Maximum Classifier Discrepancy for Unsupervised Domain Adaptation

Kuniaki Saito, Kohel Watanabe, Yoshitaka Ushiku, and Tatsuya Harada

1The University of Tokyo 2RIKEN AIP

Introduction

Unsupervised Domain Adaptation (UDA)

- Transfer knowledge from label-rich domain (source) to unlabeled domain (target).
- Goal is to obtain a good classifier for the target domain.

Popular Approaches: Aligning features by a domain classifier

- Domain classifier tries to predict domain's label (source or target).
- Feature generator tries to deceive the domain classifier to extract domain-invariant features.

Problem

- Task-specific classifier is not considered to align features.
- Features of target can be generated near the decision boundary.

We propose a new method, Maximum Classifier Discrepancy, which utilizes the discrepancy (disagreement) of two classifiers.

Method Overview

Maximizing discrepancy

- We train two classifiers to increase the discrepancy on the target.
- The discrepancy is high when a domain classifier is used.

Minimizing discrepancy

- We update task-specific classifier to reduce discrepancy.

Maximize discrepancy (Fix G)

\[ \text{Maximize} \quad \mathbb{E}_{x \sim \mathcal{P}_s} D(F_1(x), F_2(x)) \]

Minimize discrepancy (Fix \( F_1 \), \( F_3 \))

\[ \text{Minimize} \quad \mathbb{E}_{x \sim \mathcal{P}_t} D(G(F_1(x)), F_2(x)) \]

Adversarial Learning Steps

1. Fix \( \theta \), Update \( \phi \)
2. Fix \( \phi \), Update \( \theta \)

Experiments 1: Digits Image Classification

<table>
<thead>
<tr>
<th>METHOD</th>
<th>SVHN to MNIST</th>
<th>SYNSEG to GTSRB</th>
<th>MNIST to USPS</th>
<th>MNIST* to USPS*</th>
<th>USPS to MNIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>71.1</td>
<td>86.7</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
<tr>
<td>MMD</td>
<td>82.7</td>
<td>93.3</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
<tr>
<td>DANN</td>
<td>82.7</td>
<td>93.3</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
<tr>
<td>ADA</td>
<td>82.7</td>
<td>93.3</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
<tr>
<td>MMD</td>
<td>82.7</td>
<td>93.3</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
<tr>
<td>Target</td>
<td>71.1</td>
<td>86.7</td>
<td>88.1</td>
<td>88.1</td>
<td>67.2</td>
</tr>
</tbody>
</table>

Results:

- Our method outperformed domain classifier based methods.
- The accuracy improves with the decrease of discrepancy loss.

Summary

- A new adversarial learning for UDA using task-specific classifiers.
- Method effective for image classification and semantic segmentation.
- The effectiveness is verified through extensive experiments.
- Gradient reversal layer [2] also works in many settings.
- We proposed to sample two networks from a network with dropout [7].