent and Tag Spaces in Music Auto-Tagging**

Philip Tovstogan, Xavier Serra and Dmitry Bogdanov

Topic: ML4MD Poster: Tovstogan et al.

https://netflix.zoom.us/j/91482530587?pwd=dmZnV1d5a221RlZWa0Vv5F2SVBU5UT09

Meeting ID: 914 8253 0587

Password: ML4MD
Password: ML4MD

Cosine Similarity of Multimodal Content Vectors for TV Programmes

Saba Nazir, Taner Cagali, Chris Newell and Mehrnoosh Sadrzadeh

Topic: ML4MD Poster: Nazir et al.
https://netflix.zoom.us/j/99906337836?pwd=WmBBVVFvMEU5KzhhSUpEeFRRb0lyQT09
Meeting ID: 999 0633 7836
Password: ML4MD

Artist biases in collaborative filtering for music recommendation

Andres Ferraro, Jae Ho Jeon, Biho kim, Xavier Serra and Dmitry Bogdanov

Topic: ML4MD Poster: Ferraro et al.
https://netflix.zoom.us/j/91972151215?pwd=QllVZ2QxTUxmaCtWZmdyRks4UW55UT09
Meeting ID: 919 7215 1215
Password: ML4MD

Self-Correcting Non-Chronological Autoregressive Music Generation

Wayne Chi, Prachi Kumar, Suri Yaddanapudi, Rahul Suresh and Umut Isik

Topic: ML4MD Poster: Chi et al.
https://netflix.zoom.us/j/91455879996?pwd=VTUvK0szb093YnlGRTFCYWydaZEZK
Meeting ID: 914 5587 9996
Password: ML4MD

Discovering X Degrees of Keyword Separation in a Fine Arts Collection

Arthur Flexer

Topic: ML4MD Poster: Flexer
https://netflix.zoom.us/j/94547050977?pwd=d3Z6NVJ6ZFRkV0lWVjWvdj1IDUT09
Meeting ID: 945 4705 0977
Password: ML4MD

Abstract 13: Poster Session #2 in Machine Learning for Media Discovery, 02:10 PM

Both poster sessions will feature the full poster program.

Bach or Mock? A Grading Function for Chorales in the Style of J.S. Bach

Alexander Fang, Alisa Liu, Prem Seetharaman and Bryan Pardo

Topic: ML4MD Poster: Fang et al.
https://netflix.zoom.us/j/99893045607?pwd=RHRRzEUZueGhYMVo1Rzk0V3NiSHV1UT09
Meeting ID: 998 9304 5607
Password: ML4MD

Generative Modelling for Controllable Audio Synthesis of Expressive Piano Performance

Hao Hao Tan, Yin-Jyun Luo and Dorien Herremans

Topic: ML4MD Poster: Tan et al.
https://netflix.zoom.us/j/94487939901?pwd=TXBFNU12d0FCZy9sQlzZktMUtUSQ09
Meeting ID: 944 8793 9901
Password: ML4MD

Incorporating Music Knowledge in Continual Dataset Augmentation for Music Generation

Alisa Liu, Alexander Fang, Gaël Hadjeres, Prem Seetharaman and Bryan Pardo

Topic: ML4MD Poster: Liu et al.
https://netflix.zoom.us/j/91482530587?pwd=d3ZnVtds2Z1RjZWaq055eSE2SVWUT09
Meeting ID: 914 8253 0587
Password: ML4MD

Web Interface for Exploration of Latent and Tag Spaces in Music Auto-Tagging

Philip Tovstogan, Xavier Serra and Dmitry Bogdanov

Topic: ML4MD Poster: Tovstogan et al.
https://netflix.zoom.us/j/91482530587?pwd=d3ZnVtds2Z1RjZWaq055eSE2SVWUT09
Meeting ID: 914 8253 0587
Password: ML4MD
### Cosine Similarity of Multimodal Content Vectors for TV Programmes

Saba Nazir, Taner Cagali, Chris Newell and Mehrnoosh Sadrzadeh

Topic: ML4MD Poster: Nazir et al.

https://netflix.zoom.us/j/99906337836?pwd=WnBBVFVxMEI5KzhhSUJpEeFRb0lyQz09

Meeting ID: 999 0633 7836

Password: ML4MD

### Self-Correcting Non-Chronological Autoregressive Music Generation

Wayne Chi, Prachi Kumar, Suri Yaddanapudi, Rahul Suresh and Umut Isik

Topic: ML4MD Poster: Chi et al.

