Universal Domain Adaptation

Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan

Domain Adaptation Setting
- Domain adaptation: transfer from the source domain to the target domain
- Domain gap: The target domain differs from the source domain
- Several domain adaptation settings
- Classified by the relationship between label sets of source and target domains

Open Set DA (Busto et al. 2017) Open Set DA (Saito et al. 2018)

Universal DA

Practical Value of Universal Domain Adaptation
- Domain adaptation in the wild
- E.g., transfer an animal recognizer to an unknown area
- Unlabeled data is easily accessible by setting up a camera
- There may exist some unknown species that are not in the training set
- Domain gap exists and the trained animal recognizer cannot be directly used

Challenges in Universal Domain Adaptation
- Category gap arises from the difference of the label sets
- Commonsness: \( \xi = \frac{|C_s \cap C_t|}{|C_s|} \)
- Domain gap between the source and target data in the common label set \( C \)
- Recognize unknown classes is hard for neural networks

Universal Adaptation Network

- Insights
  - Need a transferability criterion to identify data from \( C_s, C_t, \overline{C_s}, \overline{C_t} \)
  - Align data in the common label set \( C \) maximally
  - Ignore data in the private label set \( C_s \) and \( C_t \) during adaptation
  - Reject target examples with low transferability as "unknown" class

- Ideal transferability criterion
  - These inequalities should hold for ideal transferability criterion:
    
    \[
    \begin{align*}
    E_{x,s} w(x) &> E_{x,s} w(x) \\
    E_{x,s} w(x) &> E_{x,t} w(x)
    \end{align*}
    \]

- Architecture
  - Training phase
  - Testing phase

- x: input image
- z: extracted feature
- D: domain discriminator
- EC, ED: loss function

- End-to-End training

\[
E_C = E(x, s) - E(y, G(F(x)))
\]

\[
E_D = -E_{x,s} \log D(F(x)) - E_{x,t} \log (1 - D'(F(x)))
\]

\[
\max_{D} \min_{F} E_C - \lambda E_D \quad \min_{D} E_D
\]

- Transferability criterion
  - Domain Similarity
    - \( \hat{d} \): the domain similarity of each sample
    - For a source sample, smaller \( \hat{d} \) means it is more similar to the target domain
    - For a target sample, larger \( \hat{d} \) means it is more similar to the source domain
  - \( E_{x,s} w(x) - \hat{d} \): auxiliary classifier
  - Prediction Uncertainty
    - \( H(y) \): the entropy of each prediction
    - Predictions are certain for source samples
    - Predictions are uncertain for target samples
    - Influenced by the high entropy samples from \( \hat{d} \), the entropy of samples from \( \hat{d} \) becomes larger
    - Samples from \( \hat{d} \) keep high certainty

- \( w(x) = \frac{H(y)}{\log |C_s|} - \hat{d}(x) \)
- \( w(x) = \hat{d}(x) - \frac{H(y)}{\log |C_s|} \)

Experimental Results

- State-of-the-art performance

<table>
<thead>
<tr>
<th>Method</th>
<th>Office-Home ( \xi = 0.15 )</th>
<th>Office-Home ( \xi = 0.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet</td>
<td>59.32</td>
<td>76.58</td>
</tr>
<tr>
<td>DANN</td>
<td>56.17</td>
<td>81.72</td>
</tr>
<tr>
<td>RTN</td>
<td>54.56</td>
<td>77.80</td>
</tr>
<tr>
<td>PADA</td>
<td>52.63</td>
<td>81.40</td>
</tr>
<tr>
<td>AT1</td>
<td>51.50</td>
<td>59.11</td>
</tr>
<tr>
<td>AT2</td>
<td>52.00</td>
<td>80.37</td>
</tr>
<tr>
<td>AT3</td>
<td>51.50</td>
<td>63.85</td>
</tr>
<tr>
<td>AT4</td>
<td>51.00</td>
<td>79.64</td>
</tr>
<tr>
<td>UAN</td>
<td>63.00</td>
<td>82.83</td>
</tr>
</tbody>
</table>

- Table: Results on Office-31 \( \xi = 0.32 \), ImageNet-Caltech \( \xi = 0.07 \), VisDA2017 \( \xi = 0.50 \)

- Method | Office-31 | ImageNet-Caltech | VisDA |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet</td>
<td>75.94</td>
<td>89.60</td>
<td>90.91</td>
</tr>
<tr>
<td>DANN</td>
<td>80.65</td>
<td>90.94</td>
<td>88.07</td>
</tr>
<tr>
<td>RTN</td>
<td>85.70</td>
<td>87.80</td>
<td>89.81</td>
</tr>
<tr>
<td>PADA</td>
<td>85.25</td>
<td>90.99</td>
<td>90.01</td>
</tr>
<tr>
<td>AT1</td>
<td>79.38</td>
<td>92.60</td>
<td>90.98</td>
</tr>
<tr>
<td>AT2</td>
<td>66.13</td>
<td>73.57</td>
<td>85.62</td>
</tr>
<tr>
<td>AT3</td>
<td>85.62</td>
<td>94.77</td>
<td>97.99</td>
</tr>
<tr>
<td>AT4</td>
<td>64.25</td>
<td>85.45</td>
<td>85.12</td>
</tr>
<tr>
<td>UAN</td>
<td>53.92</td>
<td>85.12</td>
<td>89.24</td>
</tr>
</tbody>
</table>

- Analysis
  - Existing methods perform well when their assumptions hold but worse when violated
  - UAN works stably without prior knowledge on the relationship of label sets
  - Mainstream methods suffer from negative transfer in most classes
  - Only UAN promotes positive transfer for all classes
  - Quality of the transferability criterion can be verified

- Hypothesis Quality (blue for common and black for private)

School of Software - Tsinghua University - China
National Engineering Lab for Big Data Software
University of California, Berkeley
Mail: youkaichao@gmail.com