Attention Mechanism on Question Answering over Knowledge Bases

Yuanzhe Zhang, **Shizhu He**, Kang Liu and Jun Zhao

National Laboratory of Pattern Recognition
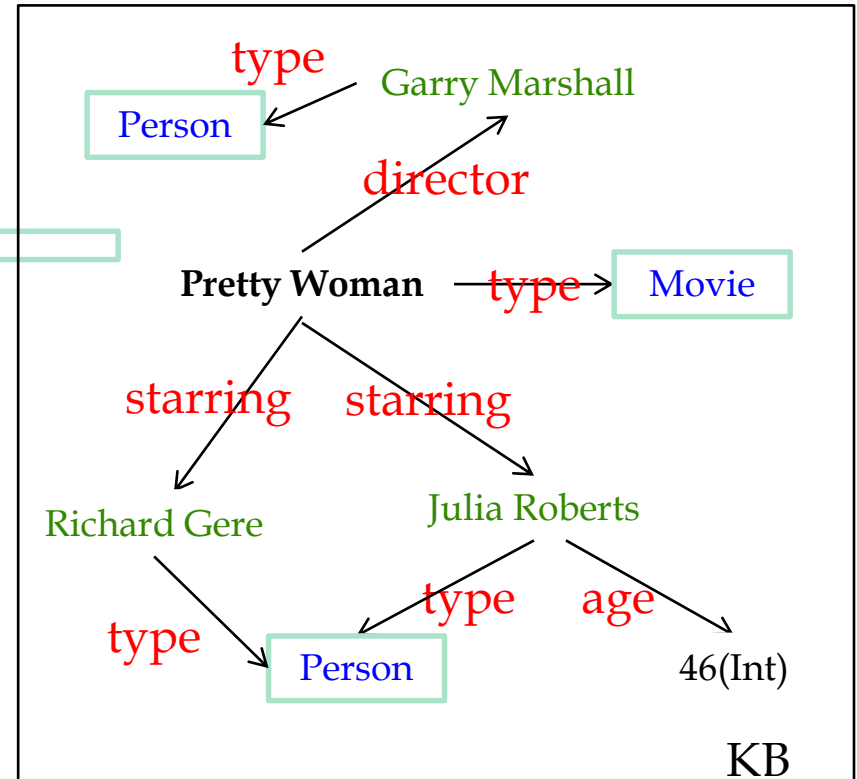Institute of Automation, Chinese Academy of Sciences

Outline

- Task

- Introduction

- Our Proposed Approach

- Experiments

Question Answering over Knowledge Base

In which films directed by Garry Marshall was Julia Roberts starring?
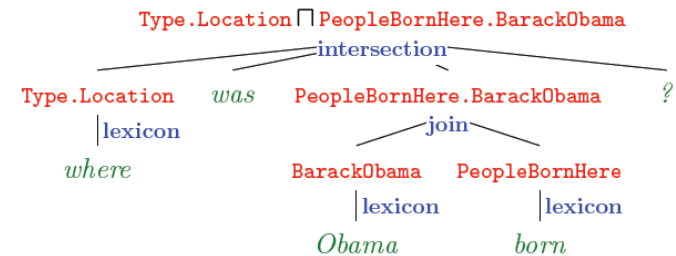
Question Answering System

Pretty Woman…

type

Garry Marshall

Person

director

Pretty Woman —type→ Movie

starring   starring

Richard Gere   Julia Roberts
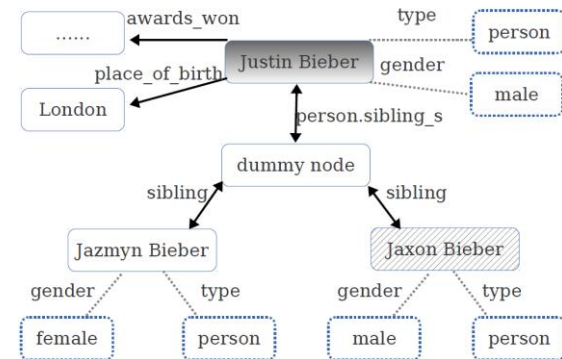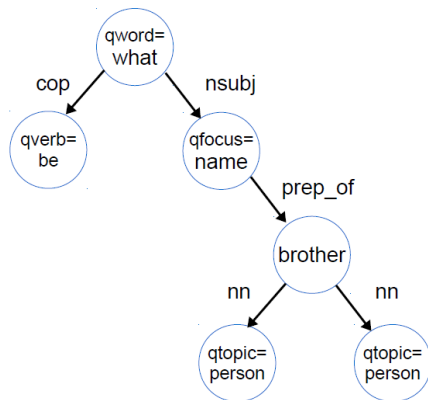
type   type   age