https://netflix.zoom.us/j/91455879996?pwd=VTUvK0szb093YniGRTFCYWhdaZEZjZ09

Meeting ID: 914 5587 9996

Password: ML4MD

---

**2nd ICML Workshop on Human in the Loop Learning (HILL)**

**Shanghang Zhang, Xin Wang, Fisher Yu, Jiajun Wu, Prof. Darrell**

**Sat Jul 18, 11:00 AM**

Recent years have witnessed the rising need for learning agents that can interact with humans. Such agents usually involve applications in computer vision, natural language processing, human computer interaction, and robotics. Creating and running such agents call for interdisciplinary research of artificial intelligence, machine learning, and software engineering design, which we abstract as Human in the Loop Learning (HILL). HILL is a modern machine learning paradigm of significant practical and theoretical interest. For HILL, models and humans engage in a two-way dialog to facilitate more accurate and interpretable learning. The workshop aims to bring together researchers and practitioners working on the broad areas of human in the loop learning, ranging from the interactive/active learning algorithms designed for real-world decision making systems (e.g., autonomous driving vehicles, robotic systems, etc.), models with strong explainability, as well as interactive system designs (e.g., data visualization, annotation systems, etc.). In particular, we aim to elicit new connections among these diverse fields, identifying theory, tools and design principles tailored to practical machine learning workflows. The target audience for the workshop includes people who are interested in using machines to solve problems by having a human be an integral part of the learning process. In this yearâ€™s HILL workshop, we emphasize on the interactive/active learning algorithms for real-world decision making systems as well as learning algorithms with strong explainability. We continue the previous effort to provide a platform for researchers to discuss approaches that bridge the gap between humans and machines and get the best of both worlds. We believe the theme of the workshop will be interesting to ICML attendees, especially those who are interested in interdisciplinary study.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00 AM</td>
<td>Opening Remarks</td>
<td>Zhang</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Invited Talk 1: Prof. Zeynep Akata from University of Tà¼bingen</td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Invited Talk 1-QA</td>
<td></td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Invited Talk 2: Prof. Tom Griffiths from Princeton University</td>
<td>Griffiths</td>
</tr>
<tr>
<td>12:40 PM</td>
<td>Invited Talk 2-QA</td>
<td></td>
</tr>
<tr>
<td>12:50 PM</td>
<td>Invited Talk 3: Prof. Christian Lebiere from Carnegie Mellon University</td>
<td>Lebiere</td>
</tr>
<tr>
<td>01:20 PM</td>
<td>Invited Talk 3-QA</td>
<td></td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Poster Session 1 with Zoom meeting links for the accepted papers</td>
<td></td>
</tr>
<tr>
<td>02:10 PM</td>
<td>Invited Talk 4: Prof. Richard Zemel from University of Toronto</td>
<td>Zemel</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Invited Talk 4-QA</td>
<td></td>
</tr>
<tr>
<td>02:50 PM</td>
<td>Invited Talk 5: Prof. Pradeep Ravikumar from Carnegie Mellon University</td>
<td></td>
</tr>
<tr>
<td>03:20 PM</td>
<td>Invited Talk 5-QA</td>
<td></td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Invited Talk 6: Prof. Raquel Urtasun from University of Toronto</td>
<td></td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Invited Talk 6-QA</td>
<td></td>
</tr>
<tr>
<td>04:10 PM</td>
<td>Invited Talk: Dr. Kalesha Bullard from Facebook AI Research</td>
<td>Bullard</td>
</tr>
<tr>
<td>04:40 PM</td>
<td>Invited Talk-Kalesha-QA</td>
<td></td>
</tr>
<tr>
<td>05:00 PM</td>
<td>Invited Talk 7: Prof. Anca Dragan from UC Berkeley</td>
<td>Dragan</td>
</tr>
<tr>
<td>05:20 PM</td>
<td>Invited Talk 7-QA</td>
<td></td>
</tr>
<tr>
<td>05:30 PM</td>
<td>Poster Session 2 with Zoom meeting links for the accepted papers</td>
<td></td>
</tr>
<tr>
<td>06:10 PM</td>
<td>Invited Talk 9: Prof. Sergey Levine from UC Berkeley</td>
<td>Levine</td>
</tr>
<tr>
<td>06:40 PM</td>
<td>Invited Talk 9-QA</td>
<td></td>
</tr>
</tbody>
</table>
06:50 PM Invited Talk 10: Prof. Wenwu Zhu from Tsinghua University

07:20 PM Invited Talk 10-QA

07:30 PM Invited Talk 11: Prof. Chelsea Finn from Stanford University

08:10 PM Poster Session 3 with Zoom meeting links for the accepted papers

08:40 PM Closing Remarks

Abstracts (12):

Abstract 2: Invited Talk 1: Prof. Zeynep Akata from University of Tübingen in 2nd ICML Workshop on Human in the Loop Learning (HILL), 11:30 AM
Pre-recorded talk video is available at:
https://slideslive.com/38930827/invited-talk-1

