Combating Label Noise in Deep Learning using Abstention

Speaker: Sunil Thulasidasan
sunil@lanl.gov
A Practical Challenge for Deep Learning

State-of-the-art models require *large amounts of clean, annotated data.*
Annotation is labor intensive!

ImageNet: 15 million labeled images; over 20,000 classes

The data that transformed AI research—and possibly the world (D. Gershgorn, quartz, magazine, 2017)

- 49k workers
- 167 countries
- 2.5 years to complete!
Approaches to large-scale labeling

• Crowdsourcing at scale – labor intensive, but relatively cheap

• Use weak labels from queries, user tags and pre-trained classifiers
Approaches to large-scale labeling

- Crowdsourcing at scale – labor intensive, but cheap
- Use weak labels from queries, user tags and pre-trained classifiers

Both approaches can lead to significant labeling errors!
• Label noise is an inconsistent mapping from features $X$ to labels $Y$
The Deep Abstaining Classifier (DAC)

**Approach:** Use learning difficulty on incorrectly labeled or confusing samples to defer on learning -- “abstain” -- till correct mapping is learned.
Training a Deep Abstaining Classifier

$$\mathcal{L}(x) = (1 - p(x)_{k+1}) \left( - \sum_{i=1}^{k} t(x)_i \log \frac{p(x)_i}{1 - p(x)_{k+1}} \right) + \alpha \log \frac{1}{1 - p(x)_{k+1}}$$

Cross entropy as usual
Training a Deep Abstaining Classifier

$$\mathcal{L}(x) = (1 - p(x)_{k+1}) \left( - \sum_{i=1}^{k} t(x)_i \log \frac{p(x)_i}{1 - p(x)_{k+1}} \right) + \alpha \log 1 - p(x)_{k+1}$$

Encourages abstention

Cross entropy over actual classes
Training a Deep Abstaining Classifier

\[
\mathcal{L}(x) = (1 - p(x)_{k+1}) \left( - \sum_{i=1}^{k} t(x)_i \log \frac{p(x)_i}{1 - p(x)_{k+1}} \right) + \alpha \log \frac{1}{1 - p(x)_{k+1}}
\]

- Encourages abstention
- Cross entropy over actual classes
- Penalizes abstention

Abstention class

Automatically tuned during learning.
Abstention Dynamics

Abstained percent on training set vs epoch with 10% label noise.

Introduce abstention after a warmup period.

Abstention reduces as the DAC makes learning progress.

Ideal rate of abstention

Overfitting regime!
The DAC gives state-of-art results in label-noise experiments.

Training protocol:
- Use DAC to identify and eliminate label noise.
- Retrain on cleaner set.

GCE: Generalized Cross-Entropy Loss (Zhang et al NIPS ‘18); Forward (Patrini et al, CVPR ‘17); MentorNet (Li et al, ICML ‘18)
Abstention in the presence of Systematic Label Noise: The Random Monkeys Experiment

All the monkey labels in the training set (STL-10) are randomized.

Can the DAC learn that images containing monkey features have unreliable labels and abstain on monkeys in the test set?
Random Monkeys: DAC Predictions on Monkey Images

The DAC abstains on most of the monkeys in the test set!
Image Blurring

Blur a subset (20%) of the images in the training set and randomize labels.

Will the DAC learn to abstain on blurred images in the test set?
DAC Behavior on Blurred Images

DAC abstains on most of the blurred images in the test set.

For DAC, validation accuracy is calculated on non-abstained samples.
Conclusions

• Abstention training is an effective way to clean label noise in a deep learning pipeline.
  
  • Abstention can also be used as a representation learner for label noise.
    • Especially useful for interpretability in “don’t-know” decision situations.

Code available at https://github.com/thulas/dac-label-noise
Code available at https://github.com/thulas/dac-label-noise

Joint work with……..

Tanmoy Bhattacharya
Los Alamos National Lab

Jeff Bilmes
University of Washington

Gopinath Chennupati
Los Alamos National Lab

Jamal Mohd-Yusof
Los Alamos National Lab

Poster:
Tue Jun 11th
06:30 -- 09:00 PM @ Pacific Ballroom #9

Point of Contact:
Sunil Thulasidasan
(sunil@lanl.gov)