# Knowledge-aware Attentive Neural Network for Ranking Question Answer Pairs

YingShen, YangDeng, MinYang, YaliangLi, NanDu, WeiFan, KaiLei

Xiaofan Yan

# Outline

- Abstract
- Introduction
- Method
- Experiment
- Conclusion

# Abstract

- KABLSTM
  - (Knowledge-aware Bidirectional Long Short Term Memory)
  - Leverage knowledge graph to enrich the representational learning of QA sentence

# Introduction

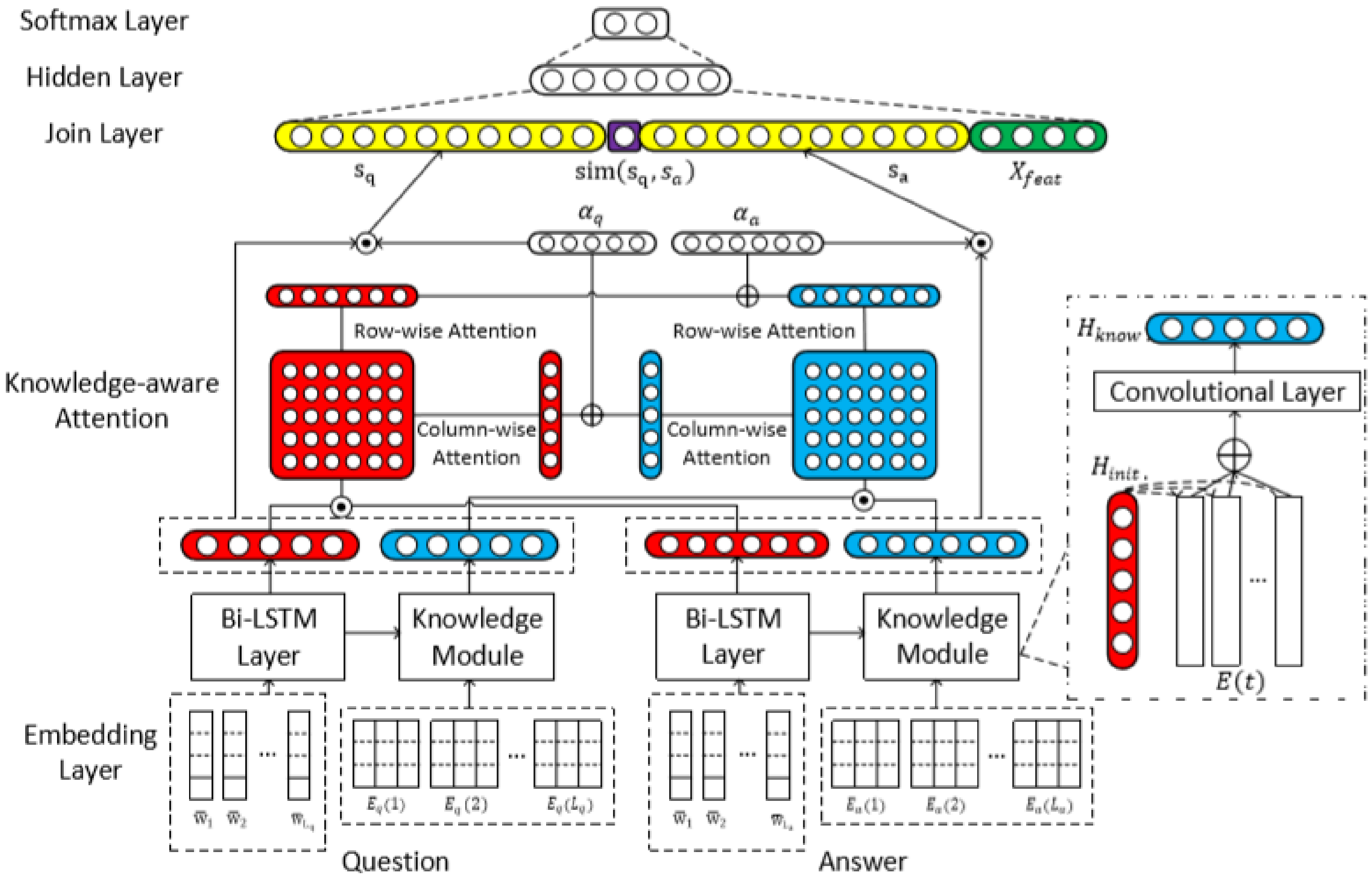
- QA(question answering)
  - Deep learning in natural language processing
  - Exploit attention mechanisms to learn interaction information
  - Challenge
    - Background knowledge from open-domain knowledge graphs, a higher score to negative answers since the negative answer is more similar to the questions
    - The issues of redundancy and noise prevalent

后面没有解释解决这个问题？

# Method

- Bidirectional Long Short-Term Memory
  - Learn the initial representations
- Knowledge model
  - Knowledge-based sentence representation learning
  - Context-guided attentive CNN to learn the knowledge-based sentence representation
- Knowledge-aware attention
  - Context-based sentence representation learning
  - Learn final knowledge-aware attentive sentence representation

# Method



Softmax Layer
Hidden Layer
Join Layer
$s_q$
$sim(s_q, s_a)$
$s_a$
$X_{feat}$
$\alpha_q$
$\alpha_a$
Row-wise Attention
Row-wise Attention
Knowledge-aware Attention
Column-wise Attention
Column-wise Attention
$H_{know}$
Convolutional Layer
$H_{init}$
$E(t)$
Bi-LSTM Layer
Knowledge Module
Bi-LSTM Layer
Knowledge Module
Embedding Layer
Question
Answer

