



Consensus Attention-based Neural Networks for Reading Comprehension

YIMING CUI, TING LIU, ZHIPENG CHEN, SHIJIN WANG AND GUOPING Hu

JOINT LABORATORY OF HIT AND iFLYTEK RESEARCH(HFL), CHINA
2016-12-15 OSAKA, JAPAN

OUTLINE

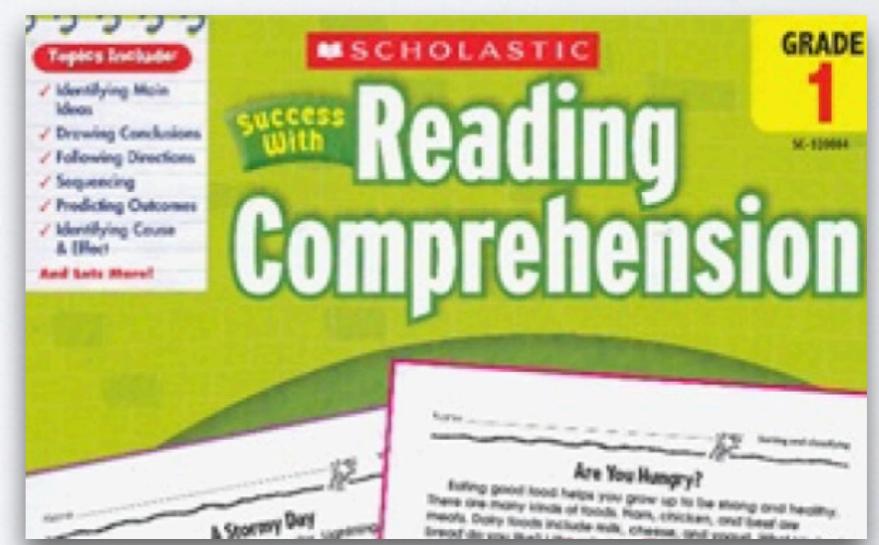
- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

OUTLINE

- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

INTRODUCTION

- Definition of RC
- Macro-view
 - To learn and do reasoning over world knowledge
- Micro-view
 - Read an article, and answer the questions based on it



INTRODUCTION

- Key points in RC
- → Document
- Query
- Candidates
- Answer

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

- 1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

*Example is chosen from the MCTest dataset ()

INTRODUCTION

- Key points in RC
 - Document
 - →Query
- Candidates
- Answer

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

- 1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

*Example is chosen from the MCTest dataset ()

INTRODUCTION

- Key points in RC
 - Document
 - Query
- → Candidates
- Answer

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

- 1) What is the name of the trouble making turtle?
- A) Fries
 - B) Pudding
 - C) James
 - D) Jane

*Example is chosen from the MCTest dataset ()

INTRODUCTION

- Key points in RC
 - Document
 - Query
 - Candidates
 - →Answer

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

- 1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

*Example is chosen from the MCTest dataset ()

INTRODUCTION

- A main obstacle in the research on RC
 - NO MUCH DATA !
- The related works are often started from providing the relevant corpus, and then proposing some technical insights in solving them
- Recently, *Cloze-style Reading Comprehension* has become enormously popular in the community

INTRODUCTION

- Why cloze-style reading comprehension?
 - Representative (as we all have done these things during our youth) and relatively easy (the answer is a single word) to start with
 - Explore the general relationship between the document and query
 - The data is relatively easy to collect

INTRODUCTION

- Cloze-style RC comprises of
 - Document: the same as the general RC
 - Query: a sentence with a blank
 - Candidate (optional): several candidates to fill in
 - Answer: a single word that exactly match the query
(the answer word should appear in the document)

OUTLINE

- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

RELATED WORKS

- CNN & Daily Mail (Hermann et al., 2015)

Original Version	Anonymised Version
Context <p>The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.” ...</p>	<p>the <i>ent381</i> producer allegedly struck by <i>ent212</i> will not press charges against the “<i>ent153</i>” host , his lawyer said friday . <i>ent212</i> , who hosted one of the most - watched television shows in the world , was dropped by the <i>ent381</i> wednesday after an internal investigation by the <i>ent180</i> broadcaster found he had subjected producer <i>ent193</i> “ to an unprovoked physical and verbal attack . ” ...</p>
Query <p>Producer X will not press charges against Jeremy Clarkson, his lawyer says.</p>	<p>producer X will not press charges against <i>ent212</i> , his lawyer says .</p>
Answer <p>Oisin Tymon</p>	<p><i>ent193</i></p>

RELATED WORKS

- Children's book test (Hill et al., 2015)

Step1: Choose 21 sentences

"Well, Miss Maxwell, I think it only fair to tell you that you may have trouble with those boys when they do come. Forewarned is forearmed, you know. Mr. Cropper was opposed to our hiring you. Not, of course, that he had any personal objection to you, but he is set against female teachers, and when a Cropper is set there is nothing on earth can change him. He says female teachers can't keep order. He 's started in with a spite at you on general principles, and the boys know it. They know he'll back them up in secret, no matter what they do, just to prove his opinions. Cropper is sly and slippery, and it is hard to corner him."

"Are the boys big ?" queried Esther anxiously.

"Yes. Thirteen and fourteen and big for their age. You can't whip 'em -- that is the trouble. A man might, but they'd twist you around their fingers. You'll have your hands full, I'm afraid. But maybe they'll behave all right after all."

Mr. Baxter privately had no hope that they would, but Esther hoped for the best. She could not believe that Mr. Cropper would carry his prejudices into a personal application. This conviction was strengthened when he overtook her walking from school the next day and drove her home. He was a big, handsome man with a very suave, polite manner. He asked interestedly about her school and her work, hoped she was getting on well, and said he had some rascals of his own to send soon. Esther thought that Mr. Baxter had exaggerated matters a little.

Step3: Choose 21st sentence as Query

S: 1 Mr. Cropper was opposed to our hiring you .
2 Not , of course , that he had any personal objection to you , but he is set against female teachers , and when a Cropper is set there is nothing on earth can change him .
3 He says female teachers ca n't keep order .
4 He 's started in with a spite at you on general principles , and the boys know it .
5 They know he 'll back them up in secret , no matter what they do , just to prove his opinions .
6 Cropper is sly and slippery , and it is hard to corner him . ..
7 `` Are the boys big ? ''
8 queried Esther anxiously .
9 `` Yes .
10 Thirteen and fourteen and big for their age .
11 You ca n't whip 'em -- that is the trouble .
12 A man might , but they 'd twist you around their fingers .
13 You 'll have your hands full , I 'm afraid .
14 But maybe they 'll behave all right after all . ..
15 Mr. Baxter privately had no hope that they would , but Esther hoped for the best .
16 She could not believe that Mr. Cropper would carry his prejudices into a personal application .
17 This conviction was strengthened when he overtook her walking from school the next day and drove her home .
18 He was a big, handsome man with a very suave, polite manner .
19 He asked interestedly about her school and her work .
20 Esther felt relieved

Step3:With a BLANK

Q: She thought that Mr. _____ had exaggerated matters a little .