Person   46(Int)

KB

Question Answering over Knowledge Base (KB-QA)

- Semantic Parsing : transform questions into formal expressions, and query answer over KBs

  where was Obama born?

- Retrieval-based QA : search candidate answers, and ranking

  what is the name of justin bieber brother?

Retrieval-based QA

topic

What is the name of Justin Bieber brother?

Freebase

Search

Ranking
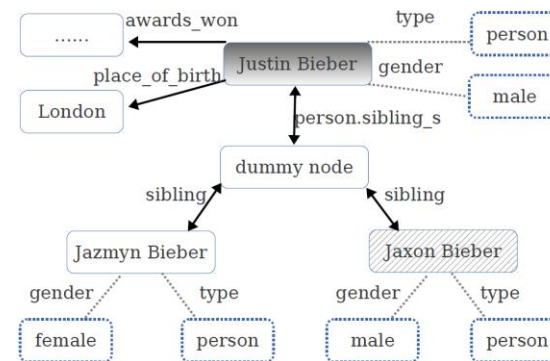
5

Retrieval-based QA

[Yao et al, ACL 2014]

- # Feature-based Ranking
  - From question
  - From knowledge base
  - From matching
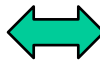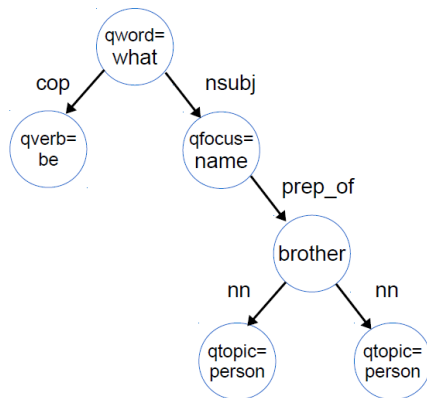
wgt.	feature
5.56	qfocus=money|type=Currency
5.35	qverb=die|type=Cause_Of_Death
5.11	qword=when|type=datetime
4.56	qverb=border|rel=location.adjoins
3.90	qword=why|incoming_relation_rank=top_3
2.94	qverb=go|qtopic=location|type=Tourist_attraction
-3.94	qtopic=location|rel=location.imports_exports.date
-2.93	qtopic=person|rel=education.end_date

Table 6: A sample of the top 50 most positive/negative features. Features are production between question and node features (c.f. Figure 1).

7 million features

what is the name of justin bieber brother?

