An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge

Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhangyi Liu, Hua Wu and Jun Zhao

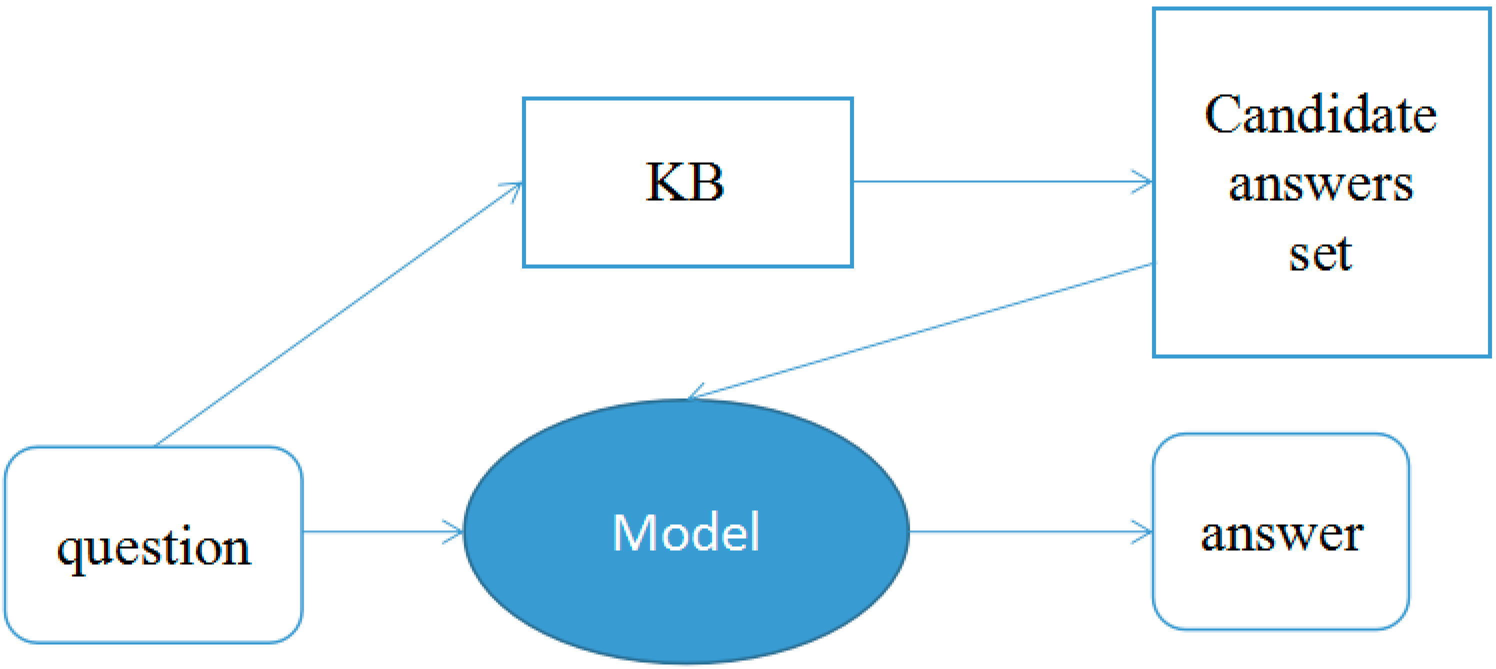
National Laboratory of Pattern Recognition, Institute of Automation,

Chinese Academy of Sciences, Beijing, 100190, China

University of Chinese Academy of Sciences, Beijing, 100049, China

Baidu Inc., Beijing, 100085, China

Task



Motivation

1.The information of answers is important for question encoding.

2.The different aspects of answer make different influences.

ex1.
Q: Who is the president of France?
A: Francois Hollande
Focus: "president" and "France"
ex2.
Answer type: <business/board_member>
Focus: Who

Challenges

1.How to conduct question representation according to different answer aspects.

2.How to solve the KB resources limitation problem.