C: Baxter, Cropper, Esther, course, fingers, manner, objection, opinion, right, spite.

a: Baxter

Step2: Choose first 20 sentences as Context

Step4: Choose other 9 similar words from Context as Candidate

Step3:The word removed from Query

OUTLINE

- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

PD & CFT

- A Chinese Reading Comprehension dataset: People Daily and Children's Fairy Tale (PD&CFT)
- Features
 - First Chinese cloze-style RC datasets, which add language diversity in this task
 - Along with the traditional news datasets (People Daily), we also provide a out-of-domain dataset (Children's Fairy Tale)

PD & CFT

- People Daily
 - Web-crawled news data, about 60k documents
- Children's Fairy Tale
 - Web-crawled children's reading material, about 1K documents
 - Contains virtualized characters, which is unable to use the common knowledge learned by large-scale data
 - Auto-set: automatically generated; Human-set: manually selected, those questions that depend on LM or cooccurrence is removed

PD & CFT

- Statistics of PD&CFT
 - Note that, the CFT dataset is only served as the out-of-domain test sets.

	People Daily			Children's Fairy Tale	
	Train	Valid	Test	Test-auto	Test-human
# Query	870,710	3,000	3,000	1,646	1,953
Max # tokens in docs	618	536	634	318	414
Max # tokens in query	502	153	265	83	92
Avg # tokens in docs	379	425	410	122	153
Avg # tokens in query	38	38	41	20	20
Vocabulary	248,160			N/A	

PD & CFT

- Example

Document	<p>1 人民网 1月 1日 讯 据《纽约时报》报道，美国华尔街股市在 2013年的最后一天继续上涨，和全球股市一样，都以最高纪录或接近最高纪录结束本年的交易。</p> <p>2 《纽约时报》报道说，标普 500 指数今年上升 29.6%，为 1997年以来的最大涨幅；</p> <p>3 道琼斯工业平均指数上升 26.5%，为 1996年以来的最大涨幅；</p> <p>4 纳斯达克 上涨 38.3%。</p> <p>5 就 12月 31日 来说，由于就业前景看好和经济 增长明年可能加速，消费者信心上升。</p> <p>6 工商协进会 报告，12月 消费者信心上升到 78.1，明显高于 11月的 72。</p> <p>7 另据《华尔街日报》报道，2013年是 1995年以来美国股市表现最好的一年。</p> <p>8 这一年里，投资美国股市的明智做法是追着“傻钱”跑。</p> <p>9 所谓的“傻钱”XXXXX，其实就是买入并持有美国股票这样的普通组合。</p> <p>10 这个策略要比对冲基金和其它专业投资者使用的更为复杂的投资方法效果好得多。</p>	1 People Daily (Jan 1). According to report of "New York Times", the Wall Street stock market continued to rise as the global stock market in the last day of 2013, ending with the highest record or near record of this year. 2 "New York times" reported that the S&P 500 index rose 29.6% this year, which is the largest increase since 1997. 3 Dow Jones industrial average index rose 26.5%, which is the largest increase since 1996. 4 NASDAQ rose 38.3%. 5 In terms of December 31, due to the prospects in employment and possible acceleration of economy next year, there is a rising confidence in consumers. 6 As reported by Business Association report, consumer confidence rose to 78.1 in December, significantly higher than 72 in November. 7 Also as "Wall Street journal" reported that 2013 is the best U.S. stock market since 1995. 8 In this year, to chase the "silly money" is the most wise way to invest in U.S. stock. 9 The so-called "silly money" XXXXX is that, to buy and hold the common combination of U.S. stock. 10 This strategy is better than other complex investment methods, such as hedge funds and the methods adopted by other professional investors.
Query	所谓的“傻钱”XXXXX，其实就是买入并持有美国股票这样的普通组合。	The so-called "silly money" XXXXX is that, to buy and hold the common combination of U.S. stock.
Answer	策略	strategy

PD & CFT

- Step I: select one sentence in the (truncated) document

1 ||| People Daily (Jan 1). According to report of "New York Times", the Wall Street stock market continued to rise as the global stock market in the last day of 2013, ending with the highest record or near record of this year.

2 ||| "New York times" reported that the S&P 500 index rose 29.6% this year, which is the largest increase since 1997.

3 ||| Dow Jones industrial average index rose 26.5%, which is the largest increase since 1996.

4 ||| NASDAQ rose 38.3%.

5 ||| In terms of December 31, due to the prospects in employment and possible acceleration of economy next year, there is a rising confidence in consumers.

6 ||| As reported by Business Association report, consumer confidence rose to 78.1 in December, significantly higher than 72 in November.

7 ||| Also as "Wall Street journal" reported that 2013 is the best U.S. stock market since 1995.

8 ||| In this year, to chase the "silly money" is the most wise way to invest in U.S. stock.

9 ||| The so-called "silly money" strategy is that, to buy and hold the common combination of U.S. stock.

10 ||| This strategy is better than other complex investment methods, such as hedge funds and the methods adopted by other professional investors.

PD & CFT

- Step2: choose one word in this sentence
 - Only named entity and common noun is considered

1 ||| People Daily (Jan 1). According to report of “New York Times”, the Wall Street stock market continued to rise as the global stock market in the last day of 2013, ending with the highest record or near record of this year.

2 ||| “New York times” reported that the S&P 500 index rose 29.6% this year, which is the largest increase since 1997.

3 ||| Dow Jones industrial average index rose 26.5%, which is the largest increase since 1996.

4 ||| NASDAQ rose 38.3%.

5 ||| In terms of December 31, due to the prospects in employment and possible acceleration of economy next year, there is a rising confidence in consumers.

6 ||| As reported by Business Association report, consumer confidence rose to 78.1 in December, significantly higher than 72 in November.

7 ||| Also as “Wall Street journal” reported that 2013 is the best U.S. stock market since 1995.

8 ||| In this year, to chase the “silly money” is the most wise way to invest in U.S. stock.

9 ||| The so-called “silly money” strategy is that, to buy and hold the common combination of U.S. stock.

10 ||| This strategy is better than other complex investment methods, such as hedge funds and the methods adopted by other professional investors.

PD & CFT

- Step3: Leave out that word, and the sentence will become the query

1 ||| People Daily (Jan 1). According to report of “New York Times”, the Wall Street stock market continued to rise as the global stock market in the last day of 2013, ending with the highest record or near record of this year.

2 ||| “New York times” reported that the S&P 500 index rose 29.6% this year, which is the largest increase since 1997.

