Abstract

- Elastic Weight Consolidation (EWC) is a technique used in overcoming catastrophic forgetting between successive tasks.
- Domain Adaptation (DA) aims to build algorithms that leverage information from source domains to facilitate performance on an unseen target domain.
- We propose a model-independent framework - Sequential Domain Adaptation (SDA).
- SDA draws on EWC for training on successive source domains to move towards a general domain solution, thereby solving the problem of domain adaptation.
- We test SDA on convolutional, recurrent, and attention-based architectures. Our experiments show that the proposed framework enables simple architectures such as CNNs to outperform complex state-of-the-art models in domain adaptation of SA.
- In addition, we observe that the effectiveness of a harder first Anti-Curriculum ordering of source domains leads to maximum performance.

Experimental Setup

Datasets: Experiments on the standard Multi-Domain Sentiment Dataset. It contains reviews from 4 domains, namely Books (B), DVD (D), Electronics (E), and Kitchen (K). Each domain has 2000 reviews.

Architectures: CNN, LSTM, Attention LSTM (ALSTM) and Transformer Encoder (TE).

SoTA Domain Adaptation baselines:
1. PBLM: Pivot Based Language Model (PBLM)
2. DSR: Domain-Specific Representations
3. BLSE: Bilingual Sentiment Embeddings
4. DAS: Domain Adaptive Semi-supervised learning
5. ACAN: Adversarial Category Alignment Network

Results and Discussion

<table>
<thead>
<tr>
<th></th>
<th>Kitchen</th>
<th>Electronics</th>
<th>DVD</th>
<th>Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBLM [4]</td>
<td>68.5</td>
<td>52.4</td>
<td>59.55</td>
<td>61.4</td>
</tr>
<tr>
<td>DSR [10]</td>
<td>56</td>
<td>55</td>
<td>58.3</td>
<td>52</td>
</tr>
<tr>
<td>BLSE [9]</td>
<td>80</td>
<td>74.25</td>
<td>75.25</td>
<td>71.25</td>
</tr>
<tr>
<td>GAS [9]</td>
<td>79.05</td>
<td>75.35</td>
<td>60.35</td>
<td>62.6</td>
</tr>
<tr>
<td>BLSE [9]</td>
<td>79.35</td>
<td>73.65</td>
<td>70.8</td>
<td>73.95</td>
</tr>
<tr>
<td>EWC-CNN</td>
<td>80.25</td>
<td>77.25</td>
<td>72.8</td>
<td>72.35</td>
</tr>
<tr>
<td>EWC-LSTM</td>
<td>79.7</td>
<td>76.25</td>
<td>71.35</td>
<td>70.8</td>
</tr>
<tr>
<td>EWC-ALSTM</td>
<td>75</td>
<td>79.2</td>
<td>69.05</td>
<td>67.95</td>
</tr>
<tr>
<td>EWC-TE</td>
<td>71.35</td>
<td>72.35</td>
<td>65.85</td>
<td>66.6</td>
</tr>
</tbody>
</table>

TABLE II: Comparison of proposed framework SDA with state-of-the-art architectures on the Multi-Domain Sentiment Dataset.

For the SoA, best performing source domain is chosen, while for SDA anti-curriculum source domain ordering is chosen.

1. Proposed outperforms all.
2. Some use Domain adaptation in a semi-supervised setting. In other words, they use large quantities of target domain unlabelled data yet SDA outperforms all.
3. DSR, similar to the proposed framework, uses multiple source domains for learning domain representations.
4. Their model relies heavily on domain classification. However, training a robust domain classifier requires much more data,
5. BLSE outperforms SDA in one target domain, they utilize labelled target domain data for training their architecture, whereas as SDA keeps target domain strictly unseen. Compared to standard CNN on different text classification datasets.

References