KETM:A Knowledge-Enhanced Text Matching method

IJCNN2023

Kexin Jiang[1], Yahui Zhao[1,*], Guozhe Jin[1], Zhenguo Zhang[1], Rongyi Cui[1]

[1]Inst. of intelligent information processing,Yanbian University,Yanji 133002,China

*Corresponding author

One Slide Summary

- Text Matching is a fundamental problem for Natural Language Understanding

- External knowledge can enhance the understanding of text and in turn improve Text Matching models. However, existing methods lack the ability to harness external knowledge (from KGs and Wiktionary)
- **Contributions**
- We propose a knowledge-enhanced text matching method——KETM, which adds the interpretation of words to the model and fuses them with the text information using a fusion of gating mechanisms.
- Our proposed method has good generality and can be applied to other text-matching models to improve the performance of the model without adding additional model parameters.
- We experiment on several well-known datasets. Our proposed text-matching model achieves very good performance, and the experimental results are improved compared to the baseline model on these datasets.

Text Matching

Text matching refers to taking two texts as input and determining their relationship by understanding their semantics

- Example
  - p: Two men are using two horses to help with farm work.
  - h: The farmers are ploughing the fields

Text Matching with Knowledge Enhance

Motivation

- Example
  - p: The man is holding a saxophone
  - h: The man is holding an instrument

- Important to understand the the relationship between terms that can be fetched using external knowledge
  - saxophone and instrument

Goal

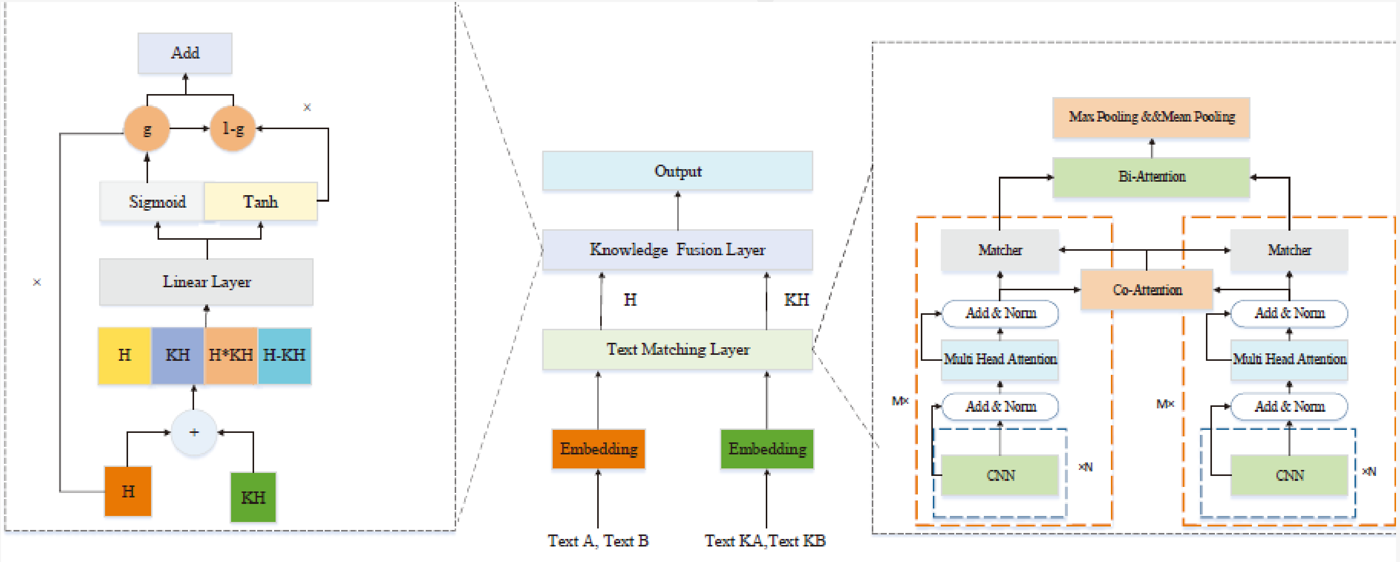
- Analyzing the impact of relevant external knowledge for Text Matching
- Develop hybrid models that can exploit information both from text and external knowledge

External Knowledge

Enhance Text using Wiktionary form External Knowledge

P: The man is holding a saxophone. ‘saxophone’: “A single-reed instrument musical instrument of the woodwind family, usually made of brass and with a distinctive loop bringing the bell upwards.”
H:The man is holding an instrument ‘instrument’: “A device used to produce music.”