3 ||| Dow Jones industrial average index rose 26.5%, which is the largest increase since 1996.

4 ||| NASDAQ rose 38.3%.

5 ||| In terms of December 31, due to the prospects in employment and possible acceleration of economy next year, there is a rising confidence in consumers.

6 ||| As reported by Business Association report, consumer confidence rose to 78.1 in December, significantly higher than 72 in November.

7 ||| Also as “Wall Street journal” reported that 2013 is the best U.S. stock market since 1995.

8 ||| In this year, to chase the “silly money” is the most wise way to invest in U.S. stock.

9 ||| The so-called “silly money” XXXXX is that, to buy and hold the common combination of U.S. stock.

10 ||| This strategy is better than other complex investment methods, such as hedge funds and the methods adopted by other professional investors.

Document

Query

The so-called “silly money” XXXXX is that, to buy and hold the common combination of U.S. stock.

PD & CFT

- Step4: the removed word becomes the answer to the query

1 ||| People Daily (Jan 1). According to report of “New York Times”, the Wall Street stock market continued to rise as the global stock market in the last day of 2013, ending with the highest record or near record of this year.

2 ||| “New York times” reported that the S&P 500 index rose 29.6% this year, which is the largest increase since 1997.

3 ||| Dow Jones industrial average index rose 26.5%, which is the largest increase since 1996.

4 ||| NASDAQ rose 38.3%.

5 ||| In terms of December 31, due to the prospects in employment and possible acceleration of economy next year, there is a rising confidence in consumers.

6 ||| As reported by Business Association report, consumer confidence rose to 78.1 in December, significantly higher than 72 in November.

7 ||| Also as “Wall Street journal” reported that 2013 is the best U.S. stock market since 1995.

8 ||| In this year, to chase the “silly money” is the most wise way to invest in U.S. stock.

9 ||| The so-called “silly money” XXXXX is that, to buy and hold the common combination of U.S. stock.

10 ||| This strategy is better than other complex investment methods, such as hedge funds and the methods adopted by other professional investors.

Document

Query

Answer

strategy

PD & CFT

- Comparison of three Cloze-style RC datasets

Dataset	Language	Genre	Blank Type	Doc	Query
CNN/DM	English	News	NE	News Article	Summary w/ a blank
CBTest	English	Story	NE,CN,VP	20 consecutive sentences	21th sentence w/ a blank
PD&CFT	Chinese	News, story	NE,CN	Doc w/ a blank	the sentence that blank belongs to

OUTLINE

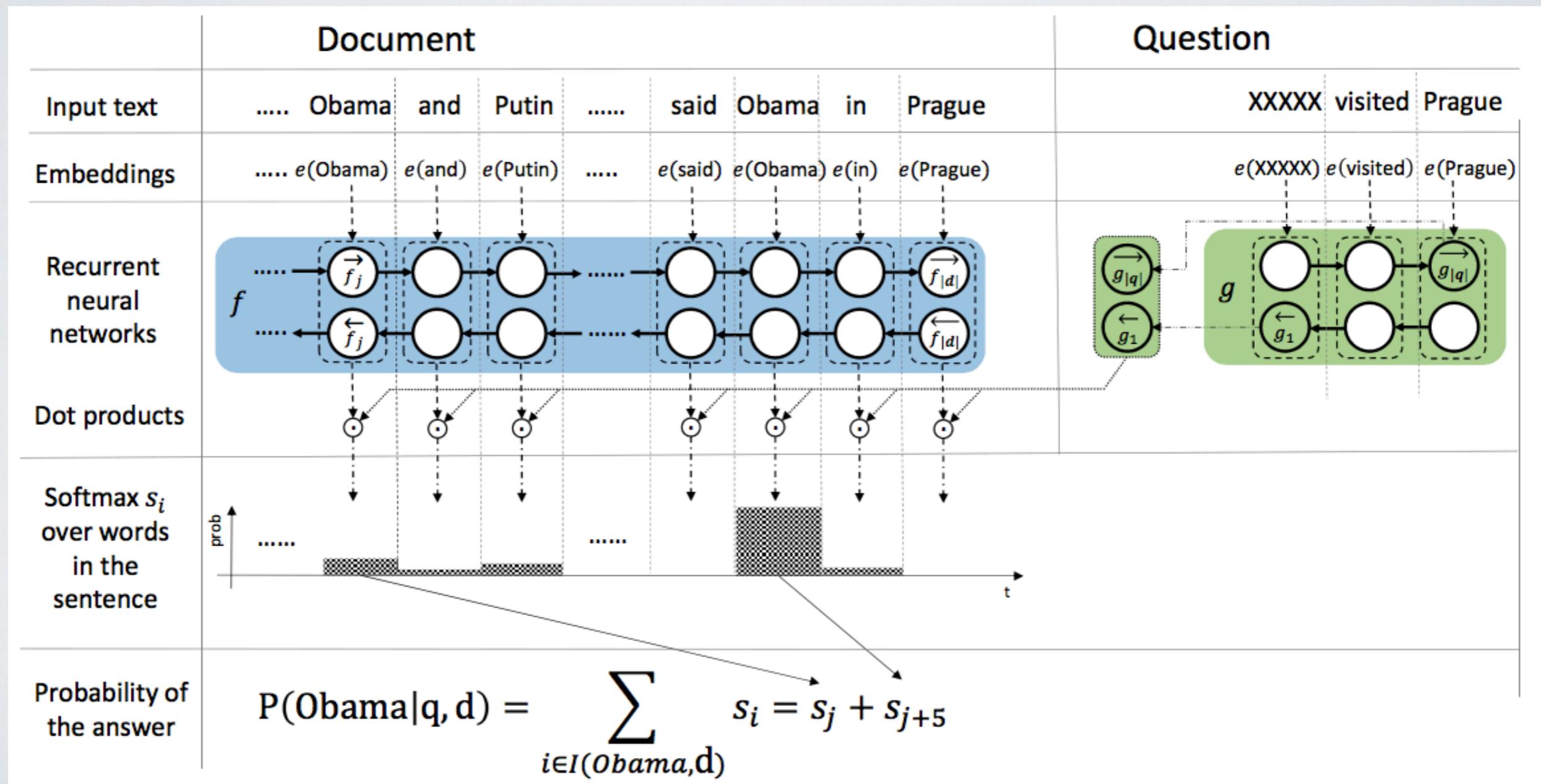
- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

CAS READER

- We propose an extension to the AS Reader (Kadlec et al., 2016), which is a popular framework on close-style reading comprehension task
- Modification
 - Instead of blending query representations into one, we can take EVERY individual query words to generate a document-level attention respectively