# Experiment

- Experiment setup
  - Dataset: TREC QA& WIkiQA
  - Metrics: MAP&MRR
- Implementation details
  - Word embedding: GloVE
  - Knowledge graph: FB5M

# Experiment

- Experiment result

**Table 3: Result on TREC QA(original) (with ablation study)**

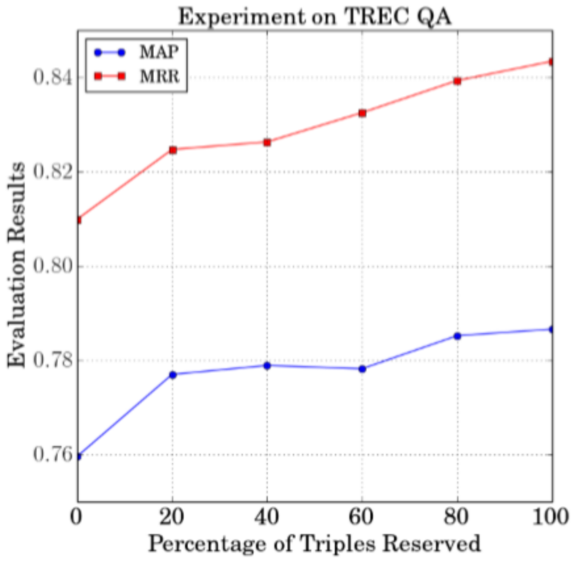
| Model | MAP | MRR |
|---|---|---|
| Wang & Nyberg (2015) [11] | 0.7134 | 0.7913 |
| Severyn & Moschitti (2015) [6] | 0.7459 | 0.8078 |
| Tay et al. (2017) [8] | 0.7499 | 0.8153 |
| Rao et al. (2016) [5] | 0.7800 | 0.8340 |
| Tay et al. (2018) [9] | 0.7712 | 0.8384 |
| KABLSTM | **0.7921** | **0.8444** |
| w/o attention | 0.7805 | 0.8309 |
| w/o KG | 0.7596 | 0.8099 |

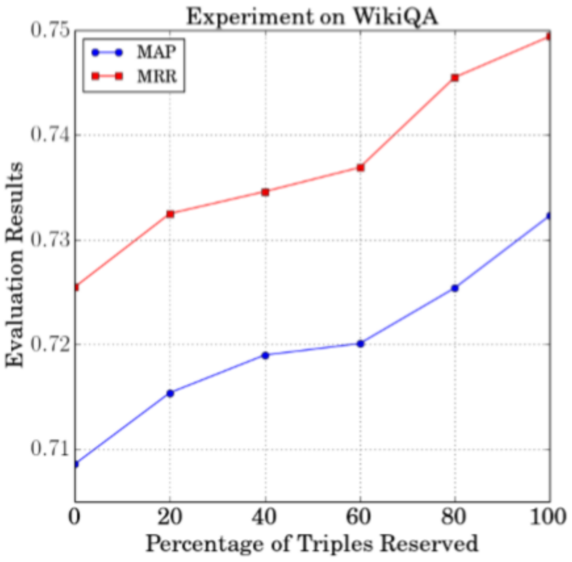
**Table 4: Result on TREC QA(cleaned) and WikiQA (with ablation study)**

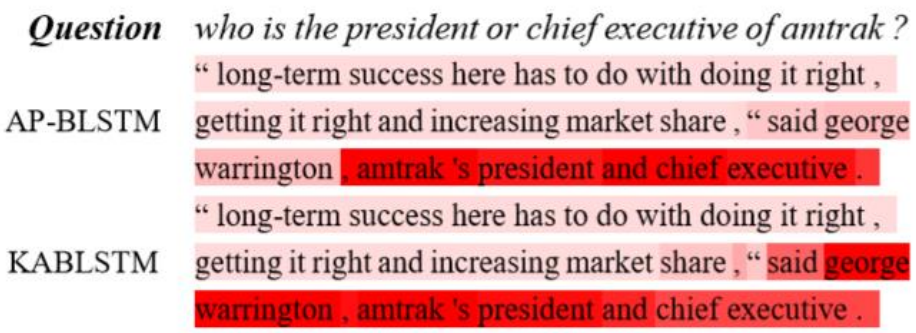
| Models | TREC QA(cleaned) | | WikiQA | |
|---|---|---|---|---|
| | MAP | MRR | MAP | MRR |
| Yang et al. (2015) [14] | 0.6951 | 0.7633 | 0.6520 | 0.6652 |
| Santos et al. (2016) [4] | 0.7530 | 0.8511 | 0.6886 | 0.6957 |
| Rao et al. (2016) [5] | 0.8010 | 0.8770 | 0.7010 | 0.7180 |
| Chen et al. (2017) [2] | 0.7814 | 0.8513 | 0.7212 | 0.7312 |
| Wang et al. (2016) [10] | 0.7369 | 0.8208 | **0.7341** | 0.7418 |
| KABLSTM | **0.8038** | **0.8846** | 0.7323 | **0.7494** |
| w/o attention | 0.7821 | 0.8654 | 0.7214 | 0.7363 |
| w/o KG | 0.7633 | 0.8320 | 0.7086 | 0.7255 |

# Experiment

- Analyze
  - Completeness of Knowledge Graph -- incomplete knowledge graph
  - Case Study -- visualizing the attention weight with baseline model



# Conclusion

- Propose a knowledge-aware attentive neural network for ranking QA pairs, which effectively incorporate external knowledge from KGs into sentence representational learning.