Semi-Supervised Active Learning with Temporal Output Discrepancy

Siyu Huang, Tianyang Wang, Haoyi Xiong, Jun Huan, Dejing Dou

Baidu Research & Austin Peay State University & Styling AI
Active Learning

Diversity-aware Approaches

Uncertainty-aware Approaches
- Margin between posterior probabilities
- Entropy of posterior probabilities
- Expected model change
- Expected error reduction
- Distance to decision boundary
- Adversarial discrimination

Hybrid Methods

Settles 2010
Active Learning

Diversity-aware Approaches

Uncertainty-aware Approaches
- Margin between posterior probabilities
- Entropy of posterior probabilities
- Expected model change
- Expected error reduction
- Distance to decision boundary
- Adversarial discrimination

Hybrid Methods

Settles 2010
Loss Estimation via *Temporal Output Discrepancy (TOD)*

- **Temporal Output Discrepancy (TOD):**
  The discrepancy of outputs of a neural network $f$ at different GD steps.

\[
D_t^{[T]}(x) \overset{\text{def}}{=} \| f(x; w_{t+T}) - f(x; w_t) \|
\]
Why can TOD estimate sample loss?

*With an appropriate setting of learning rate $\eta$, we have*

*• Theorem 1*

\[
D_t^{(1)}(x) \leq \eta \sqrt{2L_t(x)} \| \nabla_w f(x; w_t) \|^2
\]

*• Corollary 2*

\[
D_t^{(T)}(x) \leq \sqrt{2T \eta C} \sqrt{\sum_{\tau=t}^{t+T-1} L_{\tau}(x)}
\]
TOD-based Active Learning

• Cyclic Output Discrepancy (COD) as data sampling strategy
TOD-based Active Learning

- Semi-supervised task training

Baseline model

Mean Teacher [Tarvainen and Harri, NeurIPS 2017]
The exponential moving average of historical model parameters
Experimental Comparison of Active Learning Methods

Active learning performance on four datasets

<table>
<thead>
<tr>
<th>Method</th>
<th>Cifar10</th>
<th>SVHN</th>
<th>Caltech-101</th>
<th>Extra model?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreset (ICLR'18)</td>
<td>91.4s</td>
<td>168.7s</td>
<td>48.2s</td>
<td>✗</td>
</tr>
<tr>
<td>VAAL (ICCV'19)</td>
<td>13.0s</td>
<td>17.2s</td>
<td>32.6s</td>
<td>✓</td>
</tr>
<tr>
<td>LL4AL (CVPR'19)</td>
<td>7.7s</td>
<td>10.8s</td>
<td>39.6s</td>
<td>✓</td>
</tr>
<tr>
<td>Ours</td>
<td>7.2s</td>
<td>10.1s</td>
<td>26.9s</td>
<td>✗</td>
</tr>
</tbody>
</table>

Efficiency of active sampling strategies
(One active sampling iteration)
Summary

• Temporal Output Discrepancy (TOD) estimates the loss of unlabeled samples by evaluating the discrepancy of neural network outputs at different GD steps.

• TOD is a lower-bound of accumulated sample loss.

• Based on TOD, we develop an unlabeled data sampling strategy (COD) and a semi-supervised training scheme for active learning.

https://github.com/siyuhuang/TOD