Action-Sufficient State Representation Learning for Control with Structural Constraints

Presented by: Biwei Huang

Biwei Huang

Chaochao Lu

Leqi Liu

José Miguel Hernández-Lobato

Clark Glymour

Bernhard Schölkopf

Kun Zhang

What low-dimensional representation to find & why:

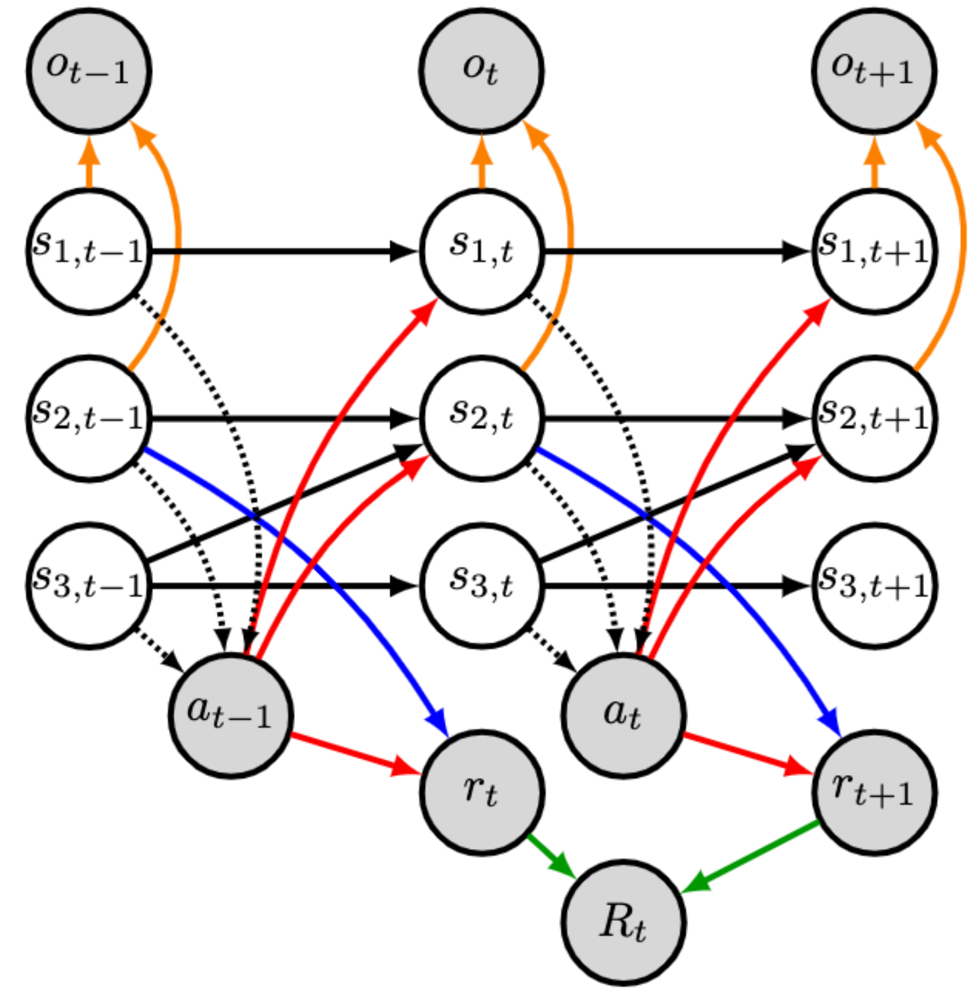
- Perceived signals are usually high-dimensional, with irrelevant information for decision-making
- Finding and using essential and sufficient information helps **improve computational efficiency and generalization ability**

Decision on when to cross relies on the color of traffic lights, which can be represented by a one-dimensional var.

