Employing External Rich Knowledge for Machine Comprehension

中国科学院自动化研究所
模式识别国家重点实验室

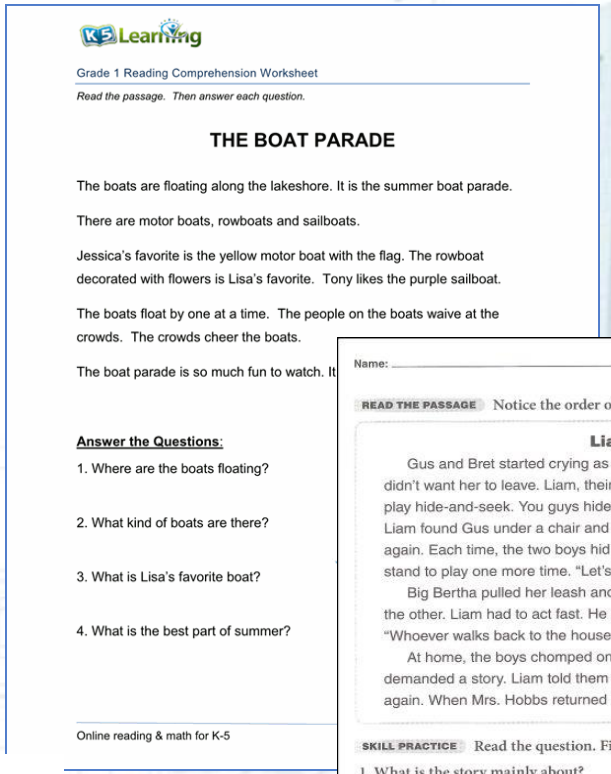
王炳宁 郭尚敏 刘康 赵军

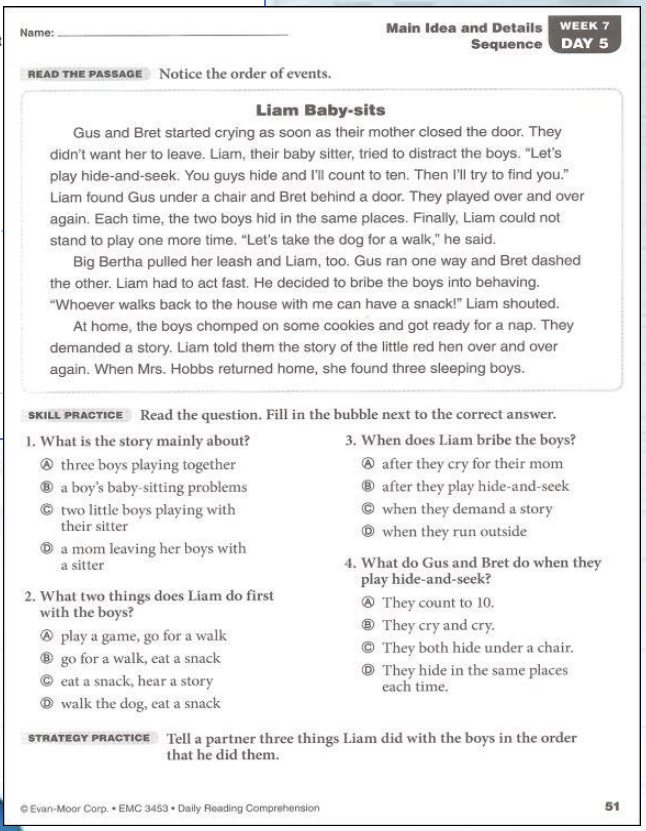
What is machine comphrehension?

QA

MC

VS.





Dataset

Document

One night I was at my friend's house where he threw a party. We were enjoying our dinner at night when all of a sudden we heard a knock on the door. I opened the door and saw this guy who had scar on his face. (.....)As soon as I saw him I ran inside the house and called the cops. The cops came and the guy ran away as soon as he heard the cop car coming. We never found out what happened to that guy after that day.

Question

1: What was the strange guy doing with the friend?

Candidate answers

A) enjoying a meal
B) talking about his job
C) talking to him
*D) trying to beat him

2: Why did the strange guy run away?

*A) because he heard the cop car
B) because he saw his friend
C) because he didn't like the dinner
D) because it was getting late

Dataset

Mctest： Richardson M, Burges C J C, Renshaw E. MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text //EMNLP. 2013, 1: 2.

