Neural Natural Language Inference Models Enhanced with External Knowledge

Qian Chen[1], Xiaodan Zhu[2], Zhen-Hua Ling[1], Diana Inkpen[3], Si Wei[4]

[1]University of Science and Technology of China [2]Queen's University [3]University of Ottawa [4]iFLYTEK 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!!!

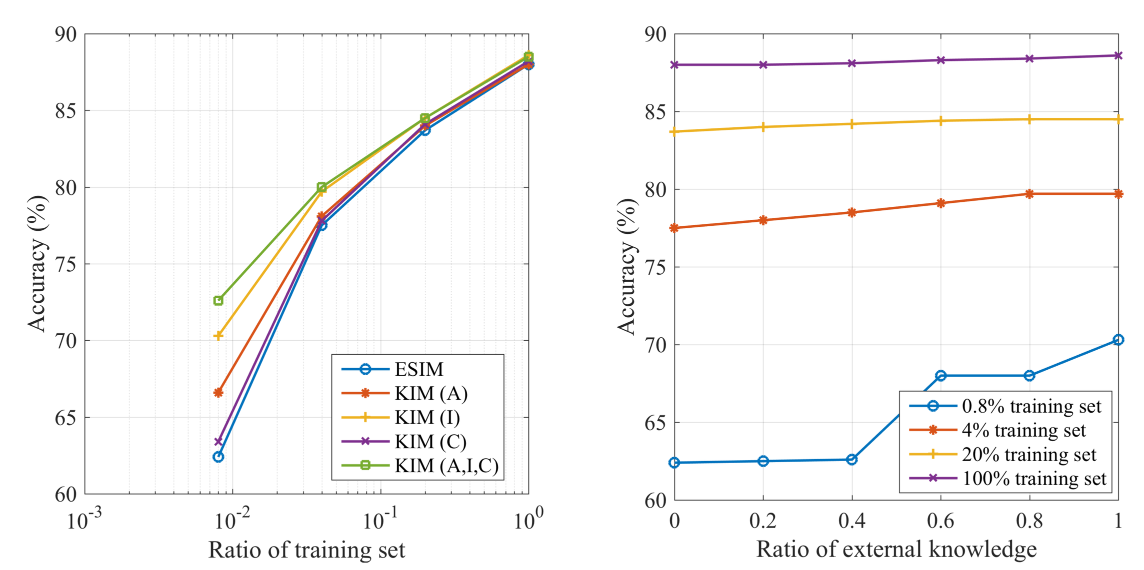
https://github.com/lukecq1231/kim

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

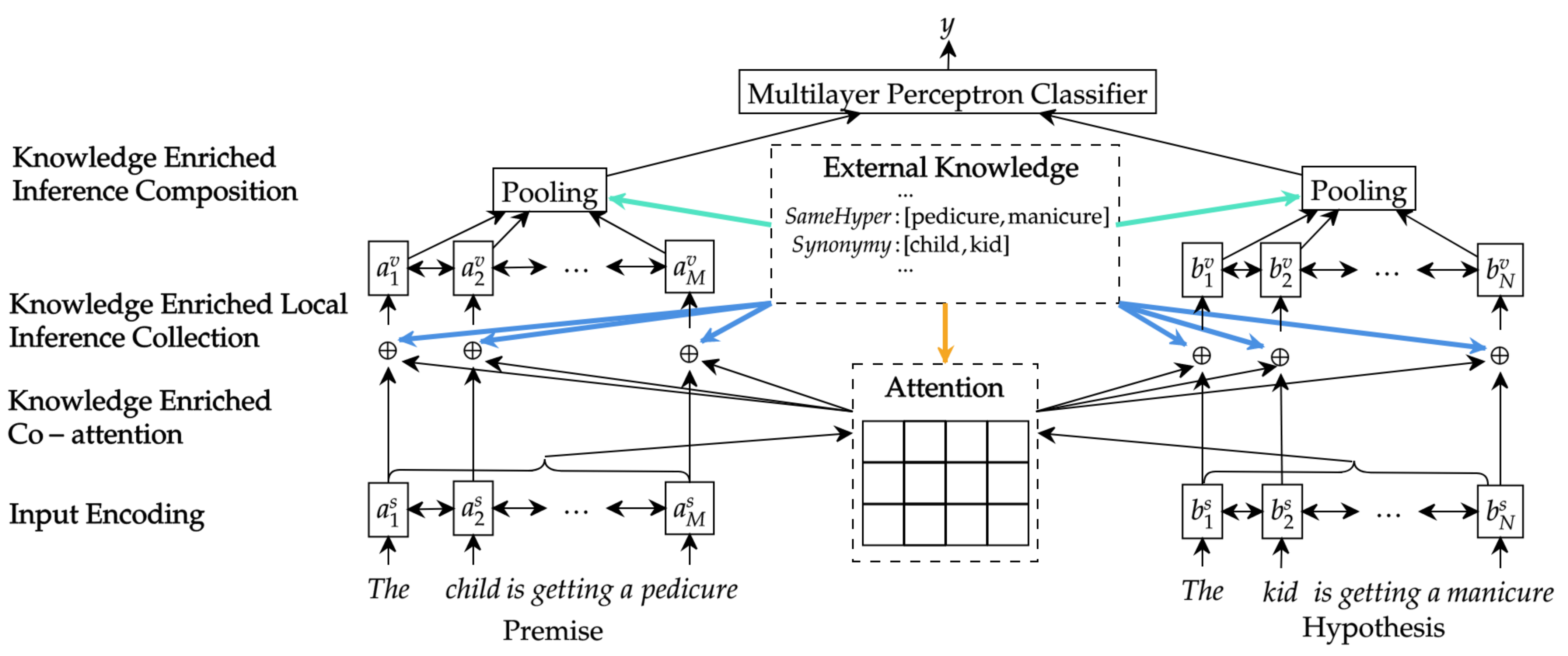
An example

P/G	Sentences
e/c	p: An African person standing in a **wheat** field. h: A person standing in a **corn** field.
e/c	p: Little girl is flipping an **omelet** in the kitchen. h: A young girl cooks **pancakes**.
c/e	p: A middle eastern **marketplace**. h: A middle easten **store**.
c/e	p: Two boys are swimming with boogie **boards**. h: Two boys are swimming with their **floats**.

Analysis



Our model — KIM (Knowledge-based Inference Model)



Detail of KIM

1. External Knowledge

$\boldsymbol{r}_{ij}$= [Syn, Ant, Hyper, Hypon, Co-hypon]