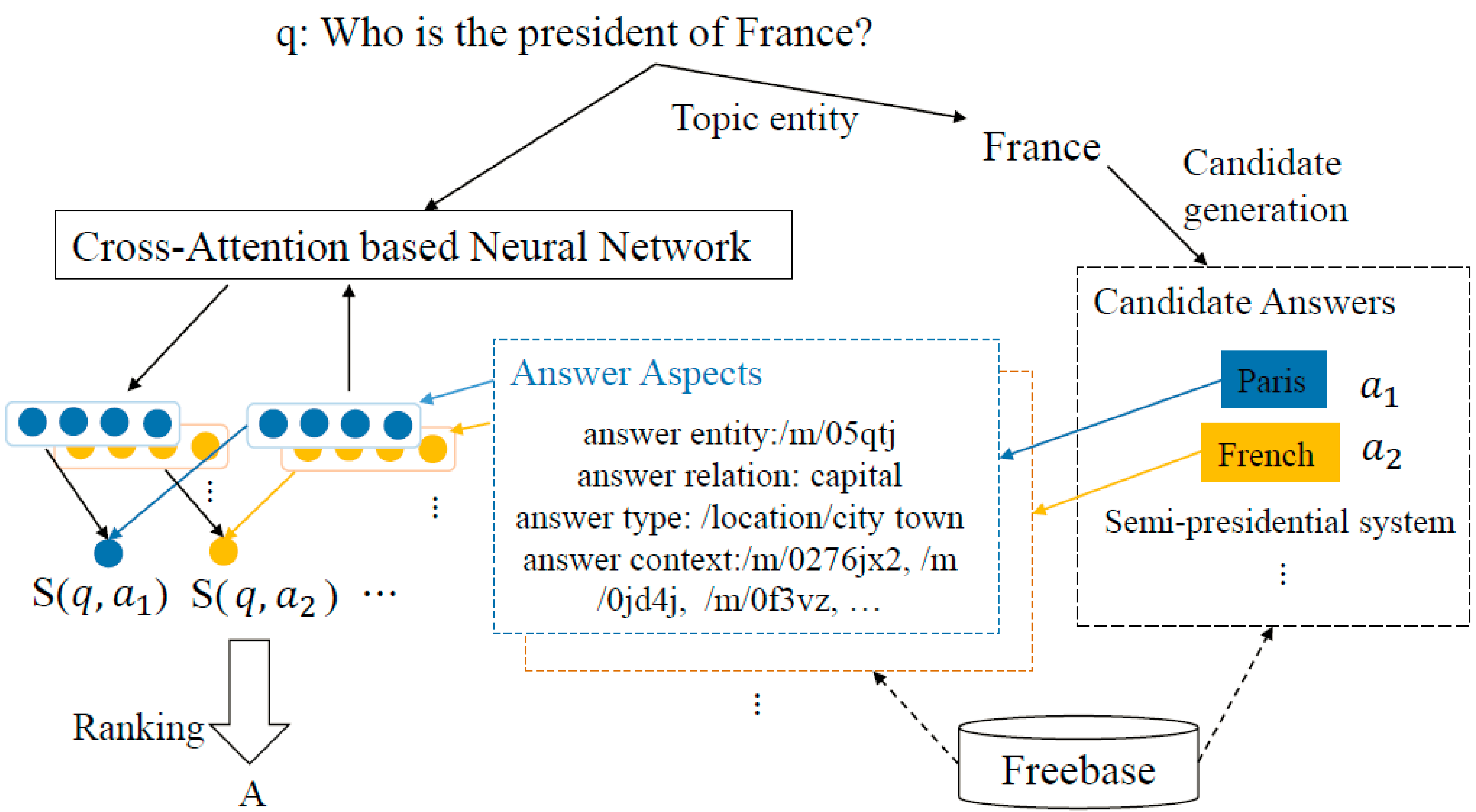
(1) The global information of KB is deficient.