分为MC160和MC500

Split	MC160		MC500	
	Documents	Questions	Documents	Questions
Train	70	280	300	1200
Dev	30	120	50	200
Test	60	240	150	600

Document: 150-300 words

Question: About 10 words

Other Resources:

Facebook : bAbI[1]

Google Deepmind: CNN and Daily Mail articles[2]

Facebook: CBTest[3]

Stanford: ProcessBank [4]

[1] Weston J, Bordes A, Chopra S, et al. Towards ai-complete question answering: A set of prerequisite toy tasks[J]. arXiv preprint arXiv:1502.05698, 2015.
[2] Hermann K M, Kocisky T, Grefenstette E, et al. Teaching machines to read and comprehend[C]//Advances in Neural Information Processing Systems. 2015: 1684-1692.
[3] Hill F, Bordes A, Chopra S, et al. The Goldilocks Principle: Reading Children‘s Books with Explicit Memory Representations[C].// ICLR. 2016.
[4] Berant J, Srikumar V, Chen P C, et al. Modeling Biological Processes for Reading Comprehension[C]//EMNLP. 2014.

Traditional

由于MCtest数据集较小，基于大数据的深度学习技术很难应用到其中去。以往的方法往往是依靠一些现有的自然语言处理工具抽取文档特征然后以此对文档进行建模。

Lexical Matching

Discourse Relation

Syntactic Parsing

Dependency Parsing

Word Representation Similarity

.....

这些特征直观清楚，但是构建这些特征费时费力。并且在实际中虽然能对阅读理解的一部分问题有效的处理，但是特征工程的方法相对于最基本的非监督的基线系统的提升非常有限。

Proposed Method

和传统问答系统一样，标准阅读理解也可以分为两部分。即答案选择（Answer Selection）和答案生成（Answer Generation）

答案选择（Answer Selection ）：根据问题，一系列文档或者句子中选择能回答这个问题的片段。

答案生成（Answer Generation）：根据问题和上一步选择出的片段生成出最后的答案。

One night I was at my friend's house where he threw a party. We were enjoying our dinner
at night when all of a sudden we heard a knock on the door. I opened the door and saw this
guy who had scar on his face. (.....)As soon as I saw him I ran inside the
house and called the cops. The cops came and the guy ran away as soon as he heard the cop
car coming. We never found out what happened to that guy after that day.

1: What was the strange guy doing with the friend?
A) enjoying a meal
B) talking about his job
C) talking to him
*D) trying to beat him

Proposed Method

Document: D
Document sentences: $\{s_1,\ldots,s_n\}$
Question: q
Candidate Answers: $\{a_1,\ldots,a_4\}$

$$p(a|q,D) = p(S|q,D)p(a|q,S)$$

答案选择 答案生成

Proposed Method

One night I was at my friend's house where he threw a party.
We were enjoying our dinner at night when all of a sudden we heard a knock on the door.
(.....)
A few seconds later I heard a noise from outside and when I ran outside I saw that the strange guy was trying to beat my friend.
As soon as I saw him I ran inside the house and called the cops.
The cops came and the guy ran away as soon as he heard the cop car coming.
We never found out what happened to that guy after that day.

$|q, S)$

生成

What was the strange guy doing with the friend?

