Negative Transfer in Deep Multi-task Learning

Shengchao Liu
Yingyu Liang
Anthony Gitter
Multi-task Learning Roadmap

Transfer Learning:

Transfer knowledge from source tasks to target tasks

Multi-task Learning:

Model the task relatedness

Learn all tasks simultaneously

Learn shared representation
Deep Multi-task Learning

(Deep) Single-task learning

(Deep) Multi-task learning

Negative Transfer
Reinforced Multi-task Learning (RMTL)

Reinforcement Learning

1. policy
2. reward function
**Algorithm 3: Reward Function**

Given a focused task.

Initialize neural network $W$, with depth $d$.

For batch of data $B$ do

For all candidate policy $\pi$. do

Get gradient w.r.t. batch of data on the last shared layer $\nabla_{d-1}B$.

Compute cosine-similarity $v$ w.r.t. the focused task based on $\nabla_{d-1}B$.

Fetch reward, $r = \max(v, 0)$.

End for

Select best $\pi$ according to $r$.

Multiply gradients $\nabla B$ by $\pi \cdot r$.

Update $W$ by backpropagation.

End for
Experiments: PCBA

3 similar pairs from 128 tasks
Appendix

1. More Focused Learning (RMTL) results, randomly sampled pairs
2. Negative transfer in 128 PCBA tasks
3. GradNorm: GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks
4. Pre-training and Fine-tuning
5. Plottings: STL, MTL, GradNorm, Focused (RMTL), FineTuning
<table>
<thead>
<tr>
<th>Transfer Learning Settings</th>
<th>Related Areas</th>
<th>Source Domain Labels</th>
<th>Target Domain Labels</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inductive Transfer Learning</strong></td>
<td>Multi-task Learning</td>
<td>Available</td>
<td>Available</td>
<td>Regression, Classification</td>
</tr>
<tr>
<td></td>
<td>Self-taught Learning</td>
<td>Unavailable</td>
<td>Available</td>
<td>Regression, Classification</td>
</tr>
<tr>
<td><strong>Transductive Transfer Learning</strong></td>
<td>Domain Adaptation, Sample Selection Bias, Co-variate Shift</td>
<td>Available</td>
<td>Unavailable</td>
<td>Regression, Classification</td>
</tr>
<tr>
<td><strong>Unsupervised Transfer Learning</strong></td>
<td></td>
<td>Unavailable</td>
<td>Unavailable</td>
<td>Clustering, Dimensionality Reduction</td>
</tr>
</tbody>
</table>

![Diagram of Transfer Learning settings and related tasks]

- **Inductive Transfer Learning**
  - Multi-task Learning: Available Labels (Source), Available Labels (Target) → Regression, Classification
  - Self-taught Learning: Unavailable Labels (Source), Available Labels (Target) → Regression, Classification

- **Transductive Transfer Learning**
  - Domain Adaptation, Sample Selection Bias, Co-variate Shift: Available Labels (Source), Unavailable Labels (Target) → Regression, Classification

- **Unsupervised Transfer Learning**
  - Sample Selection Bias / Covariance Shift: Unavailable Labels (Source), Unavailable Labels (Target) → Clustering, Dimensionality Reduction