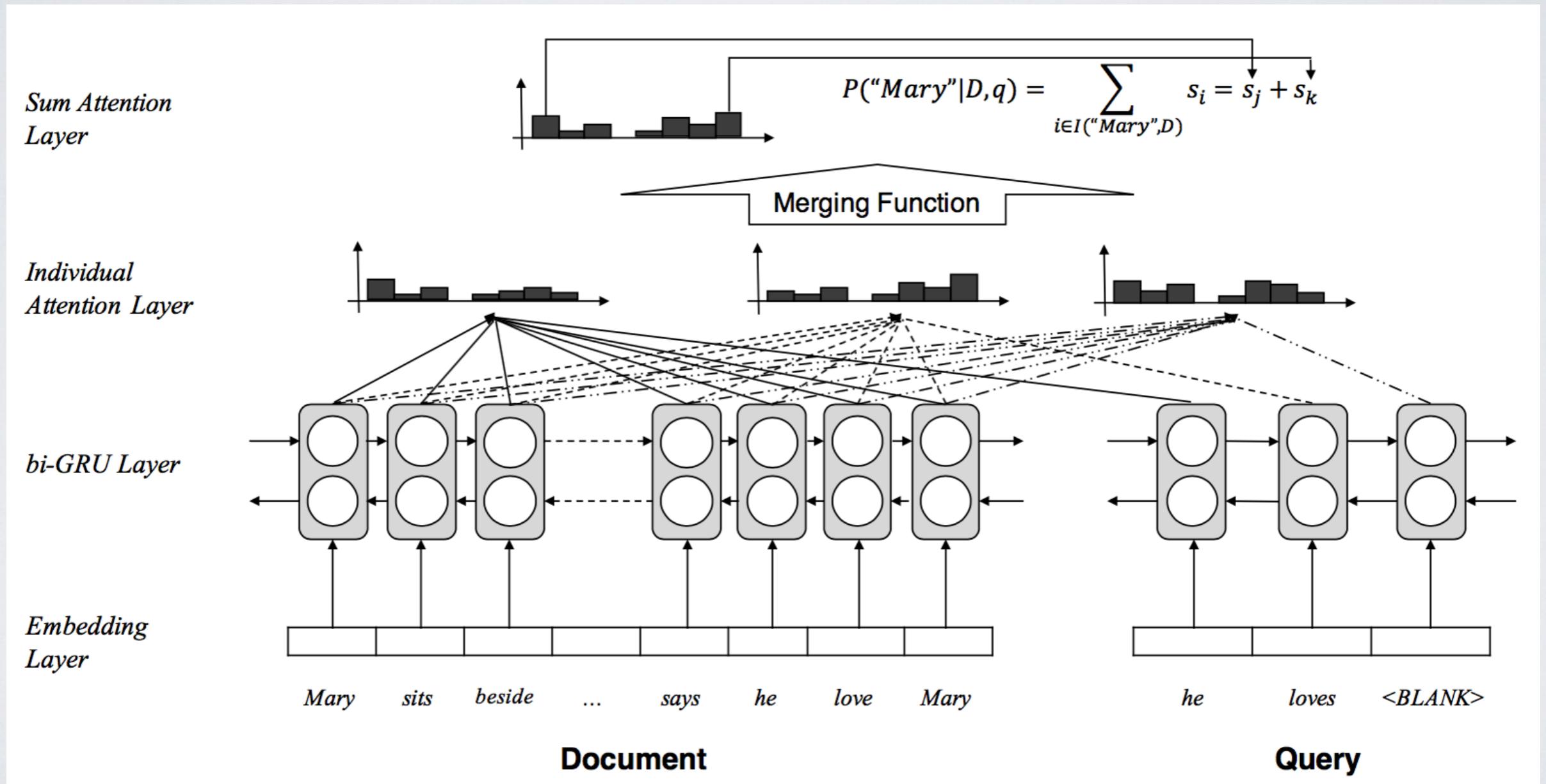
CAS READER

- AS Reader (Kadlec et al., 2016)