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$$

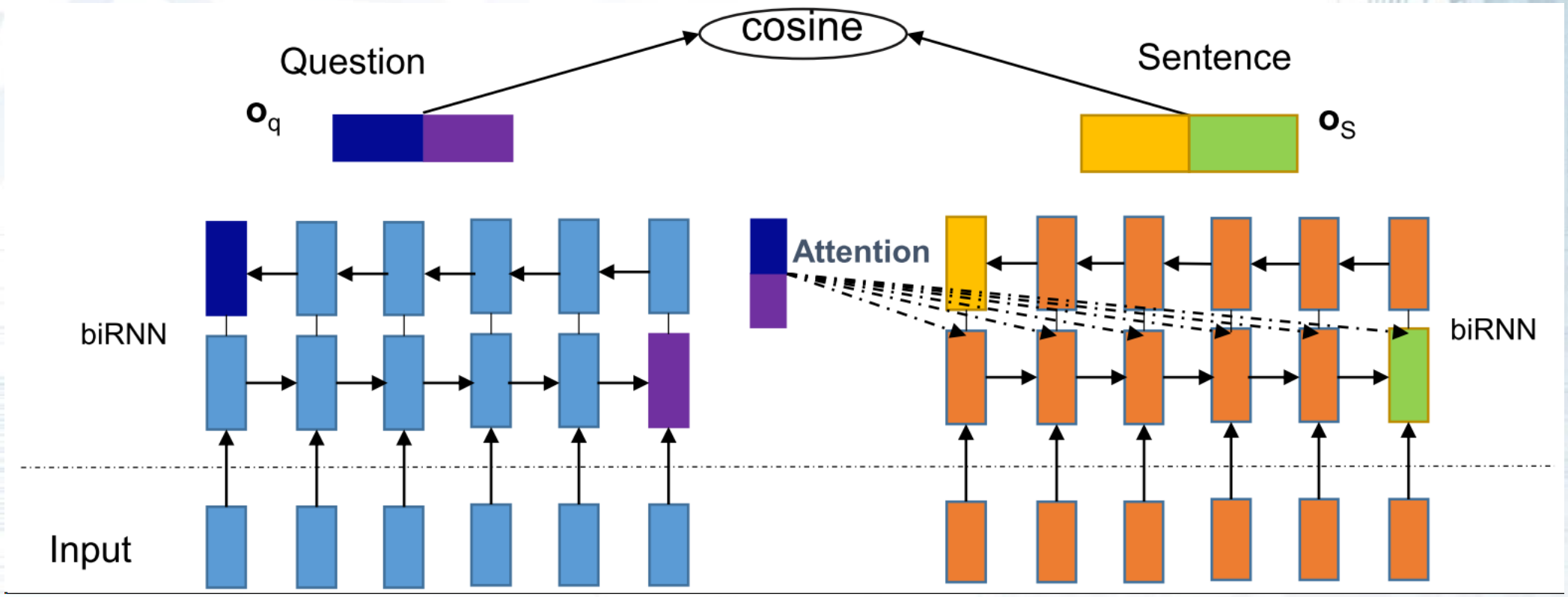
$$\mathbf{h}_t = \sigma(\mathbf{W}_{ih}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h)$$

$$s_t \propto \mathbf{h}_t^T \mathbf{W}_{qo} \mathbf{h}_n^q$$

$$\tilde{\mathbf{h}}_t = s_t \mathbf{h}_t$$

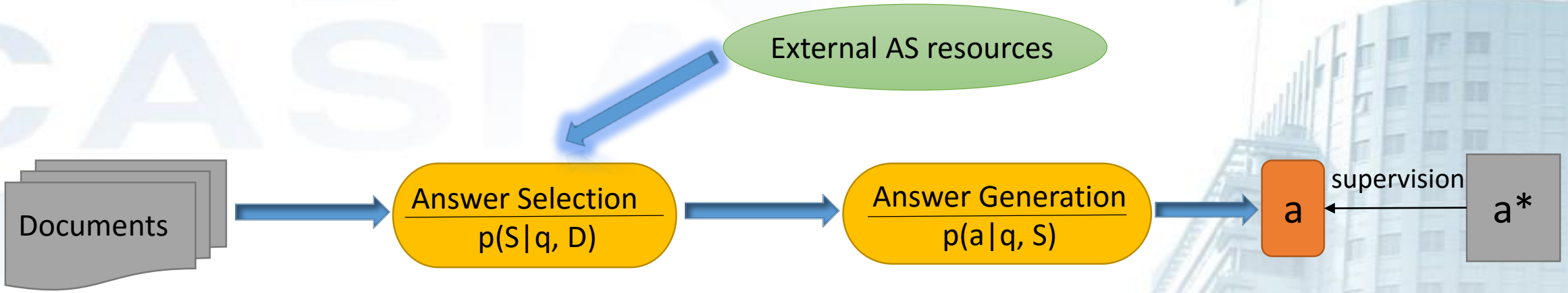
Proposed Method

Answer selection model architecture



Proposed Method

WikiQA InsuranceQA TrecQA ...



$$L_{AS}(q, D) = \sum_{s \in D} P(s|q, D; \theta_{RNN}) \log Q(s|q, D)$$

在外部的Answer selection数据上训练的一个模型

After we had dinner the strange guy asked my friend to come outside because he wanted to talk to him. A few seconds later I heard a noise from outside and when I ran outside I saw that the strange guy was trying to beat my friend. I ran towards him and asked him to leave our house. He left the house but after half an hour he came back and this time he was dressed up as superhero and he also had a sword. As soon as I saw him I ran inside the house and called the cops. The cops came and the guy ran away as soon as he heard the cop car coming. We never found out what happened to that guy after that day.