Contributions:

Action-Sufficient state Representations (ASRs): learning a minimal set of state representations that capture sufficient information for decision making

Environment model with structural constraints:



$$\begin{cases} o_t = f(D_{\vec{s} \to o} \odot \vec{s}_t, e_t), \\ r_t = g(D_{\vec{s} \to r} \odot \vec{s}_{t-1}, D_{a \to r} \odot a_{t-1}, \epsilon_t), \\ s_{i,t} = h_i(D_{\vec{s}(\cdot,i)} \odot \vec{s}_{t-1}, D_{a \to \vec{s}(\cdot,i)} \odot a_{t-1}, \eta_{i,t}), \end{cases}$$

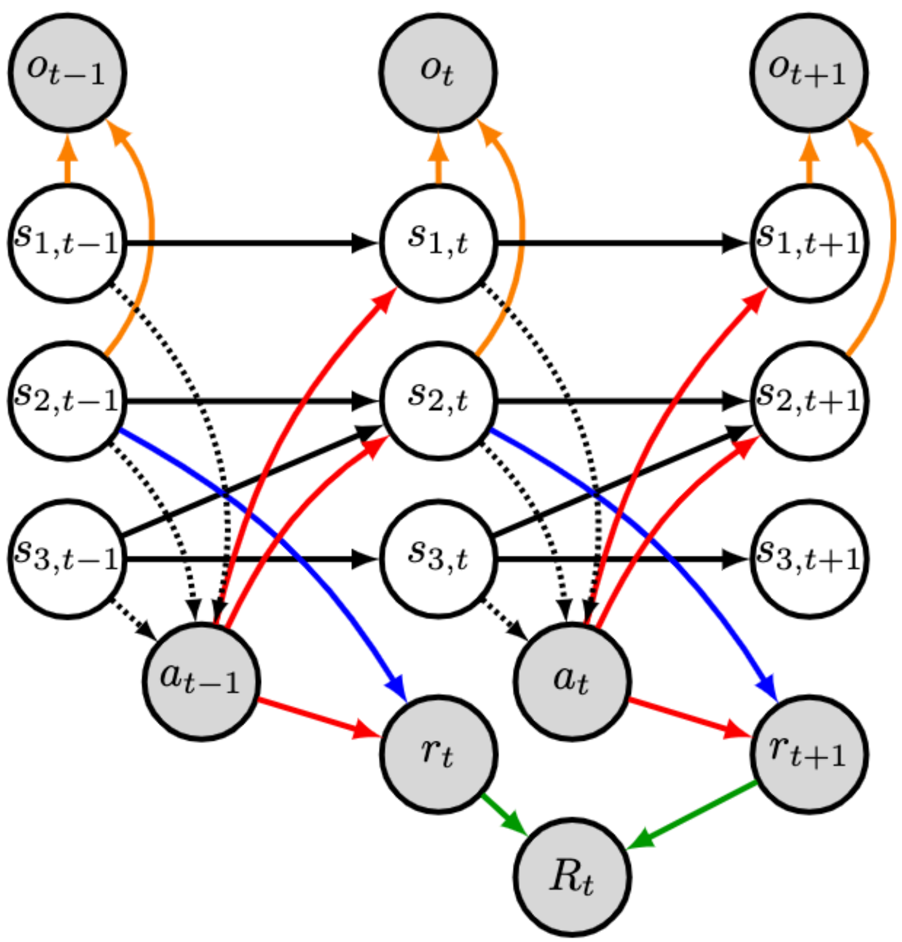
for $i = 1, \cdots, d$.

- **Encode structural relationships with masks $D_{\cdot \to \cdot}$ to characterize independence constraints**, including the structure
  - over different dimensions of $\vec{s}_t$,
  - from the action variable a_{t-1} to different dimensions of s_t, and
  - from different dimensions of s_{t-1} to the reward r_t

Minimal sufficient state representations for control

Graphical view: $s_{i,t}$ is in ASRs iff it satisfies one of the following conditions:

- $s_{i,t} \in \vec{s}_t^{\text{ASR}}$ has an edge to the reward in the next time-step r_{t+1}, or
- $s_{i,t} \in \vec{s}_t^{\text{ASR}}$ has an edge to another state dimension in the next time-step $s_{j,t+1}$, such that the same component at time t is in ASRs, i.e., $s_{j,t} \in \vec{s}_t^{\text{ASR}}$.



$$\vec{s}_t^{\text{ASR}} = (s_{2,t}, s_{3,t})^\top$$

Minimal sufficient state representations for control

Graphical view: $s_{i,t}$ is in ASRs iff it satisfies one of the following conditions:

- $s_{i,t} \in \vec{s}_t^{\text{ASR}}$ has an edge to the reward in the next time-step r_{t+1}, or
- $s_{i,t} \in \vec{s}_t^{\text{ASR}}$ has an edge to another state dimension in the next time-step $s_{j,t+1}$, such that the same component at time t is in ASRs, i.e., $s_{j,t} \in \vec{s}_t^{\text{ASR}}$.