2. Input Encoding

Premise: $\boldsymbol{a} = (a_1, \dots, a_m)$

Hypothesis: $\boldsymbol{b} = (b_1, \dots, b_n)$

$$\boldsymbol{a}_i^s = \text{BiLSTM}(\mathbf{E}(\boldsymbol{a}), i)\,, \qquad (1)$$

$$\boldsymbol{b}_j^s = \text{BiLSTM}(\mathbf{E}(\boldsymbol{b}), j)\,. \qquad (2)$$

3. Knowledge-Enriched Co-Attention

$$e_{ij} = (\boldsymbol{a}_i^s)^{\mathrm{T}} \boldsymbol{b}_j^s + F(\boldsymbol{r}_{ij})\,. \qquad (3)$$

$$\mathbb{1}(\boldsymbol{r}_{ij}) = \begin{cases} 1 & \text{if } \boldsymbol{r}_{ij} \text{ is not a zero vector}\,; \\ 0 & \text{if } \boldsymbol{r}_{ij} \text{ is a zero vector}\,. \end{cases} \qquad (4)$$

Word pairs with semantic relationship are probably aligned together.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}\,, \quad \boldsymbol{a}_i^c = \sum_{j=1}^{n} \alpha_{ij} \boldsymbol{b}_j^s\,, \qquad (5)$$

$$\beta_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{m} \exp(e_{kj})}\,, \quad \boldsymbol{b}_j^c = \sum_{i=1}^{m} \beta_{ij} \boldsymbol{a}_i^s\,, \qquad (6)$$

