Progress and Future Prospects of Machine Reading Comprehension

Yiming Cui

Joint Laboratory of HIT and iFLYTEK Research (HFL)

September 2nd, 2017

OUTLINE

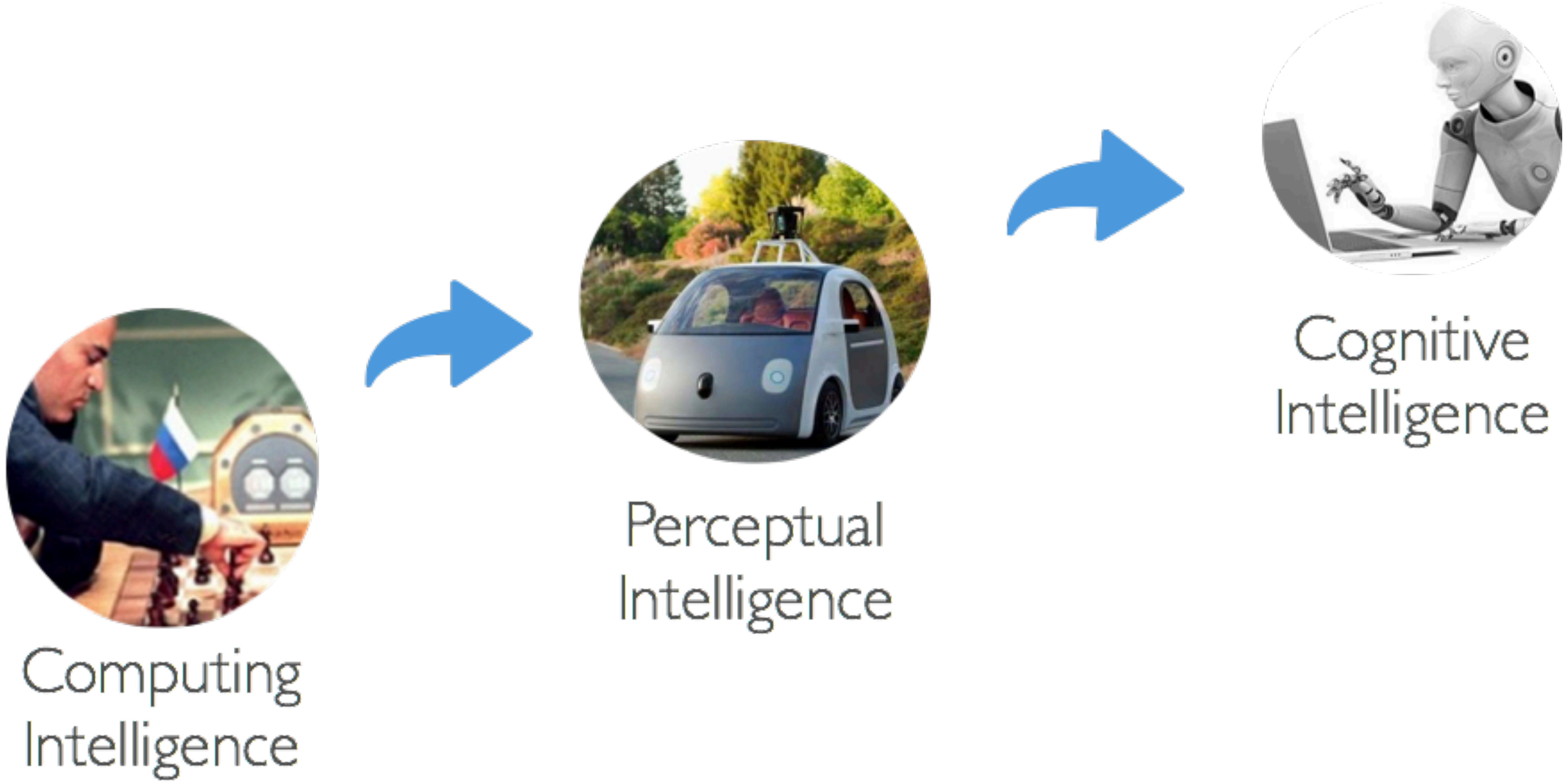
- Introduction to Machine Reading Comprehension (MRC)
- (Almost) Recent Progress of MRC
    - Cloze-style MRC
    - Complex MRC
- Future Prospects
- Conclusion