Conditional independence view: $s_{i,t}$ is in ASRs if and only if

$$s_{i,t} \not\perp R_{t+1} | a_{t-1:t}, \vec{s}_{t-1}^{\text{ASR}}$$

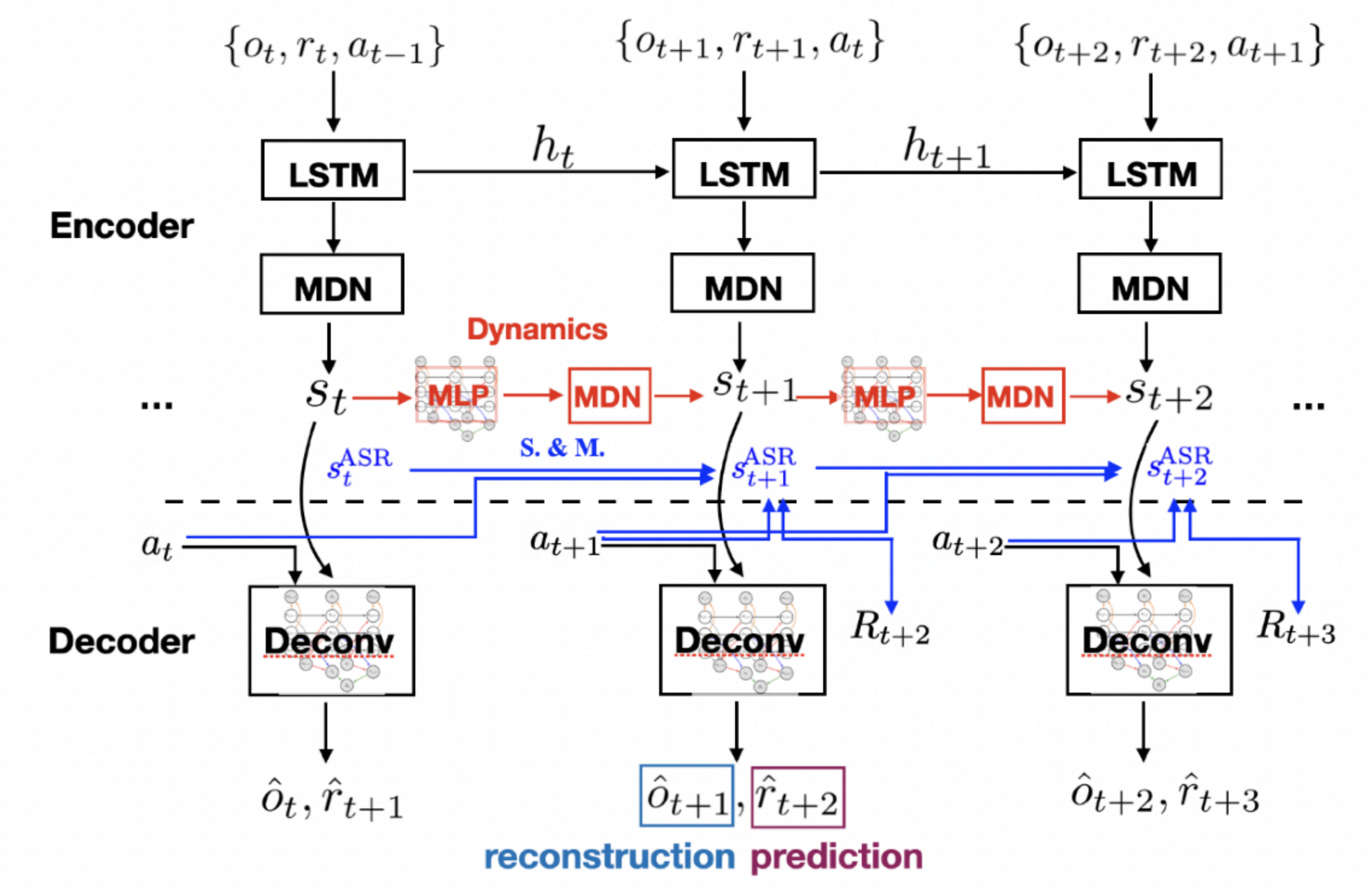
Objective for ASRs: Learn the ASRs by maximizing the following mutual information