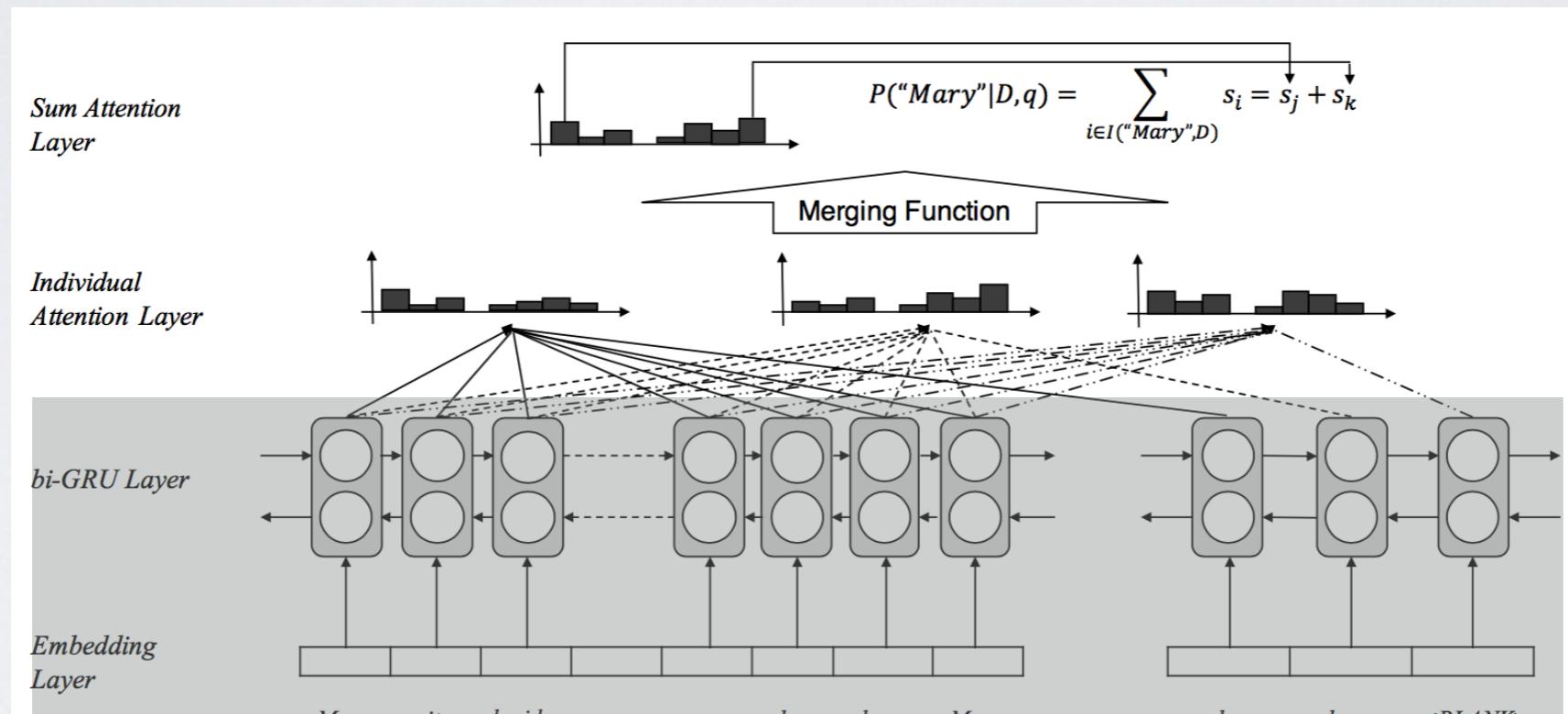
CAS READER

- Neural Architecture



CAS READER

- Step I: Transform document and query into contextual representations using GRU



Document

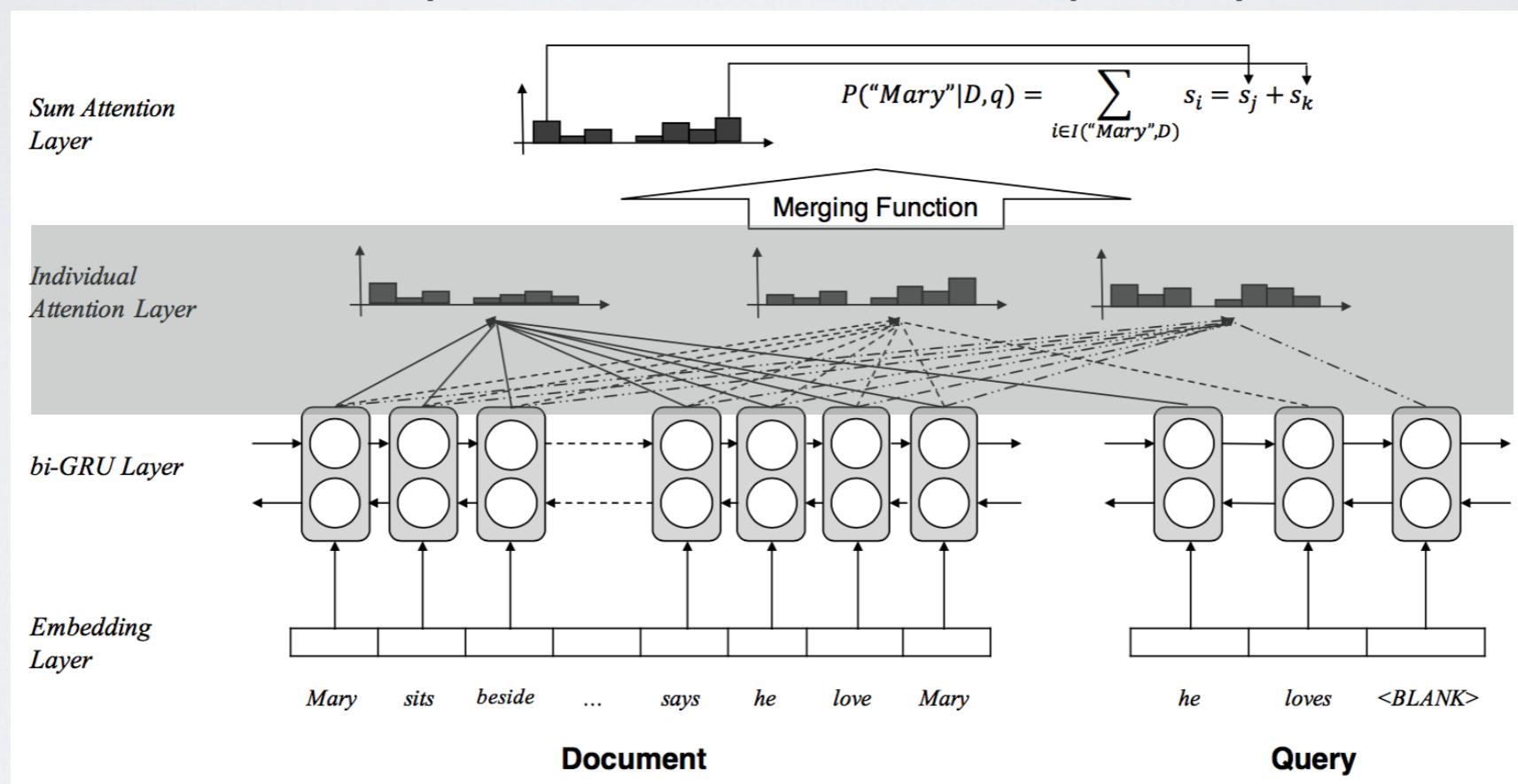
$$e(x) = W_e * x, \text{ where } x \in \mathcal{D}, \mathcal{Q} \quad (1)$$

$$\overrightarrow{h_s(x)} = \overrightarrow{GRU}(e(x)) ; \overleftarrow{h_s(x)} = \overleftarrow{GRU}(e(x)) \quad (2)$$