OUTLINE

- **Introduction to Machine Reading Comprehension (MRC)**
- (Almost) Recent Progress of MRC
  - Cloze-style MRC
  - Complex MRC
- Future Prospects
- Conclusion

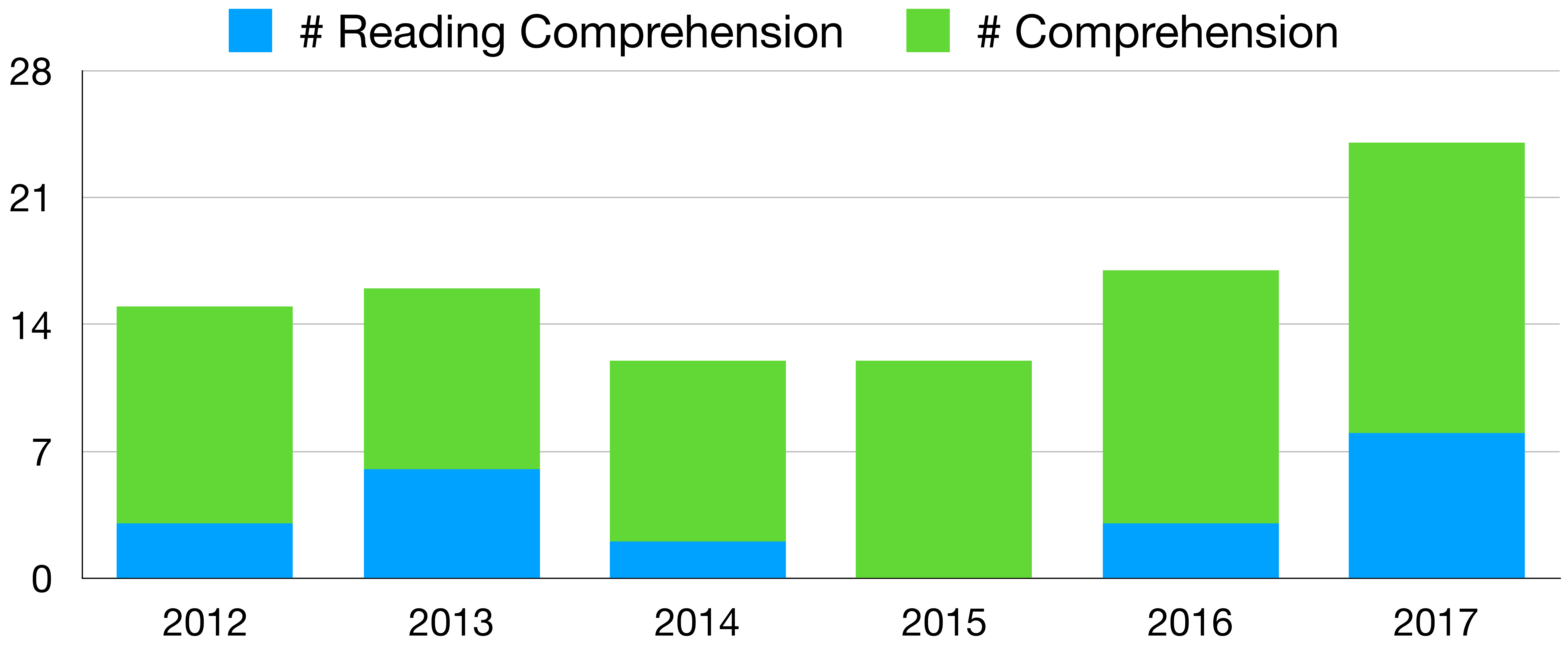
Introduction

- To comprehend human language is essential in AI
- As a typical task in cognitive intelligence, Machine Reading Comprehension (MRC) has attracted attentions from NLP field



Research Trends

- Research trends of MRC (from 2012)
  - Keywords: Reading Comprehension, Comprehension



