Hybrid semi-Markov CRF for Neural Sequence Labeling

Zhi-Xiu Ye, Zhen-Hua Ling
University of Science and Technology of China

Background

Sequence labeling is a type of pattern recognition task that involves the algorithmic assignment of a categorial label to each member of a sequence of observed values. Take named entity recognition as an example: sentence: Barack Obama was born in Hawaii.

1. Input data
   a. Input sentence $x = \{x_1, ..., x_n\}$
   b. Word-level label: $y = \{y_1, ..., y_n\}$
   c. Segment-level label: $s = \{s_1, s_2, ..., s_p\}$

2. Word-level representations

   a. $w_i$ for CRF and $a_i$, $b_i$ for HSCRF.

3. Score computation

   a. In CRFs, compute the score of word-level label $m_k$ via the representation of $i$-th word $w_i$.
   b. In HSCRFs, compute the score of segment-level label $m_k$ via the representation of $i$-th word $w_i$.

4. Joint training and decoding

   a. A CRF output layer and a HSCRF output layer are integrated into an unified neural network.
   b. The model parameters are shared and optimized by minimizing the summation of the loss functions of the CRF layer and the HSCRF layer with equal weights as follows:

   $\text{loss} = \text{loss}_{\text{CRF}} + \text{loss}_{\text{HSCRF}}$

Experiments

Dataset: CoNLL 2003 shared task: English named entity recognition.

<table>
<thead>
<tr>
<th>Model</th>
<th>Ent. Length</th>
<th>Test Set F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM-BLSTM-CRF</td>
<td></td>
<td>91.68</td>
</tr>
<tr>
<td>LM-BLSTM-HSCRF</td>
<td>91.65</td>
<td>92.05</td>
</tr>
<tr>
<td>LM-BLSTM-HSCRF(JNT)</td>
<td>91.79</td>
<td>92.05</td>
</tr>
<tr>
<td>CNN-BLSTM-JNT(JNT)</td>
<td>mean</td>
<td>91.26 ± 0.10</td>
</tr>
<tr>
<td>LM-BLSTM-JNT(JNT)</td>
<td>mean</td>
<td>91.34</td>
</tr>
</tbody>
</table>

Table 2: Comparison with existing works

1. Word-level labels may supervise models to learn word-level descriptions which tend to benefit the recognition of short entities.

2. Segment-level labels may guide models to capture the descriptions of combining words for whole entities which help to recognize long entities.

3. By utilizing both labels, the proposed joint model can achieve better overall performance of recognizing entities with different lengths.

Source code available!!!

https://github.com/ZhixiuYe/HSCRF-pytorch

Our implementation is based on python and the PyTorch library.