Neural Natural Language Inference Models Enhanced with External Knowledge

Qian Chen¹, Xiaodan Zhu², Zhen-Hua Ling¹, Diana Inkpen³, Si Wei⁴

¹University of Science and Technology of China  ²Queen’s University  ³University of Ottawa  ⁴FLYTEK Research

Contributions

★ Enrich the state-of-the-art neural natural language inference models with external knowledge.
★ The proposed models improve neural NLI models to achieve the state-of-the-art performance on the SNLI and MultiNLI datasets.

Source code available!!!

https://github.com/lukecq1231/kim

Our implementation uses python and is based on the Theano library.

An example

<table>
<thead>
<tr>
<th>P/G Sentences</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>e/c p: An African person standing in a <strong>wheat</strong> field.</td>
<td>The child is getting a <em>pouc</em>re.</td>
</tr>
<tr>
<td>h: A person standing in a <strong>corn</strong> field.</td>
<td>The child is getting a manicure.</td>
</tr>
<tr>
<td>e/c p: Little girl is flipping an <strong>omelet</strong> in the kitchen.</td>
<td>At the <strong>marketplace</strong>.</td>
</tr>
<tr>
<td>h: A young girl cooks <strong>pancakes</strong>.</td>
<td>At the <strong>store</strong>.</td>
</tr>
<tr>
<td>c/e p: Two boys are swimming with <strong>boogie boards</strong>.</td>
<td>The boys are swimming with their <strong>floats</strong>.</td>
</tr>
<tr>
<td>h: Two boys are swimming in the <strong>kitchen</strong>.</td>
<td>The boys are swimming in the <strong>kitchen</strong>.</td>
</tr>
</tbody>
</table>

Our model — KIM (Knowledge-based Inference Model)

1. External Knowledge

2. Input Encoding

- **Premise:** \( a = (a_1, \ldots, a_m) \)
- **Hypothesis:** \( b = (b_1, \ldots, b_n) \)

3. Knowledge-Enriched Co-Attention

4. Local Inference Collection with External Knowledge

5. Knowledge-Enhanced Inference Composition

<table>
<thead>
<tr>
<th>Detail of KIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. External Knowledge</td>
</tr>
<tr>
<td>( r_{ij} \rightarrow ) [Syn, Ant, Hyper, Hypon, Co-hypon]</td>
</tr>
<tr>
<td>2. Input Encoding</td>
</tr>
<tr>
<td>Premise: ( a = (a_1, \ldots, a_m) ) \n</td>
</tr>
<tr>
<td>3. Knowledge-Enriched Co-Attention</td>
</tr>
<tr>
<td>( e_{ij} = (a_i^<em>)^T b_j^</em> + F(r_{ij}) )</td>
</tr>
<tr>
<td>4. Local Inference Collection with External Knowledge</td>
</tr>
<tr>
<td>( a_i^m = G(a_i^e; a_i^c; a_i^* - a_i^e; a_i^c; \sum_{j=1}^m a_j r_{ij}) )</td>
</tr>
<tr>
<td>5. Knowledge-Enhanced Inference Composition</td>
</tr>
<tr>
<td>( a_i^w = \sum_{j=1}^m \exp(H(a_i^m; \alpha_i r_{ij})) a_j^w )</td>
</tr>
</tbody>
</table>

Analysis

<table>
<thead>
<tr>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>SNLI:</strong> Training: 550k sentence pairs, held-out: 10k, testing: 10k</td>
</tr>
<tr>
<td>- <strong>Clockner’s Test set:</strong> testing: 8k</td>
</tr>
<tr>
<td>- <strong>MultiNLI:</strong> Training: 400k sentence pairs, held-out: 10k/10k, testing: 10k/10k</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1: Accuracies of models on SNLI.</th>
</tr>
</thead>
<tbody>
<tr>
<td>------------------</td>
</tr>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>LSTM Att. [Rocktäschel et al., 2015]</td>
</tr>
<tr>
<td>Match-LSTM [Wang and Jiang, 2016]</td>
</tr>
<tr>
<td>Decomposable Att. [Parkhi et al., 2016]</td>
</tr>
<tr>
<td>DIEN [Gong et al., 2017]</td>
</tr>
<tr>
<td>CAFE [Tay et al., 2018]</td>
</tr>
<tr>
<td>ESIM [Chen et al., 2017a]</td>
</tr>
<tr>
<td>KIM (This paper)</td>
</tr>
</tbody>
</table>

| Table 2: Accuracies of models on the SNLI and [Glockner et al., 2018] test set. * indicates the results taken from [Glockner et al., 2018]. |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| **Model** | **SNLI** | **Glockner’s** | **Model** | **SNLI** | **Glockner’s** | **Model** | **SNLI** | **Glockner’s** |
| [Parikh et al., 2016] | 84.7 | 84.7 | 84.7 | 84.7 | 84.7 | 84.7 | 84.7 | 84.7 |
| [Nie and Bansal, 2017] | 86.0 | 86.0 | 86.0 | 86.0 | 86.0 | 86.0 | 86.0 | 86.0 |
| ESIM * | 87.9 | 87.9 | 87.9 | 87.9 | 87.9 | 87.9 | 87.9 | 87.9 |
| KIM (This paper) | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 | 88.6 |

| Table 3: Accuracies of models on MultiNLI. * indicates models using extra SNLI training set. |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| **Model** | **In** | **Cross** | **Model** | **In** | **Cross** | **Model** | **In** | **Cross** |
| BiLSTM [Williams et al., 2017] | 66.9 | 66.9 | Gated BiLSTM [Chen et al., 2017b] | 73.5 | 73.5 | DiN [Gong et al., 2017] | 77.8 | 77.8 |
| CAFE [Tay et al., 2018] | 78.7 | 78.7 | ESIM [Chen et al., 2017a] | 76.8 | 76.8 | KIM (This paper) | 77.2 | 77.2 |

★ On SNLI, Knowledge-based Inference Model (KIM), which enriches ESIM with external knowledge, obtains an accuracy of 88.6%.
★ On Clockner’s test set, KIM achieves 83.5% (with only a 5.1% drop), which demonstrates its better generalizability.
★ On MultiNLI, KIM achieve significant gains to 77.2% and 76.4% respectively.