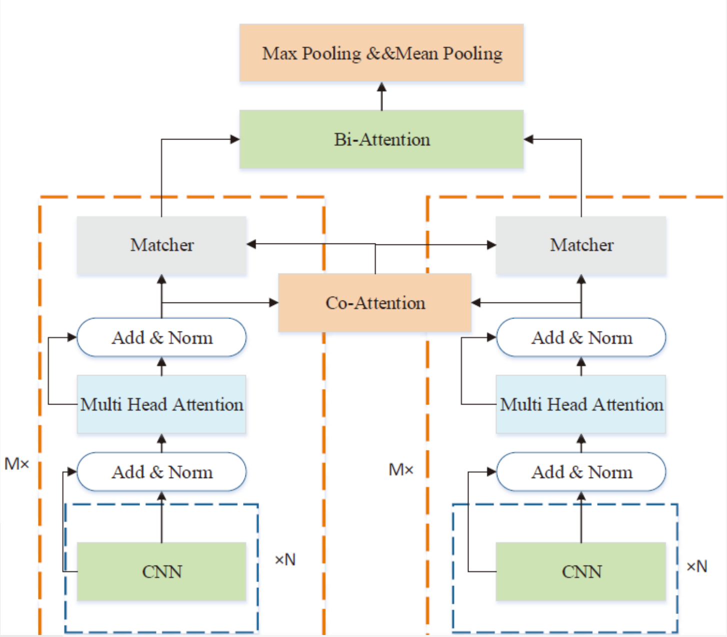
Approach Overview



- Text and Knowledge Embedding
- Extract Text and Knowledge Features
- Fusion Text and Knowledge

Approach Overview

- Extract Text and Knowledge Features



Encoder:

$$A_c = Conv\,(X)$$
$$A_m = MultiHead\,([A_c : X])$$
$$P = [A_c : A_m] \in R^{m*d}$$

Attention:

$$S = relu(W_c P^T)^T relu(W_q H^T)$$
$$a = softmax(S)$$
$$P' = a \bullet H$$
$$H' = a^T \bullet P$$

Matcher:

$$a_1 = G_1\left(\left[P; P'\right]\right)$$
$$a_2 = G_2\left(\left[P; P - P'\right]\right)$$
$$a_3 = G_3\left(\left[P; P \odot P'\right]\right)$$
$$C = G\,([a_1; a_2; a_3])$$

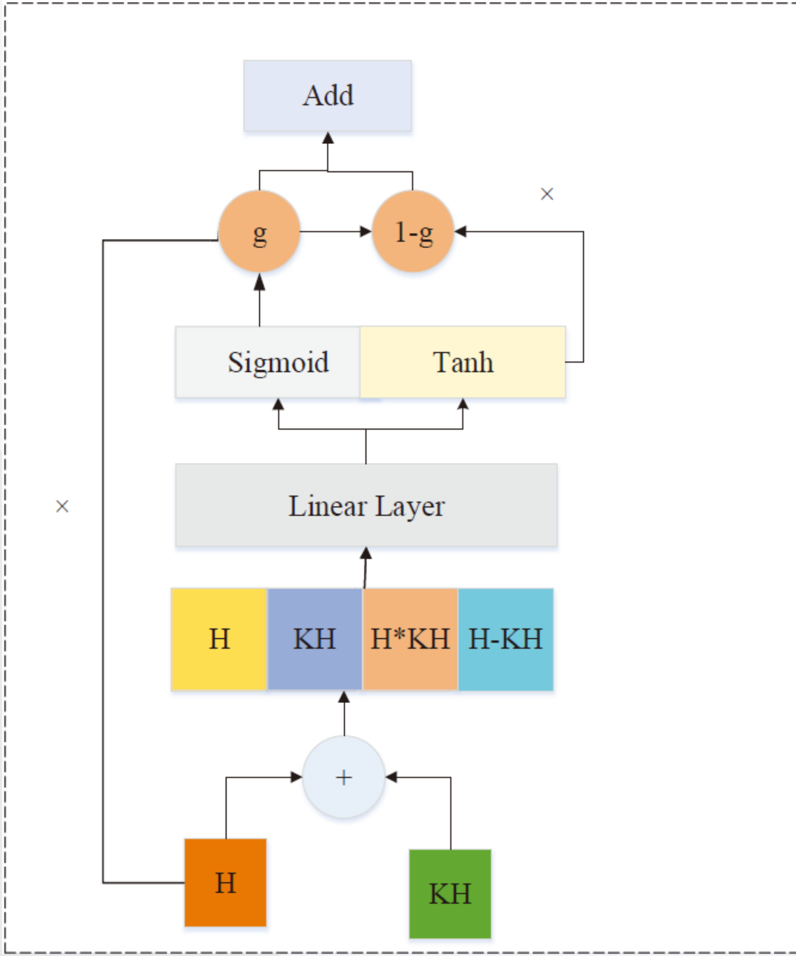
Bi-attention:

$$\alpha_t = softmax(S_{t:}) \in R^m$$
$$q_{:t} = \sum_i \alpha_{tj} P_{:j}$$
$$b = softmax(\max_{col}(S)) \in R^n$$
$$c = \sum_t b_t H_{t:} \in R^{2d}$$
$$G_{:t} = \beta(C_{:t}, H_{:t}, Q_{:t})$$
$$\beta(c, h, q) = [h; q; h \odot q; h \odot c] \in R^{8d}$$

Approach Overview

- Fusion Text and Knowledge

$$\widetilde{x} = \tanh(W_1[H; KH; H \odot KH; H - KH])$$
$$g = sigmoid(W_2[H; KH; H \odot KH; H - KH])$$
$$z = g \odot \widetilde{x} + (1 - g) \odot x$$



Experiments and Results

- baselines
  - SWEM
  - HBMP
  - DITM
  - RE2
  - DRr-Net
  - BIMPM
  - ESIM
  - BERT
  - MFAE

- datasets

dataset	train	validation	test
Snli	550152	10000	10000
SciTail	23596	1304	2126
Quora	384290	10000	10000
Sick	4500	500	4927

Experiments and Results

- Main Experiments

- Our model performs best

- the performance of representation-based model is slightly weaker than the interaction-based model

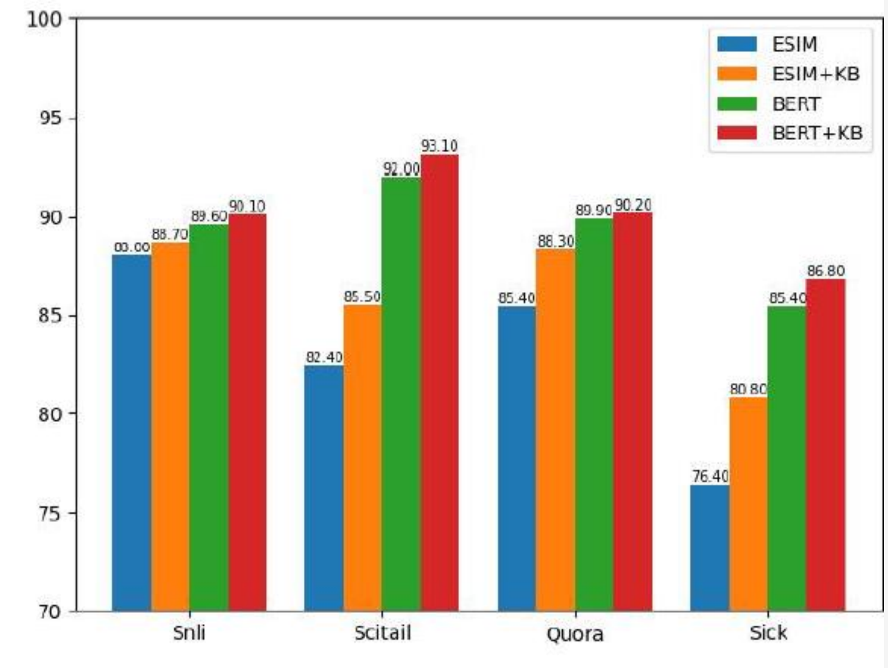
- addition of external knowledge has a positive effect on the overall performance of the model