Q: Why did the strange guy run away?

A) because he heard the cop car
B) because he saw his friend
C) because he didn't like the dinner
D) because it was getting late

The strange guy run away beacuse he heard the corp car

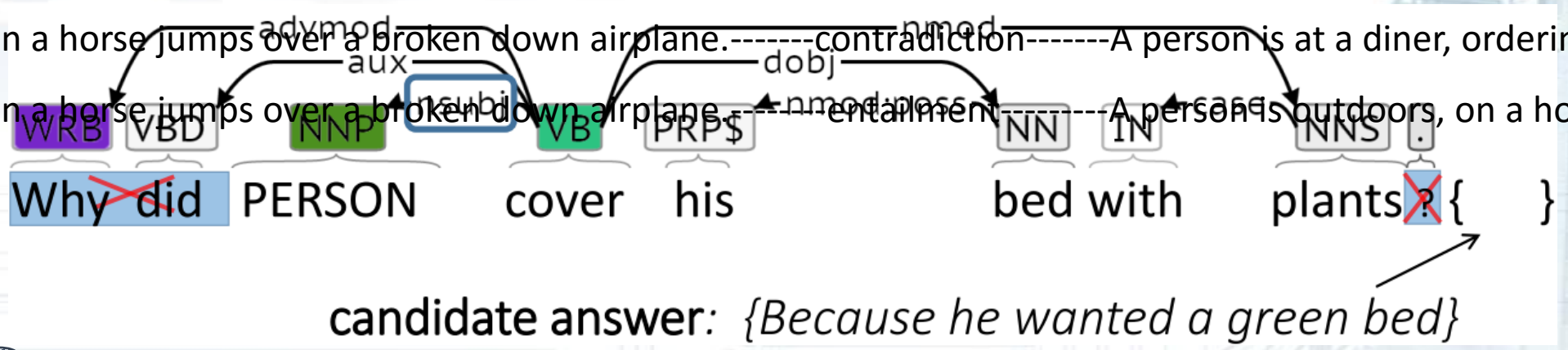
Proposed Method

Recognizing Textual Entailment-RTE

A person on a horse jumps over a broken down airplane.----------neutral----------A person is training his horse for a competition.

A person on a horse jumps over a broken down airplane.-------contradiction------A person is at a diner, ordering an omelette.

A person on a horse jumps over a broken down airplane.-------entailment--------A person is outdoors, on a horse.



Dataset:

SNLI SICK ...

9 rules to transform a question and answer candidate to a statement

Proposed Method

Combine External RTE knowledge with robust lexical feature

$$P(a|s,D) = [\beta P(s_q|s;\theta_1) + (1-\beta)P(s_q|s;\theta_{RTE})]$$

lexical matching

RTE Score

$$\beta = similarity(s_q^-, s)$$

s_q^- 代表我们将候选陈述句中的答案替换为一个符号ANSWER

similarity 是我们设计的一个词项匹配相似度，主要由n-gram constituency match,dependency match构成，用以判断两个句子之间是否有明显的词项联系

Proposed Method

$$P(a|q, s; \theta_1)$$

我们采用如下的词项特征：

- bag-of-words matching
- dependency root matching
- constituency sub-tree matching
- dependency path matching
- name entity matching
- digit matching