6

Retrieval-based QA with Subgraph Embedding

[Bordes et al, EMNLP 2014]
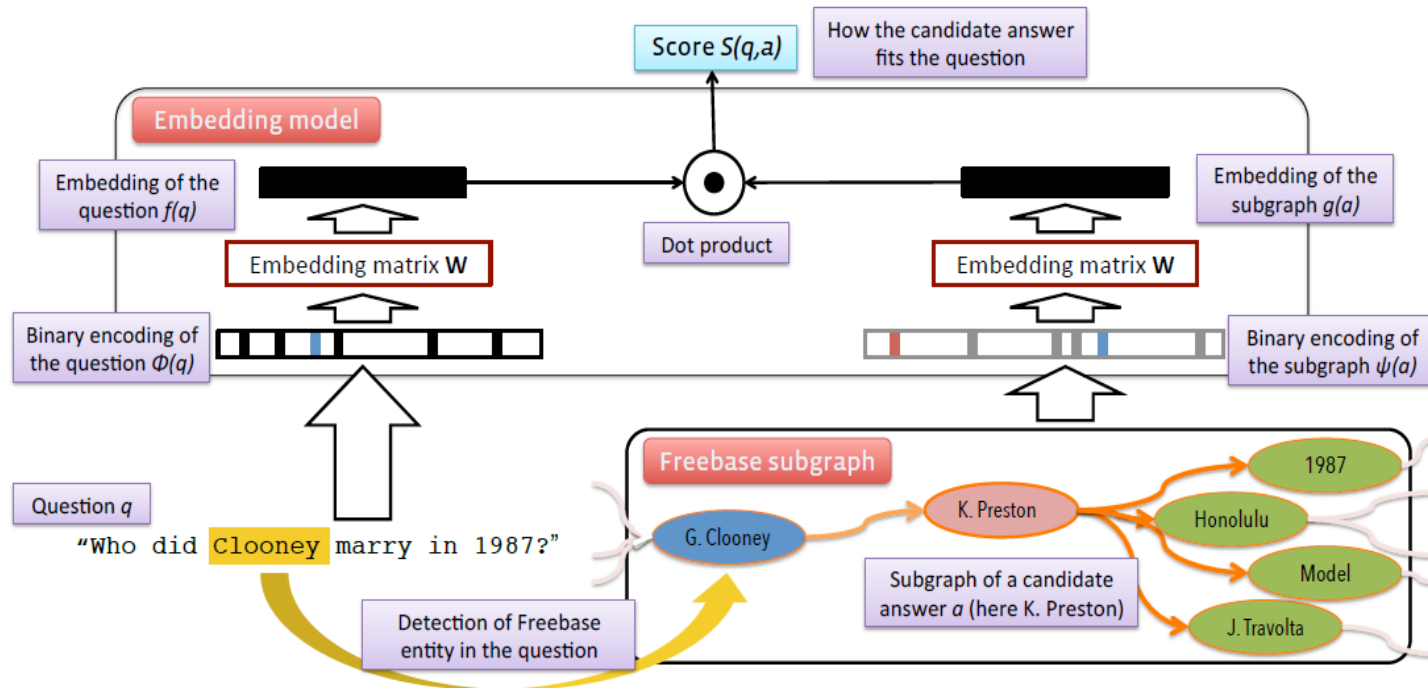
Representing candidate answers

1. single entity   [….0…..1……0…..]

2. path representation

   1-hop: question entity -> relation -> answer entity

   2-hop: question entity -> relation1 -> relation2 -> answer entity

3. subgraph representation
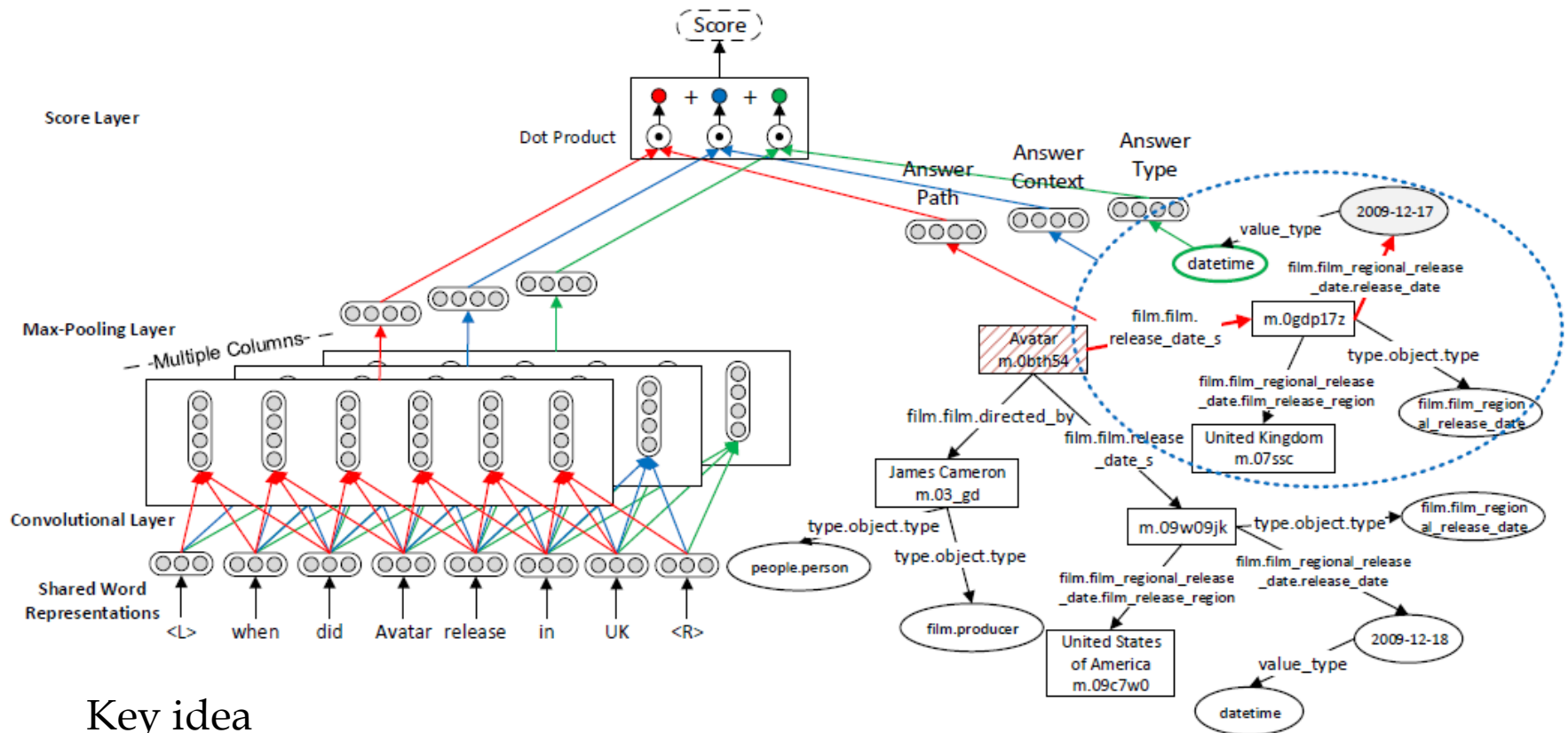
Key idea
- Represent question with CNN
- Different evidences own different compositions
  - Entity-topic path
  - Entity type
  - Other contexts