Abstract 4: Invited Talk 2: Prof. Tom Griffiths from Princeton University in 2nd ICML Workshop on Human in the Loop Learning (HILL), Griffiths 12:10 PM
Pre-recorded talk video is available at:
https://slideslive.com/38930828/predicting-and-understanding-human-decisions

Abstract 8: Poster Session 1 with Zoom meeting links for the accepted papers in 2nd ICML Workshop on Human in the Loop Learning (HILL), 01:30 PM
Zoom meeting links for the poster sessions can be found in the shared Google Doc:
https://docs.google.com/spreadsheets/d/1gegBSjv8Si766Mzh01yczfj4ExUj4-4X8wUa4mI3xQ/edit?usp=sharing
or
https://nqfnr2ysmo.feishu.cn/sheets/shtcnXo7d0IgLb2N0YBENdzsN8gg
(Same content with the google doc)
If there are Zoom meetings you cannot access, please comment on the google doc beside these meetings' links

Abstract 9: Invited Talk 4: Prof. Richard Zemel from University of Toronto in 2nd ICML Workshop on Human in the Loop Learning (HILL), Zemel 02:10 PM
Pre-recorded talk video is available at:
https://slideslive.com/38930830/wandering-within-a-world-online-contextualized-few-shot-learning

Abstract 11: Invited Talk 5: Prof. Pradeep Ravikumar from Carnegie Mellon University in 2nd ICML Workshop on Human in the Loop Learning (HILL), 02:50 PM
Pre-recorded talk video is available at:

Abstract 13: Invited Talk 6: Prof. Raquel Urtasun from University of Toronto in 2nd ICML Workshop on Human in the Loop Learning (HILL), 03:30 PM
Pre-recorded talk video is available at:
https://slideslive.com/38930835/how-should-we-train-our-robots

Abstract 15: Invited Talk 7: Prof. Anca Dragan from UC Berkeley in 2nd ICML Workshop on Human in the Loop Learning (HILL), Dragan 04:50 PM
Pre-recorded talk video is available at:
https://slideslive.com/38930935/invited-talkkalesha

Abstract 17: Invited Talk 7: Prof. Anca Dragan from UC Berkeley in 2nd ICML Workshop on Human in the Loop Learning (HILL), Dragan 04:50 PM
Pre-recorded talk video is available at:

Abstract 19: Poster Session 2 with Zoom meeting links for the accepted papers in 2nd ICML Workshop on Human in the Loop Learning (HILL), 05:30 PM
Zoom meeting links for the poster sessions can be found in the shared Google Doc:
https://docs.google.com/spreadsheets/d/1gegBSjv8Si766Mzh01yczfj4ExUj4-4X8wUa4mI3xQ/edit?usp=sharing
or
https://nqfnr2ysmo.feishu.cn/sheets/shtcnXo7d0IgLb2N0YBENdzsN8gg
(Same content with the google doc)
If there are Zoom meetings you cannot access, please comment on the google doc beside these meetings' links

Abstract 20: Invited Talk 8: Prof. Sergey Levine from UC Berkeley in 2nd ICML Workshop on Human in the Loop Learning (HILL), Levine 06:10 PM
Pre-recorded talk video is available at:
https://slideslive.com/38930835/how-should-we-train-our-robots

Abstract 22: Invited Talk 10: Prof. Wenwu Zhu from Tsinghua University in 2nd ICML Workshop on Human in the Loop Learning (HILL), zhu 06:50 PM
Pre-recorded talk video is available at:

Abstract 25: Poster Session 3 with Zoom meeting links for the accepted papers in 2nd ICML Workshop on Human in the Loop Learning (HILL), 08:10 PM
Zoom meeting links for the poster sessions can be found in the shared Google Doc:
https://docs.google.com/spreadsheets/d/1gegBSjv8Si766Mzh01yczfj4ExUj4-4X8wUa4mI3xQ/edit?usp=sharing
or
https://nqfnr2ysmo.feishu.cn/sheets/shtcnXo7d0IgLb2N0YBENdzsN8gg
(Same content with the google doc)
Zoom meeting links for the poster sessions can be found in the shared Google Doc:

https://docs.google.com/spreadsheets/d/1gegBSjv8Sf766Mzh01yczfJ4ExU4-4X8wUa4mlGxQ/edit?usp=sharing

or

https://nqfr2ysmo.feishu.cn/sheets/shtcno7d0lGbo2NCYBEndzsN8gg
(Same content with the google doc)

If there are Zoom meetings you cannot access, please comment on the google doc beside these meetings’ links