(2) OOV problem

Contributions

1. Cross-attention based NN model
2. Introduce a muli-task learning to leverage the global KB information

Overview



q: Who is the president of France?
Topic entity
France
Candidate generation
Cross-Attention based Neural Network
Candidate Answers
Paris
a_1
French
a_2
Semi-presidential system
⋮
Answer Aspects
answer entity:/m/05qtj
answer relation: capital
answer type: /location/city town
answer context:/m/0276jx2, /m /0jd4j, /m/0f3vz, …
⋮
$S(q, a_1)$ $S(q, a_2)$ ⋯
Ranking
A
Freebase

Candidate Genearation

FreeBase API(question) = topic entity (86% questions)

ex.

Q: Who is the president of France?

Topic entity: France

Then collect all the entities directly connected to it and the ones connected with 2-hops as candidate set.

Model

qustion representation

1.Embedding(question) -> question embedding

2.bi-LSTM(question embedding) -> question hidden representation

answer aspect represeatation

four answer aspects: (1)answer entiey a_e, (2)answer relation a_r, (3)answer type a_t, and (4)answer context a_c.

Embedding(a_i) = e_i , i $\in$ {e,r,t,c.}

Cross-Attention model

Intuition: view the question from different aspects of answers. Each answer aspect should focus on different words of question.

Answer-towards-question attention

For each e_i, compute a question representation.

$$\alpha_{ij} = \frac{exp(w_{ij})}{\sum_{k=1}^{n} exp(w_{ik})}$$

$$w_{ij} = f(W^T[h_j; e_i] + b)$$

$$q_i = \sum_{j=1}^{n}(\alpha_{ij} h_j)$$

$$S(q, e_i) = h(q_i, e_i)$$

Question-towards-answer attention

How to utilize the four question representation?

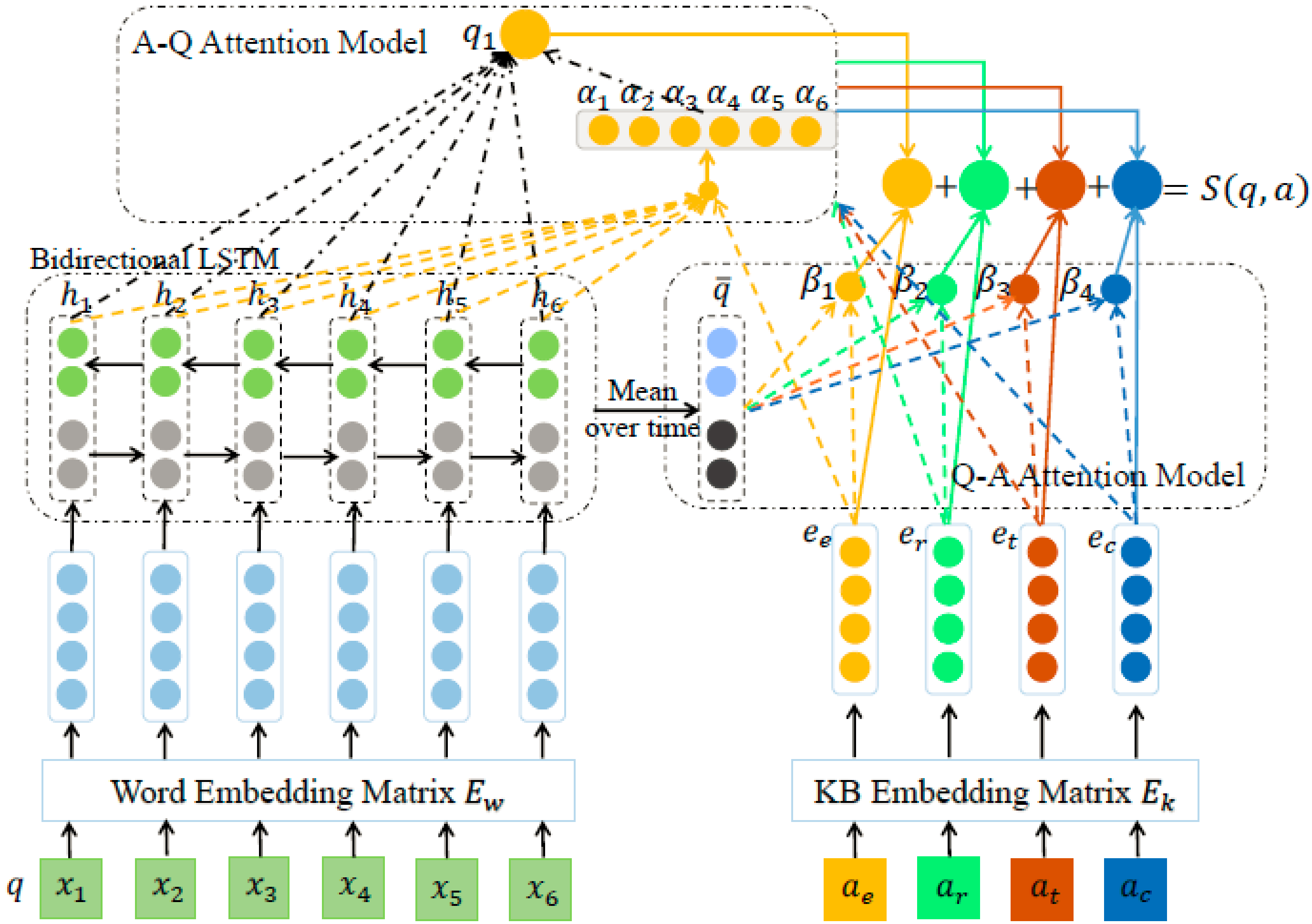
$$S\left(q,a\right)=\sum_{e_i\in\{e_e,e_r,e_t,e_c\}}\beta_{e_i}S\left(q,e_i\right)$$