$$h_s(x) = [\overrightarrow{h_s(x)}; \overleftarrow{h_s(x)}] \quad (3)$$

CAS READER

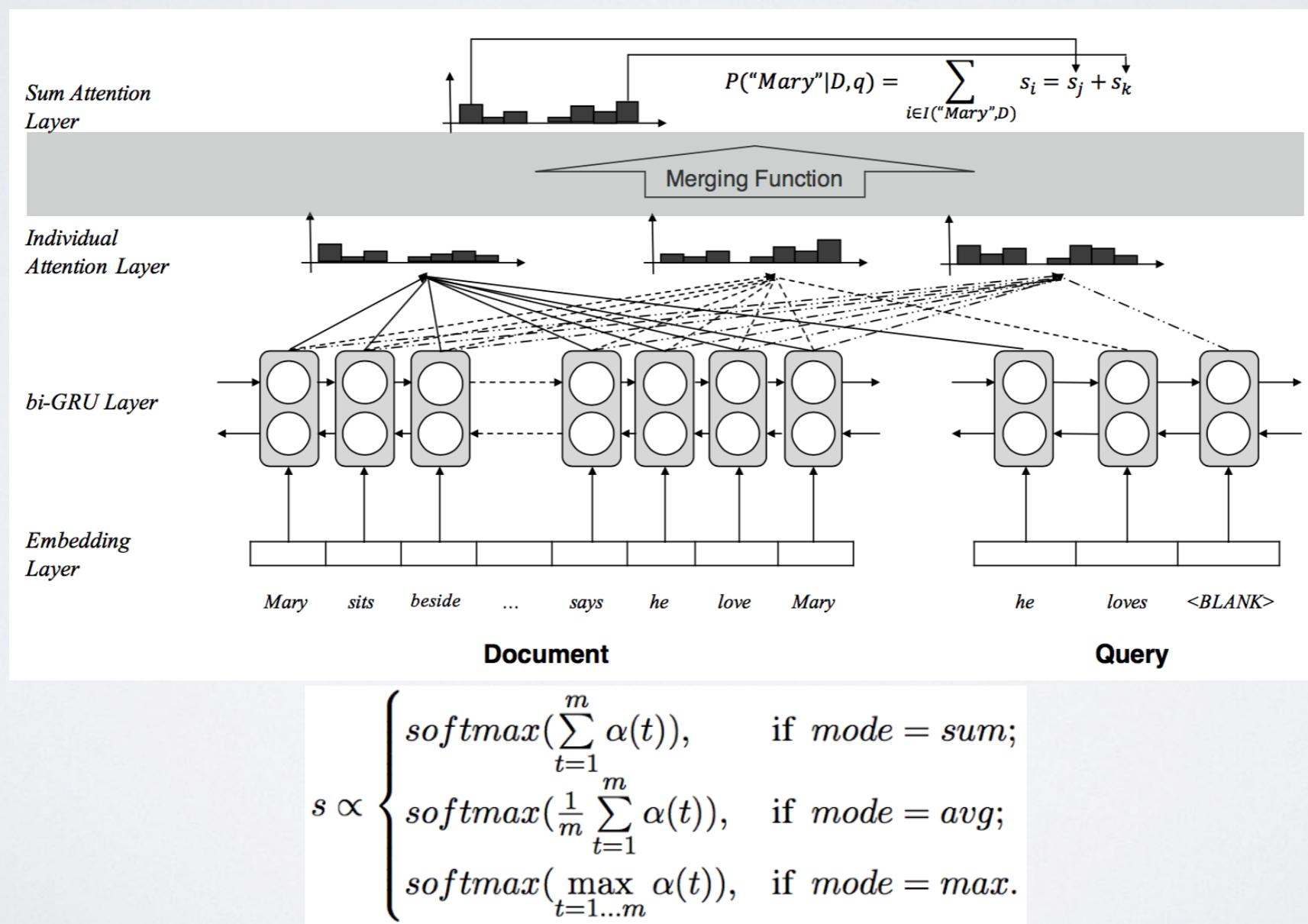
- Step2: Generate several document-level attentions in terms of every word in the query



$$\alpha(t) = \text{softmax}(h_{\text{doc}} \odot h_{\text{query}}(t))$$

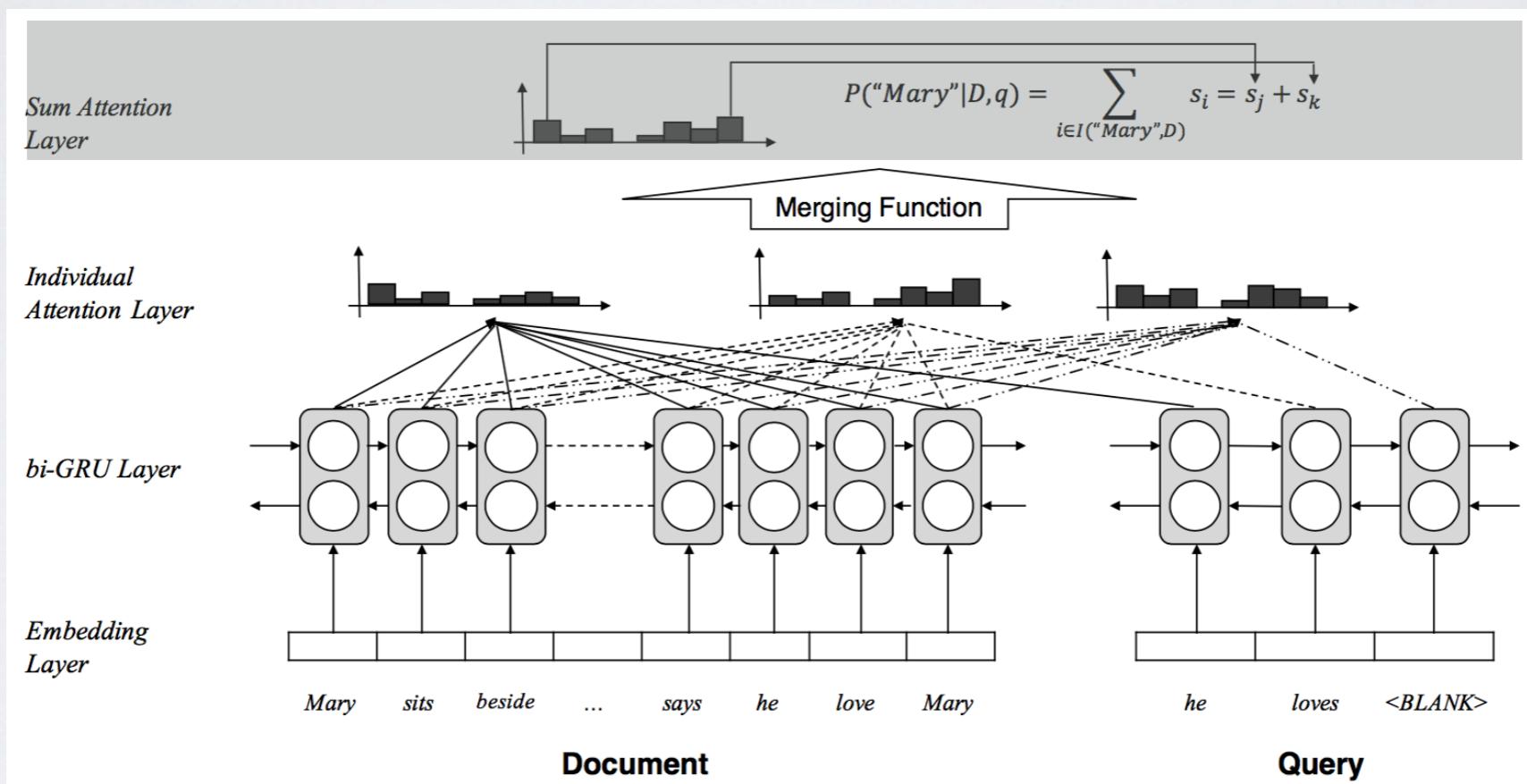
CAS READER

- Step3: Induce a consensus attention over these individual attentions with heuristic functions



CAS READER

- Step4: Applying sum-attention mechanism (Kadlec et al., 2016) to get the final probability of the answer



$$P(w | \mathcal{D}, \mathcal{Q}) = \sum_{i \in I(w, \mathcal{D})} s_i, \quad w \in V$$

OUTLINE

- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

EXPERIMENTS

- Setups
 - Embedding Layer: randomly initialized with uniformed distribution $\sim [-0.1, 0.1]$
 - Hidden Layer: GRU with random orthogonal initialization (Saxe et al., 2013), and gradient clipping to 10 (Pascanu et al., 2013)
 - Vocabulary: set a shortlist of 100k for PD&CFT condition. No vocabulary truncation on CNN and CBT.
 - Optimization: Adam (Kingma and Ba, 2014) with initial LR=0.0005. Batch size is set to 32.

