Sentence-State LSTM for Text Representation

Yue Zhang¹, Qi Liu¹ and Linfeng Song²
¹Singapore University of Technology and Design, ²University of Rochester

Introduction

1. Bi-directional LSTM

2. Disadvantages:
1. BiLSTM is slow, due to its non-parallelism caused by its sequential nature (Vaswani et al., 2017).
2. Lack of balance between local n-gram and global sequence information (Wang et al., 2016).

Method

Tasks
1. Classification (vanilla attention): \[ y = \text{softmax}(W_y g + b_y) \]
2. Sequence Labeling (vanilla CRF): \[ y_i = \text{softmax}(W_y h_i + b_y) \]

Contrast with existing work

<table>
<thead>
<tr>
<th>Model</th>
<th>Simultaneous</th>
<th>N-gram</th>
<th>Global</th>
<th>Recurrent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>×</td>
<td>×</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>CRF</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>SAN</td>
<td>√</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-LSTM</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiments

1. Data
1. Classification:
   - Movie review (Pang and Lee (2008)), 16 datasets (Liu et al. (2017))
2. Sequence Labeling
   - NER: CoNLL (Sang et al., 2003)
   - POS tagging: PTB (Marcus et al., 1993)

2. Development

3. Classification

4. Sequence labeling

5. Contrast with Bi-LSTM

References