Deep Generative and Discriminative Domain Adaptation

Han Zhao†, Junjie Hu†, Zhenyao Zhu*, Adam Coates†, & Geoffrey J. Gordon†
†Carnegie Mellon University, *Google, ‡Apple
{han.zhao, junjieh, gordon}@cs.cmu.edu, zhuzychn@gmail.com, acoates@cs.stanford.edu

AAMAS-2019

May 15, 2019

han.zhao@cs.cmu.edu

Summary
Unsupervised Domain adaptation: Source ≠ Target

Q: What if we also have unlabeled data from source domain? Can we utilize them? A: By generative modeling.
Q: Is it possible to extend domain adaptation to time-series modeling? A: Frame-wise extension.

Our Approach
Domain adversarial Auto-encoder (DAuto):

Our Approach

1. Source Images with Labels
2. Target Images with Labels
3. Adaptable Function

Joint likelihood maximization:

\[ \max_{\psi, \phi} \sum_{i=1}^{m} \log p(y_i \mid x_i, \psi) + \lambda \sum_{i=1}^{n} -||x_i - g(f(x_i), \phi) ||_2^2 \]

- \( p(x) \) given by kernel density estimation

Overall objective function:

\[ \min_{W_D, W_s} \max_{W_i, W_k} \sum_{i=1}^{m} L_p(x_i, y_i; W_f, W_k) \]

- \( L_p \): classification/regression loss
- \( L_i \): reconstruction loss
- \( L_c \): binary classification loss from domain classifier

Analysis: With probability \( \geq 1 - \delta \), \( \forall h \),

\[ \text{err}_y(h) \leq \text{err}_D(h) + \frac{1}{2} d_{h, \Delta h}(\tilde{D}_S, \tilde{D}_T) + \lambda + \frac{10\log(\delta)}{c} + O\left( \frac{1}{\sqrt{n}} \right) \]

- \( d_{h}(\cdot, \cdot) \) measures distance between two distributions
- \( \lambda \) = the optimal classification error achievable in both domains
- \( r^2 := \frac{\sum_{i=1}^{n} ||x_i - g(f(x_i))||^2}{m} \) the average reconstruction error

Experiments

Datasets (Train/Test):

- Image: 10 digit classification
  - MNIST: 60,000/10,000
  - SVHN: 73,257/26,032
  - USPS: 7,291/2,007

- Text: sentiment analysis
  - Books (B): 2,000/4,465
  - DVDs (D): 2,000/3,586
  - Electronics (E): 2,000/5,681
  - Kitchen appliances (K): 2,000/5,945

- Speech: speech recognition
  - Native: ~ 25/7 hours
  - Chinese accent: ~ 25/7 hours
  - Indian accent: ~ 25/7 hours

Carnegie Mellon University

Visualisation on Synthetic Experiments: DAuto aligns features from both source and target domains.

Sentiment Analysis: Unsupervised data/Semi-supervised Learning Helps.

Digit Classification: Works in multi-class setting.

<table>
<thead>
<tr>
<th></th>
<th>No Adapt</th>
<th>DANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>0.8553</td>
<td>0.5459</td>
</tr>
<tr>
<td>SVHN</td>
<td>0.5277</td>
<td>0.8846</td>
</tr>
<tr>
<td>USPS</td>
<td>0.6900</td>
<td>0.5426</td>
</tr>
<tr>
<td>ADDA</td>
<td>0.5983</td>
<td>0.6442</td>
</tr>
<tr>
<td>DANN</td>
<td>0.2241</td>
<td>0.9880</td>
</tr>
<tr>
<td>SVHN</td>
<td>0.6052</td>
<td>0.6500</td>
</tr>
<tr>
<td>MNIST</td>
<td>0.1585</td>
<td>0.3562</td>
</tr>
<tr>
<td>USPS</td>
<td>0.3562</td>
<td>0.9517</td>
</tr>
</tbody>
</table>

Speech Recognition: Improved results between Native and Chinese.