4. Local Inference Collection with External Knowledge

$$\boldsymbol{a}_i^m = G([\boldsymbol{a}_i^s; \boldsymbol{a}_i^c; \boldsymbol{a}_i^s - \boldsymbol{a}_i^c; \boldsymbol{a}_i^s \circ \boldsymbol{a}_i^c; \sum_{j=1}^{n} \alpha_{ij} \boldsymbol{r}_{ij}])\,, \qquad (7)$$

$$\boldsymbol{b}_j^m = G([\boldsymbol{b}_j^s; \boldsymbol{b}_j^c; \boldsymbol{b}_j^s - \boldsymbol{b}_j^c; \boldsymbol{b}_j^s \circ \boldsymbol{b}_j^c; \sum_{i=1}^{m} \beta_{ij} \boldsymbol{r}_{ji}])\,, \qquad (8)$$

Through comparing $\boldsymbol{a}_i^s$ and $\boldsymbol{a}_i^c$, in addition to their relation from **external knowledge**, we can obtain word-level inference information for each word.

5. Knowledge-Enhanced Inference Composition

$$\boldsymbol{a}_i^v = \text{BiLSTM}(\boldsymbol{a}^m, i)\,, \qquad (9)$$

$$\boldsymbol{b}_j^v = \text{BiLSTM}(\boldsymbol{b}^m, j)\,. \qquad (10)$$

Use weighted pooling based on external knowledge to obtain a fixed-length vectors.

$$\boldsymbol{a}^{\mathrm{w}} = \sum_{i=1}^{m} \frac{\exp(H(\sum_{j=1}^{n} \alpha_{ij} \boldsymbol{r}_{ij}))}{\sum_{i=1}^{m} \exp(H(\sum_{j=1}^{n} \alpha_{ij} \boldsymbol{r}_{ij}))} \boldsymbol{a}_i^v\,, \qquad (11)$$

$$\boldsymbol{b}^{\mathrm{w}} = \sum_{j=1}^{n} \frac{\exp(H(\sum_{i=1}^{m} \beta_{ij} \boldsymbol{r}_{ji}))}{\sum_{j=1}^{n} \exp(H(\sum_{i=1}^{m} \beta_{ij} \boldsymbol{r}_{ji}))} \boldsymbol{b}_j^v\,. \qquad (12)$$

Results

- **SNLI:** Training: 550k sentence pairs, held-out: 10k, testing: 10k
- **Clockner's Test set:** testing: 8k
- **MultiNLI:** Training: 400k sentence pairs, held-out: 10k/10k, testing: 10k/10k

Table 1: Accuracies of models on SNLI.

Model	Test
LSTM Att. [Rocktäschel et al., 2015]	83.5
Match-LSTM [Wang and Jiang, 2016]	86.1
Decomposable Att. [Parikh et al., 2016]	86.8
DIIN [Gong et al., 2017]	88.0
CAFE [Tay et al., 2018]	88.5
ESIM [Chen et al., 2017a]	88.0
KIM (This paper)	**88.6**

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(Δ)
[Parikh et al., 2016]*	84.7	51.9 (-32.8)
[Nie and Bansal, 2017]*	86.0	62.2 (-23.8)
ESIM *	87.9	65.6 (-22.3)
KIM (This paper)	**88.6**	**83.5** (-5.1)

Table 3: Accuracies of models on MultiNLI. * indicates models using extra SNLI training set.

Model	In	Cross
BiLSTM [Williams et al., 2017]	66.9	66.9
Gated BiLSTM [Chen et al., 2017b]	73.5	73.6
DIIN * [Gong et al., 2017]	77.8	**78.8**
CAFE [Tay et al., 2018]	**78.7**	77.9
ESIM [Chen et al., 2017a]	76.8	75.8
KIM (This paper)	**77.2**	**76.4**

★ On SNLI, Knowledge-based Inference Model (**KIM**), which enriches ESIM with external knowledge, obtains an accuracy of 88.6%.

★ On Glockner'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.