$$I(\tilde{\vec{s}}^{\text{ASR}}; R_{t+1} | a_{t-1:t}, \tilde{\vec{s}}_{t-1}^{\text{ASR}}) - I(\tilde{\vec{s}}^{\text{C}}; R_{t+1} | a_{t-1:t}, \tilde{\vec{s}}_{t-1}^{\text{ASR}}),$$

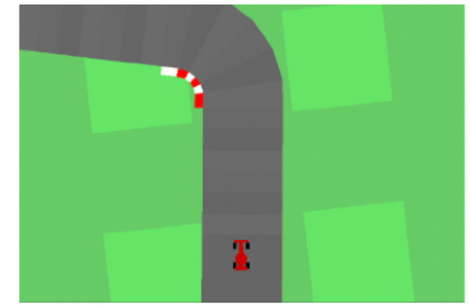
where $\tilde{\vec{s}}^{\text{C}} = \tilde{\vec{s}} \backslash \tilde{\vec{s}}^{\text{ASR}}$.

Structured Sequential VAE for estimation of ASRs

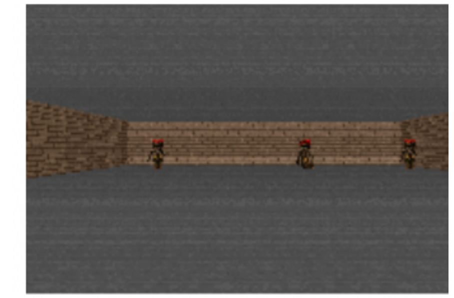
- **"Sequential VAE" component:** with LSTM & MLP
- **"Structured" component:** structural information encoded with binary masks
- **Sufficiency & minimality constraint**
- **Sparsity constraints:** sparsity constraints on structural matrices to achieve better identifiability, according to the edge-minimality property