Neural Machine Translation by Jointly Learning to Align and Translate [Dzmitry et al, ICLR 2015]

Key idea:
Tell the translator what is now translated.
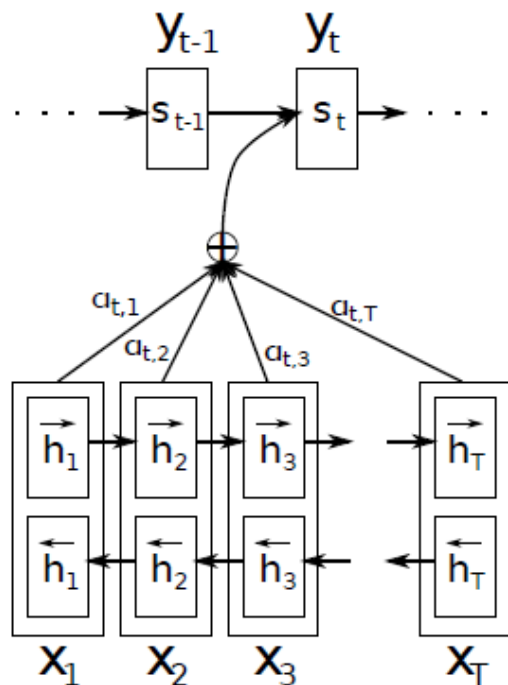
$$p(y_i|y_1,\ldots,y_{i-1},\mathbf{x}) = g(y_{i-1}, s_i, c_i)$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