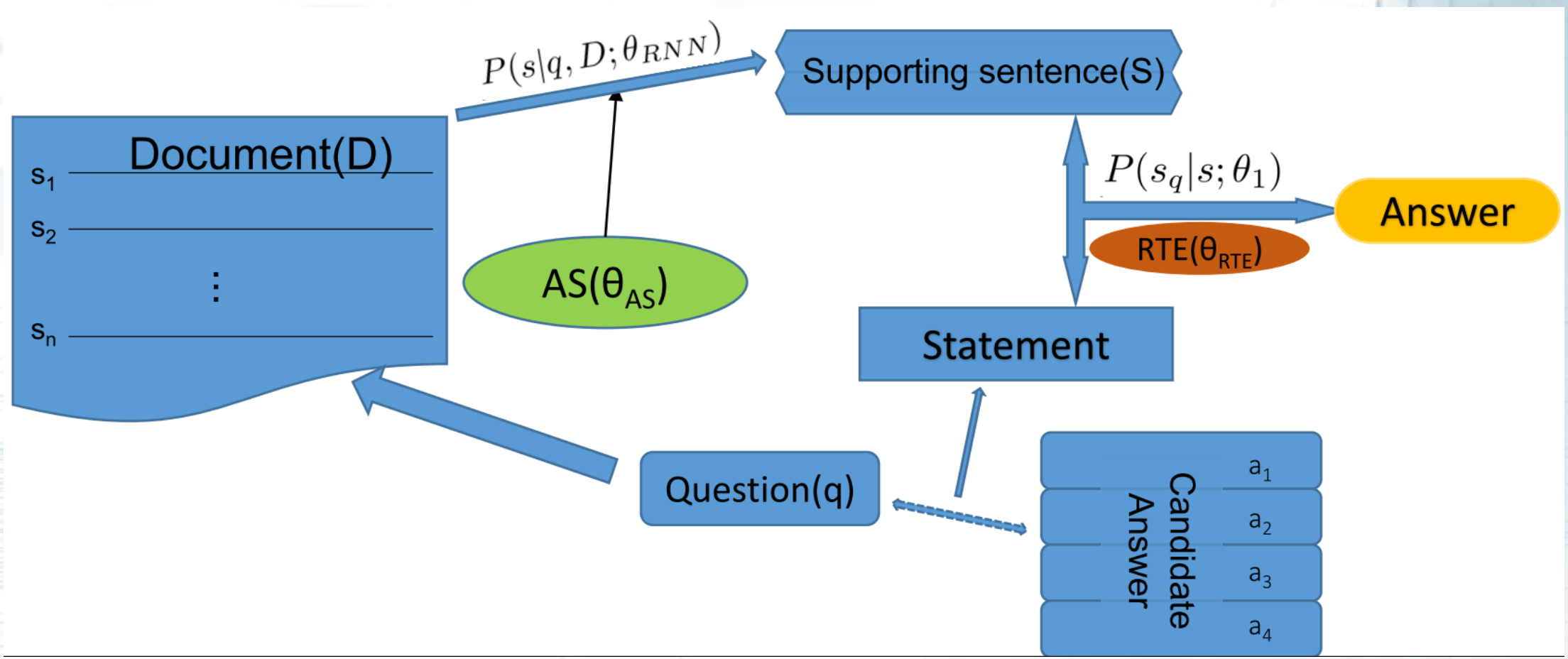
Proposed Method

$$P(a_{ij}^*|q_{ij}) = P(s|q, D; \theta_{RNN})$$
$$* [\beta P(s_q|s; \theta_1) + (1-\beta) P(s_q|s; \theta_{RTE})]$$

$$L(\theta_{+AS+RTE}; D_{train})$$
$$= \log \sum_{i=1}^{|D_{train}|} \sum_{j=1}^{|Q|} [P(a_{ij}^*|q_{ij}) + \eta L_{AS}(q_{ij}, D_i)]$$
$$- \lambda g(\theta_{+AS+RTE})$$