EXPERIMENTS

- Setups
 - Statistics of CNN & CBT NE/CN

	CNN News			CBT NE			CBT CN		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
# Query	380,298	3,924	3,198	108,719	2,000	2,500	120,769	2,000	2,500
Max # candidates	527	187	396	10	10	10	10	10	10
Avg # candidates	26	26	25	10	10	10	10	10	10
Avg # tokens	762	763	716	433	412	424	470	448	461
Vocabulary	118,497			53,063			53,185		

- Dimensions of neural units and Dropout rate (Srivastava et al., 2014)

	Embed. # units	Hidden # units	Dropout
CNN News	384	256	None
CBTest NE	384	384	None
CBTest CN	384	384	None
People Daily & CFT	256	256	0.1

- All models are trained on Tesla K40 GPU
- Implementation is done with Theano (Theano Developing Team, 2016) and Keras framework (Chollet, 2015)

EXPERIMENTS

- Results on PD&CFT

	People Valid	Daily Test	Children's Fairy Tale Test-auto	Children's Fairy Tale Test-human
AS Reader	64.1	67.2	40.9	33.1
CAS Reader (mode: avg)	65.2	68.1	41.3	35.0
CAS Reader (mode: sum)	64.7	66.8	43.0	34.7
CAS Reader (mode: max)	63.3	65.4	38.3	32.0

- Heuristic comparison: avg > sum >> max
- Dramatic drop in out-of-domain test sets

EXPERIMENTS

- Results on CNN and CBT

	CNN News		CBTest NE		CBTest CN	
	Valid	Test	Valid	Test	Valid	Test
Deep LSTM Reader [†]	55.0	57.0	-	-	-	-
Attentive Reader [†]	61.6	63.0	-	-	-	-
Impatient Reader [†]	61.8	63.8	-	-	-	-
Human (context+query) [‡]	-	-	-	81.6	-	81.6
LSTMs (context+query) [‡]	-	-	51.2	41.8	62.6	56.0
MemNN (window + self-sup.) [‡]	63.4	66.8	70.4	66.6	64.2	63.0
Stanford AR [§]	72.4	72.4	-	-	-	-
AS Reader [#]	68.6	69.5	73.8	68.6	68.8	63.4
CAS Reader (mode: avg)	68.2	70.0	74.2	69.2	68.2	65.7

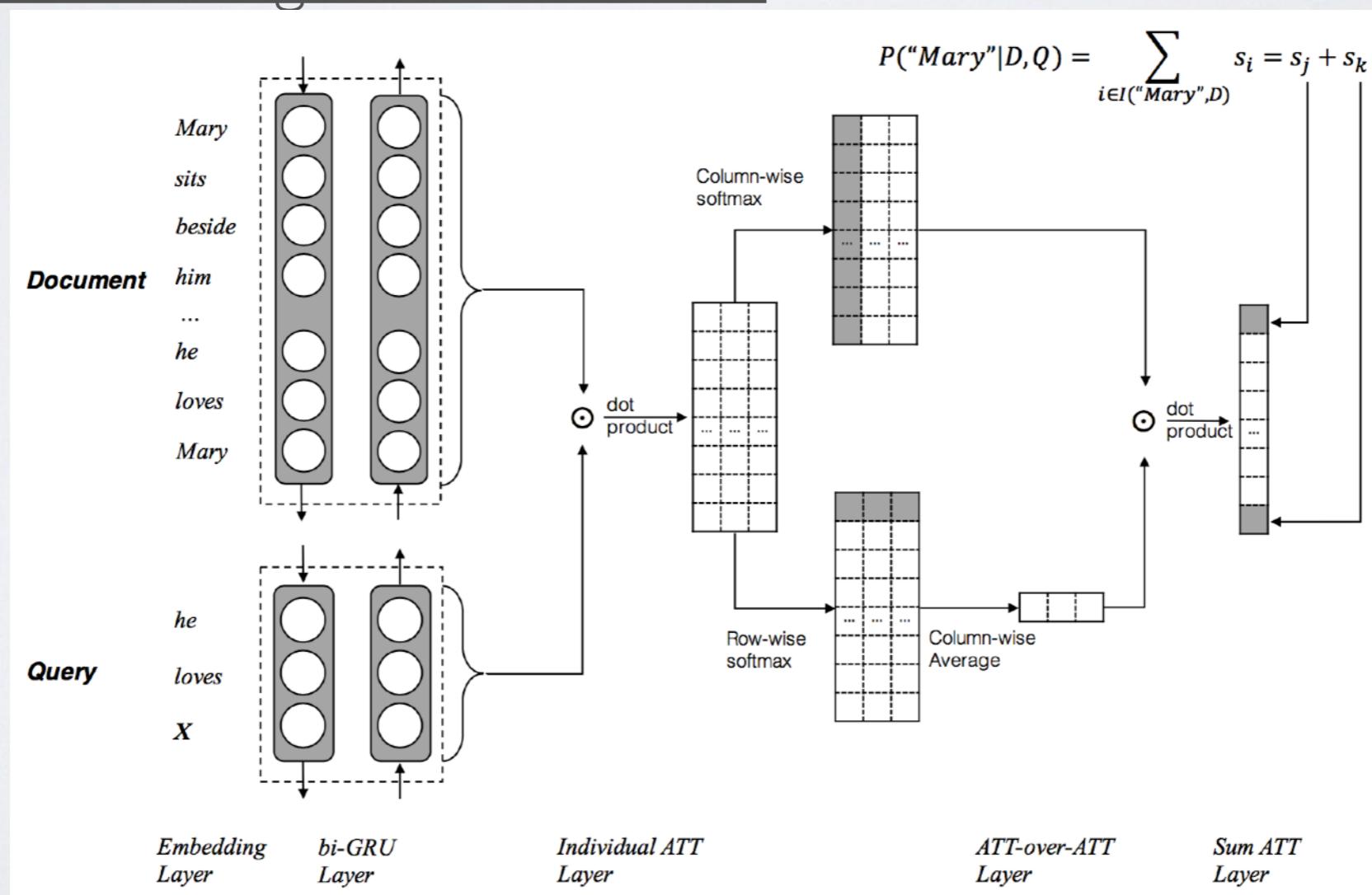
- Modest improvements over AS Reader

OUTLINE

- Introduction
- Existing Cloze-style Reading Comprehension Dataset
- Chinese Dataset: People Daily & Children's Fairy Tale (PD&CFT)
- Consensus Attention Sum Reader (CAS Reader)
- Experiments & Observations
- Further Reading & Conclusion

