Two applications of Transfer Learning

Mahdi Pakdaman

Intelligent systems Program
University of Pittsburgh
map218@pitt.edu
Outline

- Two TL algorithms
- Tasks
  - Alphabet recognition
  - Time series prediction
- Conclusion
- Future Work
Transfer Learning: Sample Transfer

\[
\min \frac{\gamma}{2} \|w\|_2^2 + \sum_i \vartheta_i + \sum_j \rho_j \vartheta_j
\]

S.t: \( y_i w^T x_i \geq 1 - \vartheta_i \quad i = 1, \ldots, n \)
\( y_j w^T x_j \geq 1 - \vartheta_j \quad j = 1, \ldots, m, \)

where \( p_j = \min \{ i \mid y_i = y_j \} \exp(-\|x_i - x_j\|) \)
Transfer Learning: Sharing Weights

\[
\min \frac{\gamma}{2} \|w\|_2^2 + \frac{\delta}{2T} \sum_t \|v_t\|_2^2 + \sum_t \sum_i \vartheta_i^{(t)}
\]

S.t: \[y_i (w + v_t)^t x_i \geq 1 - \vartheta_i \quad i = 1, \ldots, n \quad t = 1, \ldots, T\]
First Experiment: Hand written character Recognition

- Task 1: Recognize C vs O
- Task 2: Recognize F vs P
- Task 3: Recognize G vs Q
Second experiment: hpf4 order prediction

- Predicting \textit{hpf4 order} [a type of medication] for hospitalized patients

- Using time series prediction models:
  - Like HMM, CRF, ...
  - Disadvantage: We do not know exactly when the event happens
- By using accumulative profile features we can cast the time series prediction as multi-task learning problem
Hpf-4 prediction using TL

- TM-MLT (1)
  - M1: Predicting 24 hours ahead prediction
  - M2: Predicting 48 hours ahead prediction
  - M3: Predicting 72 hours ahead prediction

- TM-MLT (2)
  - M1: Predicting up to 24 hours ahead prediction
  - M2: Predicting up to 48 hours ahead prediction
  - M3: Predicting up to 72 hours ahead prediction
<table>
<thead>
<tr>
<th>Alphabet Recognition</th>
<th>Linear SVM</th>
<th>WT-SVM</th>
<th>ST-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>C vs O</td>
<td>0.93</td>
<td>0.93</td>
<td>0.98</td>
</tr>
<tr>
<td>F vs P</td>
<td>0.86</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>G vs Q</td>
<td>0.61</td>
<td>0.61</td>
<td>0.72</td>
</tr>
</tbody>
</table>
# Results: hpf4 prediction

<table>
<thead>
<tr>
<th>Hpf4 Prediction (1)</th>
<th>Linear SVM</th>
<th>WT-SVM</th>
<th>ST-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hours ahead prediction</td>
<td>0.72</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>48 hours ahead prediction</td>
<td>0.62</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td>72 hours ahead prediction</td>
<td>0.65</td>
<td>0.60</td>
<td>0.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hpf4 Prediction (2)</th>
<th>Linear SVM</th>
<th>WT-SVM</th>
<th>ST-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 24 hours ahead prediction</td>
<td>0.72</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>Up to 48 hours ahead prediction</td>
<td>0.63</td>
<td>0.61</td>
<td>0.74</td>
</tr>
<tr>
<td>Up to 72 hours ahead prediction</td>
<td>0.68</td>
<td>0.68</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Conclusion

- Generally sample transform performs better than weight sharing strategy
- Transfer Learning works well specially in cases that we do not have enough sample to learn concepts
Future Work

- Use kernel trick to have nonlinear SVMs
- Run more in depth experiment (the current results are based on simple random sampling)
- Transfer learning in time-series prediction (Transferring knowledge for predicting different medication order)
Thanks