g是正则化参数，我们使用elastic net作为正则函数

Proposed Method



Proposed Method

Preprocess and External Knowledge

我们将MCTest中文档进行指代消解，将文中的代词用命名实体代替。

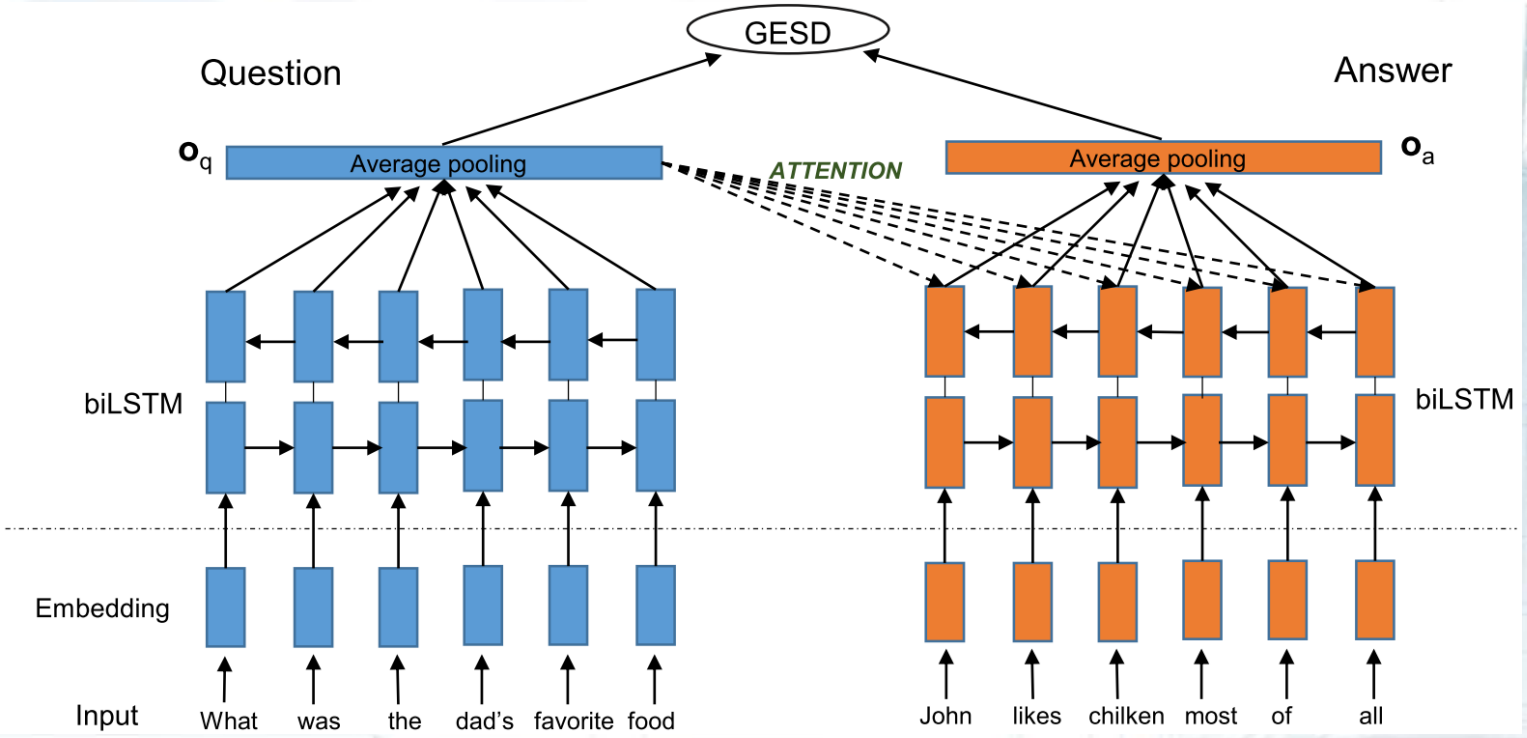
对于外部的答案选择资源（Answer selection）资源，我们使用WikiQA数据集，这个数据集基于维基百科，拥有将近2000个问题和3万个候选答案句子。
我们将wikiQA中所有的命名实体全部替换成其命名实体属性：

Barack Obama-> PERSON
The russia-> ORGANIZATION

对于外部的文本蕴涵推理（RTE）资源，我们选用斯坦福大学最新发布的超大规模语言推理数据集（SNLI）其中的训练集有50万对句子.

