Adversarially Constrained Interpolation for Unsupervised Domain Adaptation

Mohamed Azzam¹; Aurele Tohokantche Gnana¹; Hau-San Wong¹; Si Wu²
¹City University of Hong Kong, ²South China University of Technology

{m.azzam, tagnanha2-cj}@my.cityu.edu.hk, cshswong@cityu.edu.hk, cswusi@scut.edu.cn

Problem

We solve the problem of unsupervised domain adaptation, which aims to adapt models trained on source data (labeled) in order to perform well on a target data (unlabeled).

Motivation

Common solution: Learning domain-invariant representations using domain adversarial training.

The current domain adversarial training methods face two challenges:

- Data Inefficiency from the source and target domain
- Target-domain samples of different classes may become neighbors in the feature space

Contribution

- Using Inter-domain mixup samples to train the model
- Enforcing the domain discriminator to predict the mixup ratio to facilitate the training

Inter-domain mixup samples obtain better representation through:

1. Binding samples of the same class together in a single cluster
2. Encouraging smooth transition between clusters

Method

![Diagram of the method]

Source input: \( x_S \)
Mixup: \( m_p \)

Target input: \( x_T \)
Mixup ratio: \( \gamma \)

Domain mixup input: \( x_m \)
Domain mixup logit: \( y_m \)

\[ m_F(x_1, x_2) = \gamma x_1 + (1 - \gamma) x_2 \]

Results

The table shows the classification accuracy (%) of models tested on the target data.

<table>
<thead>
<tr>
<th>Source data</th>
<th>MNIST</th>
<th>SVHN</th>
<th>MNIST</th>
<th>MNIST-M</th>
<th>CIFAR-10</th>
<th>STL-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source-only</td>
<td>40.9</td>
<td>82.4</td>
<td>59.9</td>
<td>76.3</td>
<td>63.6</td>
<td></td>
</tr>
<tr>
<td>DANN</td>
<td>35.7</td>
<td>73.9</td>
<td>77.4</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>VADA</td>
<td>73.3</td>
<td>97.9</td>
<td>95.7</td>
<td>80.0</td>
<td>73.5</td>
<td></td>
</tr>
<tr>
<td>VMT</td>
<td>85.2</td>
<td>98.9</td>
<td>98.0</td>
<td>82.0</td>
<td>78.5</td>
<td></td>
</tr>
<tr>
<td>IIIM</td>
<td>-</td>
<td>97.3</td>
<td>99.5</td>
<td>83.1</td>
<td>81.6</td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td>88.7</td>
<td>99.0</td>
<td>98.1</td>
<td>83.7</td>
<td>79.7</td>
<td></td>
</tr>
</tbody>
</table>

Figures a & b shows the t-SNE plot of the last hidden layer of C in Source-only method and our method, respectively.