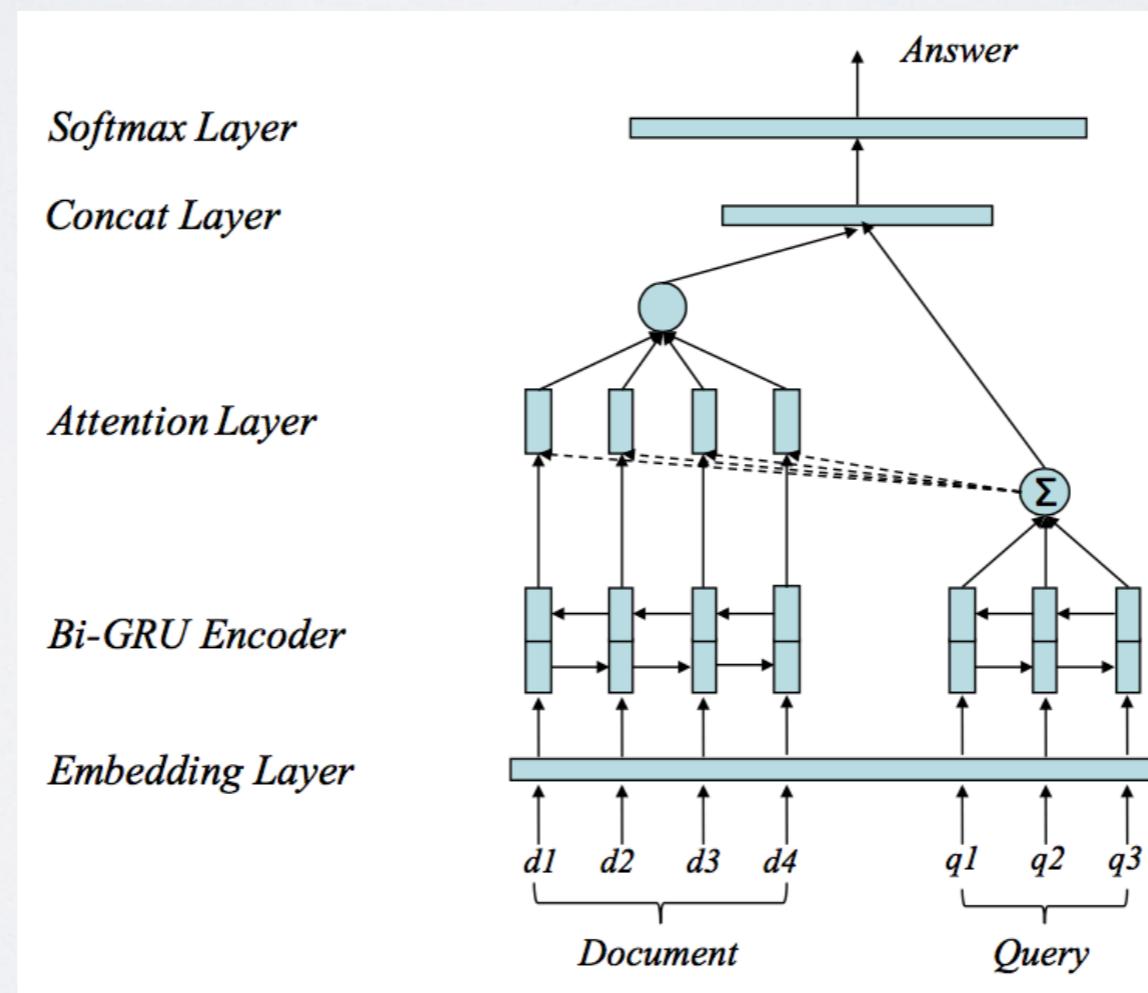
FURTHER READING

- Attention-over-Attention Neural Network for Reading Comprehension (Cui et al., 2016)
- Arxiv: <https://arxiv.org/abs/1607.04423>



FURTHER READING

- Generating and Exploiting Large-scale Pseudo Training Data for Zero Pronoun Resolution (Liu et al., 2016)
- arxiv: <https://arxiv.org/abs/1606.01603>



CONCLUSION

- PD & CFT: A Chinese Cloze-style RC dataset
 - the first Chinese RC dataset, aiming to enriching the diversity in RC task
 - Human-selected test set is much more harder than the one that is automatically generated, and brings much difficulties
- Consensus Attention-based Reader (CAS Reader)
 - By taking every word in the query, we can generate *consensus attention* via several doc-level attentions

RELATED LINKS

- PD & CFT datasets
 - <https://github.com/ymcui/Chinese-RC-Dataset>
- General training tips & Leaderboard of Cloze-style RC (updates irregularly)
 - <https://github.com/ymcui/Eval-on-NN-of-RC>
- Personal website (slides will be uploaded to this)
 - <http://ymcui.github.io>

REFERENCES

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv: 1409.0473.
- Wanxiang Che, Zhenghua Li, and Ting Liu. 2010. Ltp: A chinese language technology platform. In Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, pages 13–16. Association for Computational Linguistics.
- Danqi Chen, Jason Bolton, and Christopher D. Manning. 2016. A thorough examination of the cnn/daily mail reading comprehension task. In Association for Computational Linguistics (ACL).
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1724–1734. Association for Computational Linguistics.
- François Chollet. 2015. Keras. <https://github.com/fchollet/keras>.
- Karl Moritz Hermann, Tomas Kociský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In Advances in Neural Information Processing Systems, pages 1684–1692.
- Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. 2015. The goldilocks principle: Reading children’s books with explicit memory representations. arXiv preprint arXiv:1511.02301.
- Rudolf Kadlec, Martin Schmid, Ondřej Bajgar, and Jan Kleindienst. 2016. Text understanding with the attention sum reader network. arXiv preprint arXiv: 1603.01547.
- Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

REFERENCES

- Ting Liu, Yiming Cui, Qingyu Yin, Shijin Wang, Weinan Zhang, and Guoping Hu. 2016. Generating and exploiting large-scale pseudo training data for zero pronoun resolution. arXiv preprint arXiv:1606.01603.
- Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. ICML (3), 28:1310–1318.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543. Association for Computational Linguistics.
- Andrew M Saxe, James L McClelland, and Surya Ganguli. 2013. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv preprint arXiv:1312.6120.
- Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1):1929–1958.
- Wilson L Taylor. 1953. Cloze procedure: a new tool for measuring readability. Journalism and Mass Communication Quarterly, 30(4):415.
- Theano Development Team. 2016. Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688, May.
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In Advances in Neural Information Processing Systems, pages 2692–2700.

THANK YOU !