Model	Snli	Scitail	Quora	Sick
SWEM[28]	83.8	-	83.0	-
HBMP[29]	86.6	86.0	-	-
DITM[30]	-	89.2	86.1	-
DRr-Net[31]	87.7	87.4	89.8	-
BIMPM* [25]	87.9	75.3	88.2	76.6
ESIM* [24]	88.0	82.4	85.4	76.4
RE2* [9]	88.7	86.6	89.4	79.8
MFAE[32]	90.0	-	90.5	-
BERT_base*[13]	89.6	92.0	89.9	85.4
KETM-KB*	88.9	89.5	90.1	80.1
KETM*	89.5	90.4	90.3	84.0
KETM-KB(BERT)*	90.2	92.1	90.7	86.8
KETM(BERT)*	**90.6**	**92.6**	**91.0**	**87.1**

Experiments and Results

- Analysis of the method generality

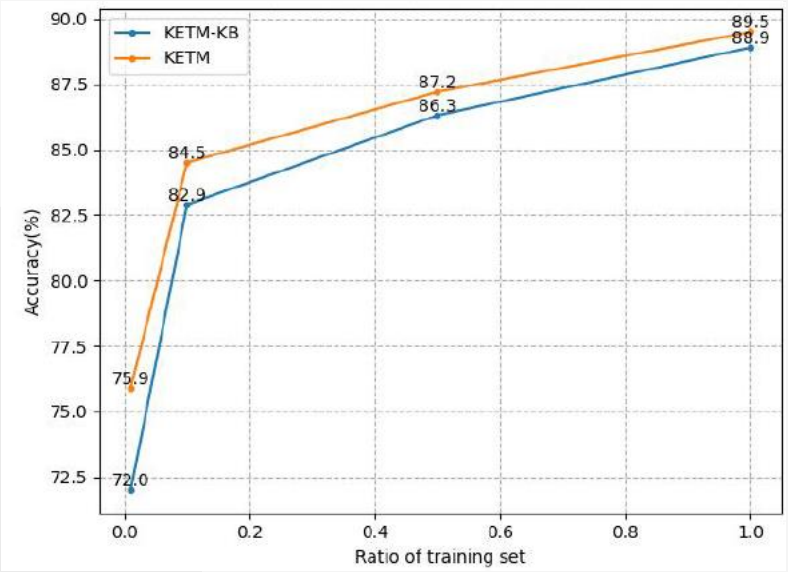
- We achieve better results on both the non-Transformer struct-based and Transformer struct-based models compared with no knowledge addition, which validates the effectiveness of our proposed knowledge enhancement framework.



Experiments and Results

- Analysis of the different train data

- external knowledge plays a important role in different training data sizes. Besides, the smaller the training data is, the greater improvement the model achieves..



Experiments and Results

- Analysis of the adversarial dataset

The cat is sleeping on the couch	The cat is resting on the couch	Entailment
The cat is sleeping on the couch	The pillow is on the couch	Neutral
The cat is sleeping on the couch	The cat is awake on the couch	Contradiction

The breakNLI dadaset is resemble SNLI, The text2 is generate be swapping words in the text1 so that world knowledge is required to make correct prediction

Experiments and Results

- Analysis of the adversarial dataset

- Our model performs best

- Adding knowledge can greatly improve the performance of the model

Model	BreakNLI
BIMPM*	68.3
RE2*	80.9
KIM	83.8
ESIM	65.8
ESIM+KB*	78.8
KETM-KB*	87.7
KETM*	**91.2**

Conclusion

- KETM,a framework to harness external knowledge for Text Macthing
- Using Wiktionary as an external knowledge for SNLI、Scitail、Quora、Sick datasets showed improved performance
- Our framework has good generality and can function as any text-matching model in the text matching layer

Thank You!