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

$$e_{ij} = a(s_{i-1}, h_j)$$

$$a(s_{i-1}, h_j) = v_a^\top \tanh(W_a s_{i-1} + U_a h_j)$$

Neural Machine Translation by Jointly Learning to Align and Translate [Dzmitry et al, ICLR 2015]
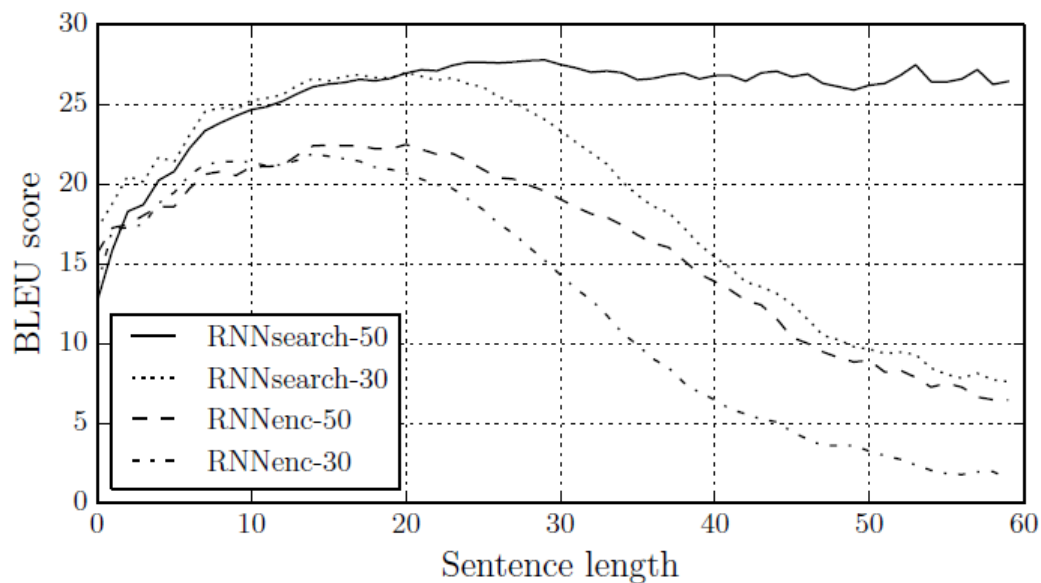
Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

Outline

- Task

- Introduction

- Our Proposed Approach

- Experiments

KB-QA with Attention-based Neural Networks

q: Who is the president of France?

Topic entity → France

Candidate generation

Attention-base Neural Network

Answer Aspects

answer entity: /m/05qtj
answer relation: capital
answer type: /location/citytown
answer context: /m/0276jx2, /m/0jd4j, /m/0f3vz, ...

Candidate Answers

Paris a_1

French a_2

Semi-presidential system

$S(q,a_1)$ $S(q,a_2)$ ...

