Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout

Yi Yang and Jacob Eisenstein

Domain Adaptation

Example: Part-of-speech Tagging

Source:
And God said, Let ... CC CNN VBD VB ...

Features:
• Mid_said source spec
• Prev_God cross domain
• Next_Let source spec

Target:
And God seid, Liyt ... CC CNN VBD VB ...

Features:
• Mid_said target spec
• Prev_God cross domain
• Next_Let target spec

Representation Learning

Learn new sets of dense features:

Source
 Learned representations
 Target

Representation learning for domain adaptation:
• Structural Corresponding Learning (SCL) [1]
• Brown Clustering
• (marginalized) Stacked Denoising Autoencoders (SDA/mSDA) [2,3]
• Latent Variable Models
• Neural Probabilistic Language Model

Denoising Autoencoders

Feature templates: And God said, let ...

• Template 1: previous token type
• Template 2: next token type

Structured Dropout

Shape of Q under different noises:

Eliminate matrix inverse for W = PQ⁻¹!

Evaluation: Same accuracy, 25X faster!

Datasets: Tycho Brahe corpus (historical Portuguese texts with 383 tags)

Results:

<table>
<thead>
<tr>
<th>Task</th>
<th>baseline</th>
<th>PCA</th>
<th>SCL</th>
<th>mDA with dropout noise</th>
<th>mDA with structured dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from 1800-1849</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1750</td>
<td>90.12</td>
<td>90.99</td>
<td>90.69</td>
<td>90.08</td>
<td>90.08</td>
</tr>
<tr>
<td>1700</td>
<td>90.43</td>
<td>90.43</td>
<td>90.16</td>
<td>91.56</td>
<td>91.57</td>
</tr>
<tr>
<td>1650</td>
<td>88.45</td>
<td>88.52</td>
<td>87.99</td>
<td>86.69</td>
<td>86.70</td>
</tr>
<tr>
<td>1600</td>
<td>87.56</td>
<td>87.58</td>
<td>88.47</td>
<td>89.60</td>
<td>89.61</td>
</tr>
<tr>
<td>1550</td>
<td>89.66</td>
<td>89.61</td>
<td>90.57</td>
<td>91.39</td>
<td>91.39</td>
</tr>
<tr>
<td>1500</td>
<td>90.58</td>
<td>90.63</td>
<td>86.99</td>
<td>88.96</td>
<td>88.95</td>
</tr>
<tr>
<td>from 1750-1849</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1700</td>
<td>94.64</td>
<td>94.62</td>
<td>94.81</td>
<td>95.08</td>
<td>95.08</td>
</tr>
<tr>
<td>1650</td>
<td>91.98</td>
<td>92.97</td>
<td>90.37</td>
<td>90.83</td>
<td>90.84</td>
</tr>
<tr>
<td>1600</td>
<td>92.95</td>
<td>92.91</td>
<td>93.17</td>
<td>93.70</td>
<td>93.78</td>
</tr>
<tr>
<td>1550</td>
<td>93.37</td>
<td>93.31</td>
<td>93.75</td>
<td>94.06</td>
<td>94.05</td>
</tr>
<tr>
<td>1500</td>
<td>95.60</td>
<td>95.75</td>
<td>90.59</td>
<td>91.71</td>
<td>91.71</td>
</tr>
</tbody>
</table>

References


Acknowledgments

This research was supported by National Science Foundation award 1349837. The first author was also supported by National Science Foundation ACL travel award.