Proposed Method

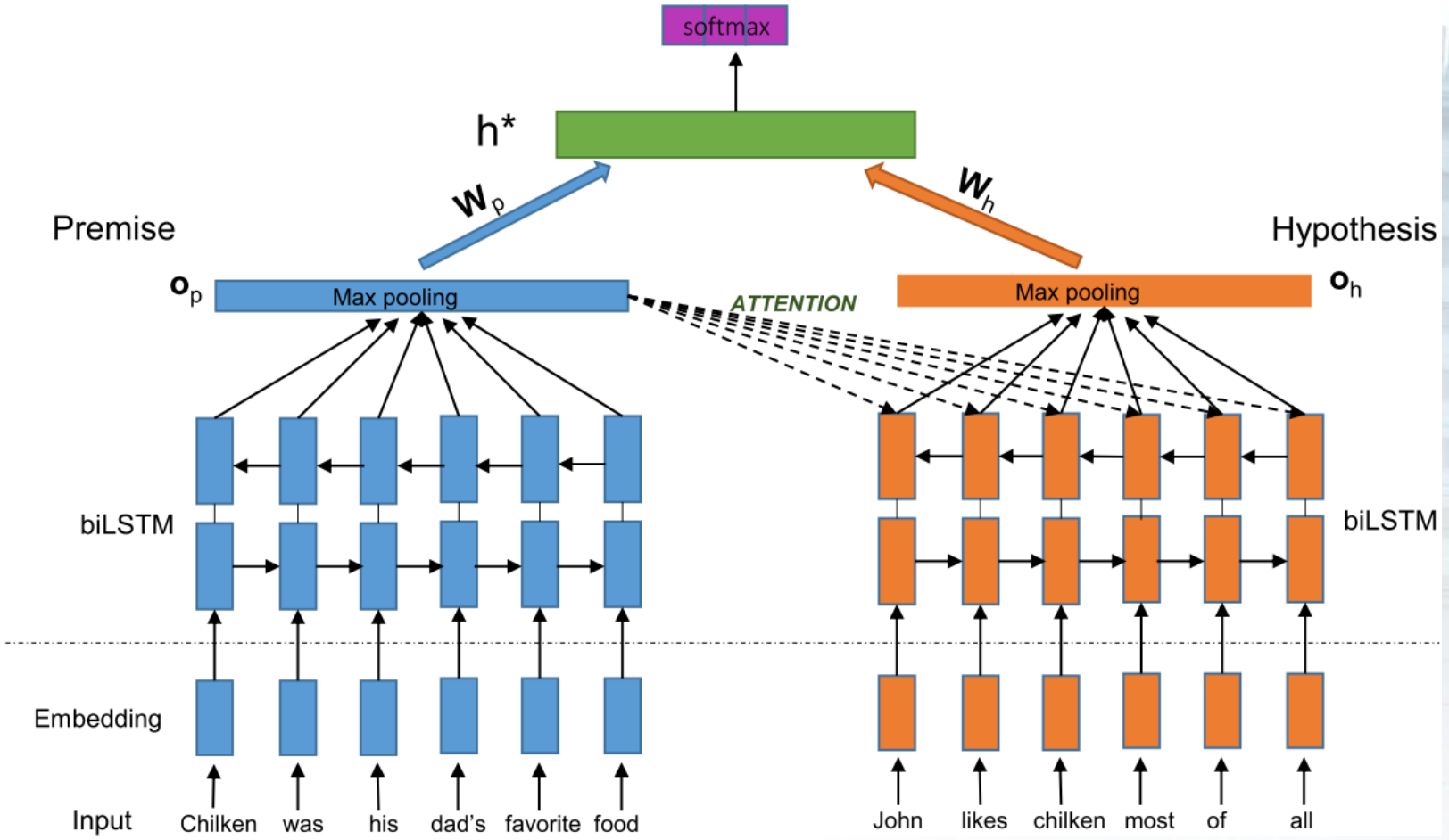
External Answer Selection model



$$GESD(x,y) = \frac{1}{1+||x-y||} \cdot \frac{1}{1+exp(-\gamma(xy^T+c))}$$

Proposed Method

External RTE model



External Model Result

	Answer Selection		RTE
	MAP	MRR	Accuracy
State of the Art	0.6921	**0.7108**	**0.835**
Our method	**0.6936**	0.7094	0.829

Yin W, Schütze H, Xiang B, et al. ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs[J]. TACL . 2016.

Rocktäschel T, Grefenstette E, Hermann K M, et al. Reasoning about Entailment with Neural Attention[C]. ICLR . 2016.

Experiment Result

System	MC160			MC500		
	One	Multiple	All	One	Multiple	All
Sliding Window	64.73	56.64	60.41	58.21	56.17	57.09
Sliding Window+Word Distance	76.78	62.50	67.50	64.00	57.46	60.43
Sliding Window+Word Distance+RTE	76.78	62.50	69.16	68.01	59.45	63.33
[Kapashi and Shah, 2015]	-	-	36.0	-	-	34.2
[Narasimhan and Barzilay, 2015]	82.36	65.23	73.23	68.38	59.90	63.75
[Wang and McAllester, 2015]	84.22	67.85	75.27	72.05	67.94	69.94
[Smith *et al.*, 2015]	78.79	**70.31**	75.77	69.12	63.34	65.43
[Sachan *et al.*, 2015]	-	-	-	67.65	**67.99**	67.83
without External Knowledge($\beta = 1, \eta = 0$)	40.39	37.94	39.08	38.40	33.13	31.33
without External AS knowledge($\eta = 0$)	41.07	40.63	40.83	49.63	28.05	32.83
without External RTE knowledge($\beta = 1$)	74.11	64.06	68.75	57.72	50.91	53.00
Final Model	**88.39**	64.84	**75.83**	**79.04**	63.51	**70.96**

Thank you!

research@bingning.wang