Ranking

A

Freebase
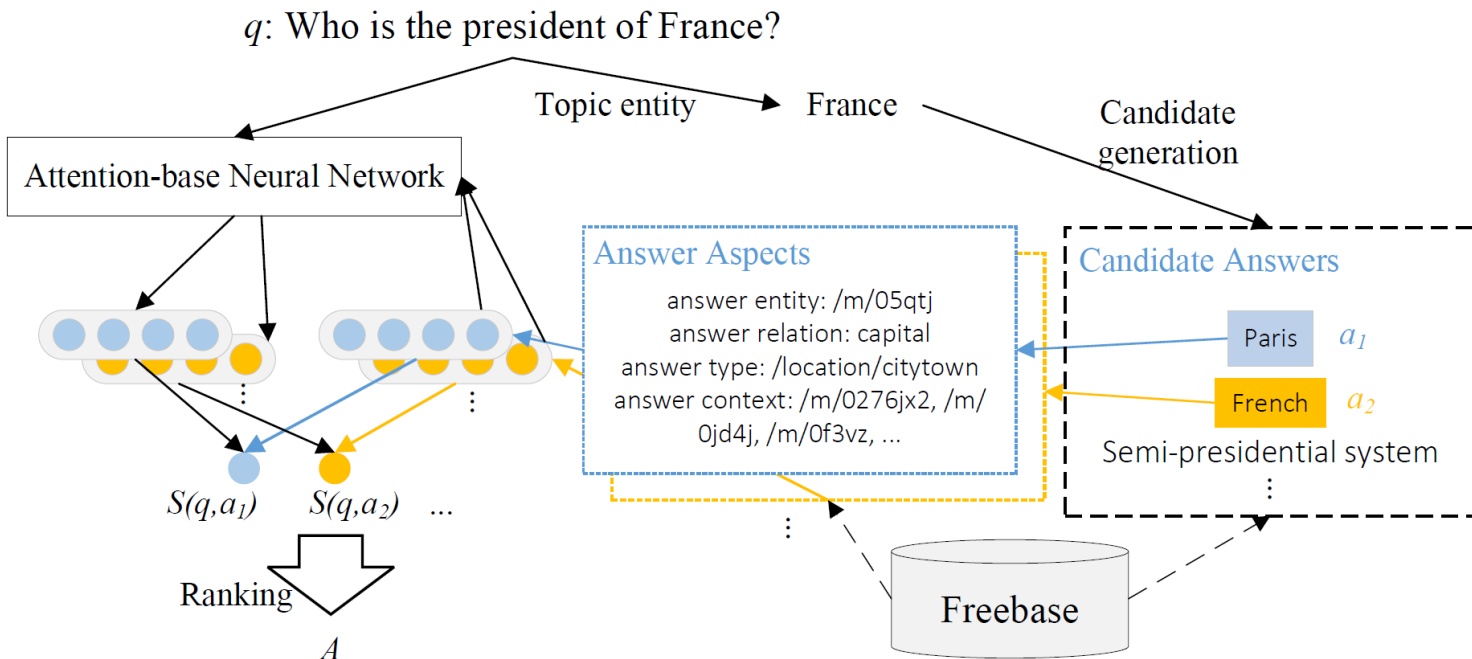
Search topic → Generate candidates → Answer Representation → **Attention Mechanism** → Question Representation
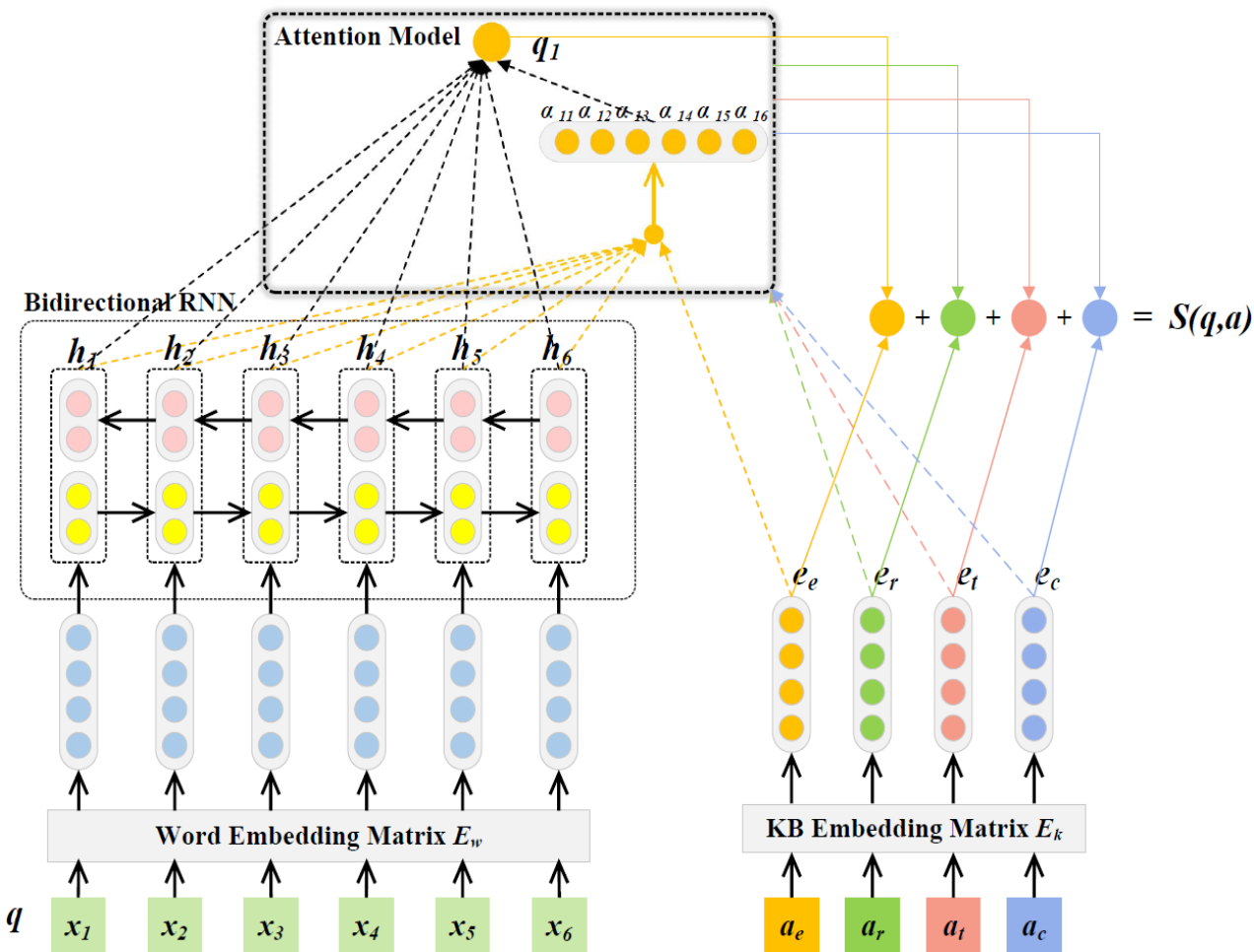
Global Information