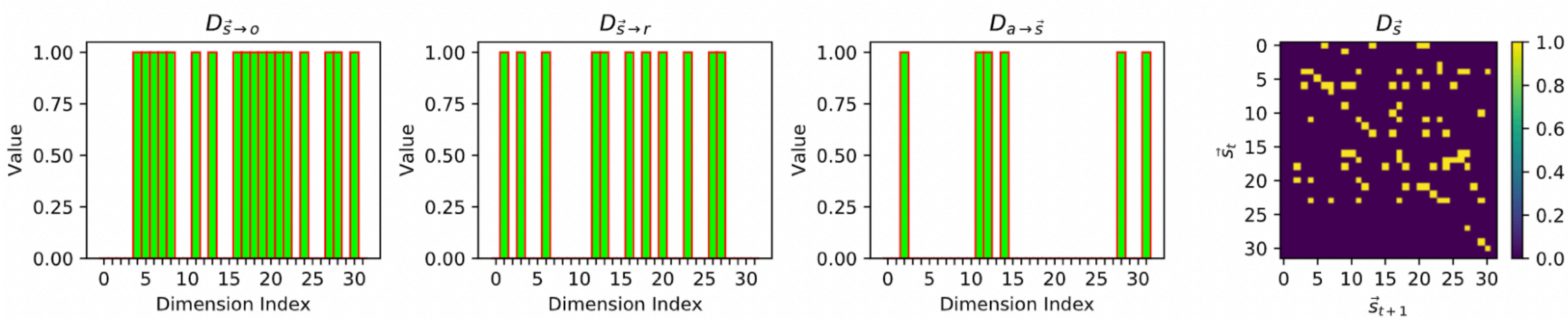
Evaluations on Car Racing & Vizdoom



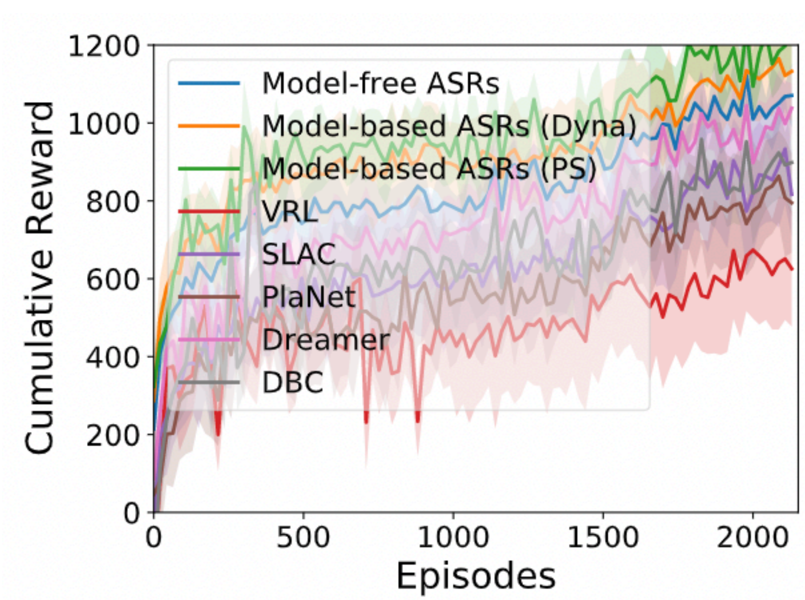
Carracing



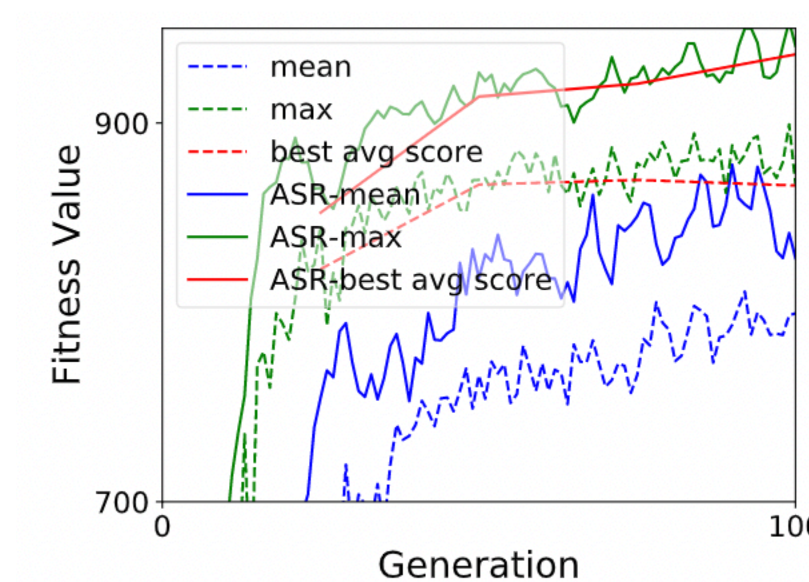
Vizdoom



Visualization of estimated structural matrices $D_{\vec{s}\to o}$, $D_{\vec{s}\to r}$, $D_{a\to\vec{s}}$, and $D_{\vec{s}}$ in Car Racing



Cumulative rewards compared
with state-with-the-art methods



Fitness Value of ASRs
compared to world models

Advantages of ASR-Based Approach:

- Structural information provides an interpretable and intuitive picture of the generating process
- and an interpretable and intuitive way to characterize a minimal sufficient set of state representations for policy learning
- No information loss when representation learning and policy learning are done separately, which is computationally more efficient
- Flexible to use a wide range of policy learning methods, including model-based RL that effectively reduces possibly risky explorations