$$\beta_{e_i}=\frac{\exp\left(\omega_{e_i}\right)}{\sum_{e_k\in\{e_e,e_r,e_t,e_c\}}\exp\left(\omega_{e_k}\right)}$$

$$\omega_{e_i}=f\left(W^T[\overline{q};e_i]+b\right)$$

$$\overline{q}=\frac{1}{n}\sum\nolimits_j^n h_j$$

Architecture



Training

Regard (q,a) pairs as supervision data, candidate set C_q can divided into P_q (Correct) and N_q (Wrong). For each correct answer $a \in P_q$, randomly select k wrong answers $a' \in N_q$.

hinge loss:
$L_{q,a,a'} = max(0, \gamma + S(q,a') - S(q,a))$

Inference

Calculate $S(q, a)$ for each $a \in C_q$, and find out the maximum value S_{max}.

$$S_{max} = argmax_{a \in C_q} S(q, a)$$

If question has more than one answer.

$$A = \{\hat{a} | S_{max} - S(q, \hat{a}) < \gamma\}$$

Combining Global Knowledge

Adopt TransE model and integrate its outcome into training process.

TransE

Relations are considered as translations in embedding space.

possitive fact: (s, p, o)

corrupted fact: (s', p, o') (randomly samle)

The distance measure $d(s+p, o)$ is defined as $||s+p-o||^2$.

Loss

$$L_k = \sum_{(s,p,o)\in S} \sum_{(s',p,o')\in S'} \left[\gamma_k + d\left(s+p, o\right) - d\left(s'+p, o'\right)\right]_+$$

Conduct KB-QA training and TransE training in turn.

Experiments

dataset: WeqQuestions

1.BOW representation

2.BOW representation + answer feature

3.SP-based

4.CNN

5.Memory-NN

Methods	Avg F_1
Bordes et al., 2014b	29.7
Bordes et al., 2014a	39.2
Yang et al., 2014	41.3
Dong et al., 2015	40.8
Bordes et al., 2015	42.2
our approach	**42.9**

Experiments

A-Q-ATT: answer-towards-question attention part

C-ATT: cross-attention

GKI: global knowledge information

Methods	Avg F_1
LSTM	38.2
Bi_LSTM	39.1
Bi_LSTM+A-Q-ATT	41.6
Bi_LSTM+C-ATT	41.8
Bi_LSTM+GKI	40.4
Bi_LSTM+A-Q-ATT+GKI	42.6
Bi_LSTM+C-ATT+GKI	42.9

Thank you