CS224N Final Project

Neural Network Joint Language Model: An Investigation and An Extension with Global Source Context

Yuhao Zhang, Charles Ruizhongtai Qi
(Advised by Thang Luong)
Objective

• Implement and evaluate a state-of-the-art joint language model.
• Extend the model to make use of global source context information.
Example

Source: <s> 他 每天 早上 走路 去 学校 </s>
he every morning walk to school

Target: <s> he walks to school every morning </s>
P(morning | to, school, every, 他, 每天, 早上, 走路, 去)
Neural Network Joint Model (NNJM)

\[ P(\text{school} \mid \text{he}, \text{walks}, \text{to, 走路, 去, 学校, } \langle\text{s}\rangle, \langle\text{s}\rangle) \]
Our extension to the basic model: NNJM with global source context (NNJM-Global)
Variants of NNJM-Global:

- Splitting the source context
- Add non-linear layer
Experiment

• Evaluation Metric
  • Perplexity on test set

\[ PP(W) = p(w_1, w_2, ..., w_N)^{-\frac{1}{N}} \]

• Data
  • French-English parallel text (25K pairs) with alignments from European Parallel Corpus.

• Training
  • Mini-batch gradient descent
  • Theano, GPU
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNJM</td>
<td>9.51</td>
</tr>
<tr>
<td>NNJM-Global</td>
<td>9.45</td>
</tr>
<tr>
<td>NNJM-Global+Split</td>
<td>9.38</td>
</tr>
<tr>
<td>NNJM-Global+Split+NonLinear</td>
<td>9.33</td>
</tr>
</tbody>
</table>
Average value of hidden layer weights for each word
Thank You!