Scoring

KB-QA with
Attention-based Neural Networks

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

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

$$q_i = \sum_{j=1}^{L_q} \alpha_{ij} h_j$$

$$S(q, a) = \sum_{e_i \in \{e_e, e_r, e_t, e_c\}} q_i \cdot e_i$$

KB-QA with Attention-based Neural Networks

Training :

Margin-based Ranking Loss

$$L_{q,a,a'} = [\gamma + S(q, a') - S(q, a)]_+$$

Objective function :

$$\min \sum_q \frac{1}{|P_q|} \sum_{a \in P_q} \sum_{a' \in N_q} L_{q,a,a'}$$

Inference:

$$S_{\mathrm{max}} = \arg\max_{a \in C_q} \{S(q, a)\}$$

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

KB-QA with
Attention-based Neural Networks

- ✓ 知识表示
    - ✓ 受限于训练语料
    - ✓ 未考虑全局信息

- ✓ 融入知识库全局信息
    - ✓ 利用了更多的知识库信息，知识库的资源表示更准确
    - ✓ 缓解未登录资源问题

TransE：将知识库资源被表示在低维向量空间上
使用(s, p, o)来表示三元组，关系是向量空间上的一种操作
尾实体o的向量应该接近头实体s的向量加上关系p的向量，即 **s + p = o**

$$d(s + p, o) = \|s + p - o\|_2^2$$

Loss Function：

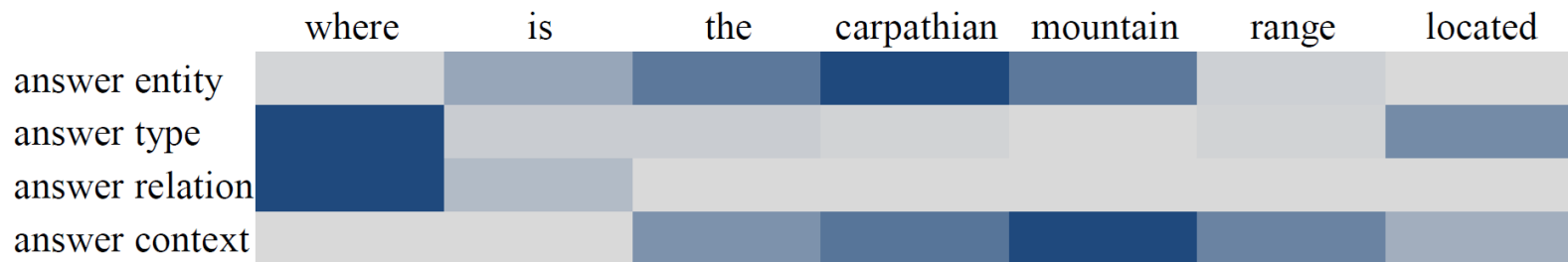$$L_k = \sum_{(s,p,o) \in S} \sum_{(s',p,o') \in S'} [\gamma_k + d(s + p, o) - d(s' + p, o')]_+$$

Experiments

Method	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
ours	**42.6**

Ours vs. NN models

Method	F_1
LSTM	38.2
Bi_LSTM	38.9
Bi_LSTM + ATT	41.6
Bi_LSTM + GKI	40.4
Bi_LSTM + ATT + GKI	42.6

+different parts

	where	is	the	carpathian	mountain	range	located
answer entity							
answer type							
answer relation							
answer context							

Attention heat map

16

Thanks

Email:
shizhu.he@nlpr.ia.ac.cn