*Statistics are obtained from ACL Anthology

RESEARCH TRENDS

- **Reading Comprehension Papers in ACL 2017 (may not an exhaustive list)**
- **Long Paper**
  - Du et al. Neural Question Generation for Reading Comprehension
  - Pengtao Xie and Eric Xing. A Constituent-Centric Neural Architecture for Reading Comprehension
  - Cui et al. Attention-over-Attention Neural Networks for Reading Comprehension
  - Wang et al. Gated Self-Matching Networks for Reading Comprehension and Question Answering
  - Dhingra et al. Gated-Attention Readers for Text Comprehension
  - Bishan Yang and Tome Mitchell. Leveraging Knowledge Bases in LSTMs for Improving Machine Reading
  - Chen et al. Reading Wikipedia to Answer Open-Domain Questions
  - Sugawara et al. Evaluation Metrics for Reading Comprehension: Prerequisite Skills and Readability
  - Joshi et al. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension
- **Short Papers**
  - Min et al. Question Answering through Transfer Learning from Large Fine-grained Supervision Data

RESEARCH TRENDS

- **Reading Comprehension Papers in EMNLP 2017 (may not an exhaustive list)**
- **Short Papers**
  - Xinya Du and Claire Cardie. Identifying Where to Focus in Reading Comprehension for Neural Question Generation
  - Noriega-Atala et al. Learning what to read: Focused machine reading
- **Long Paper**
  - Golub et al. Two-Stage Synthesis Networks for Transfer Learning in Machine Comprehension
  - Robin Jia and Percy Liang. Adversarial Examples for Evaluating Reading Comprehension Systems
  - Lai et al. RACE: Large-scale ReAding Comprehension Dataset From Examinations
  - Lin et al. Reasoning with Heterogeneous Knowledge for Commonsense Machine Comprehension
  - Liu et al. Structural Embedding of Syntactic Trees for Machine Comprehension
  - Long et al. World Knowledge for Reading Comprehension: Rare Entity Prediction with Hierarchical LSTMs Using External Descriptions
  - Yin et al. Document-Level Multi-Aspect Sentiment Classification as Machine Comprehension

INTRODUCTION

- Reading Comprehension Definition
- Macro-view
  - To learn and do reasoning with world knowledge and common knowledge while growing up
- Micro-view
  - Read an article/several articles, and answer the questions based on it

INTRODUCTION

- Key components 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 (Richardson et al., 2013)

INTRODUCTION

- Key components 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 (Richardson et al., 2013)

INTRODUCTION

- Key components 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 (Richardson et al., 2013)

INTRODUCTION

- Key components in RC
  - Document
  - Query
  - Candidates
  - **→Answer**

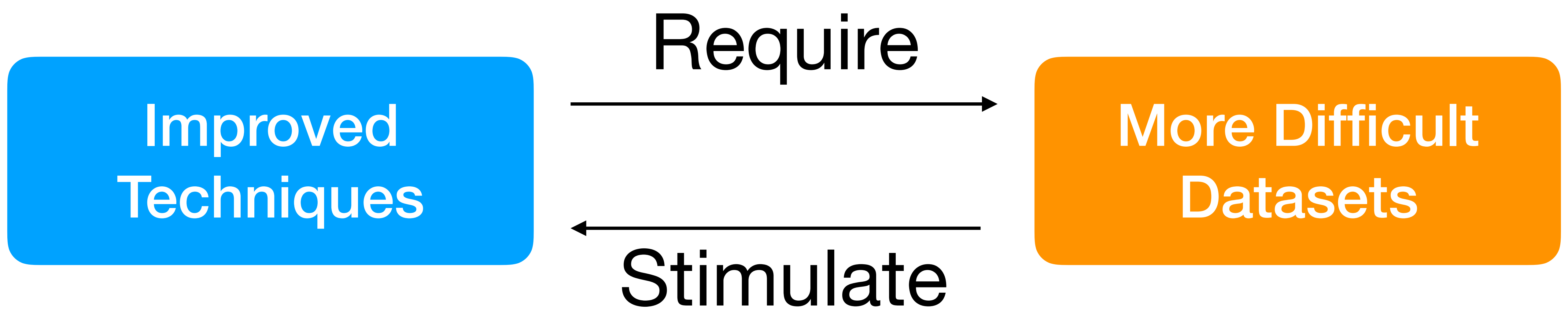
James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.
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 (Richardson et al., 2013)

INTRODUCTION

- Why MRC has become enormously popular in recent year?
- Mutual effect by
  - Growing interest in DL techniques
  - Availability of large-scale RC data

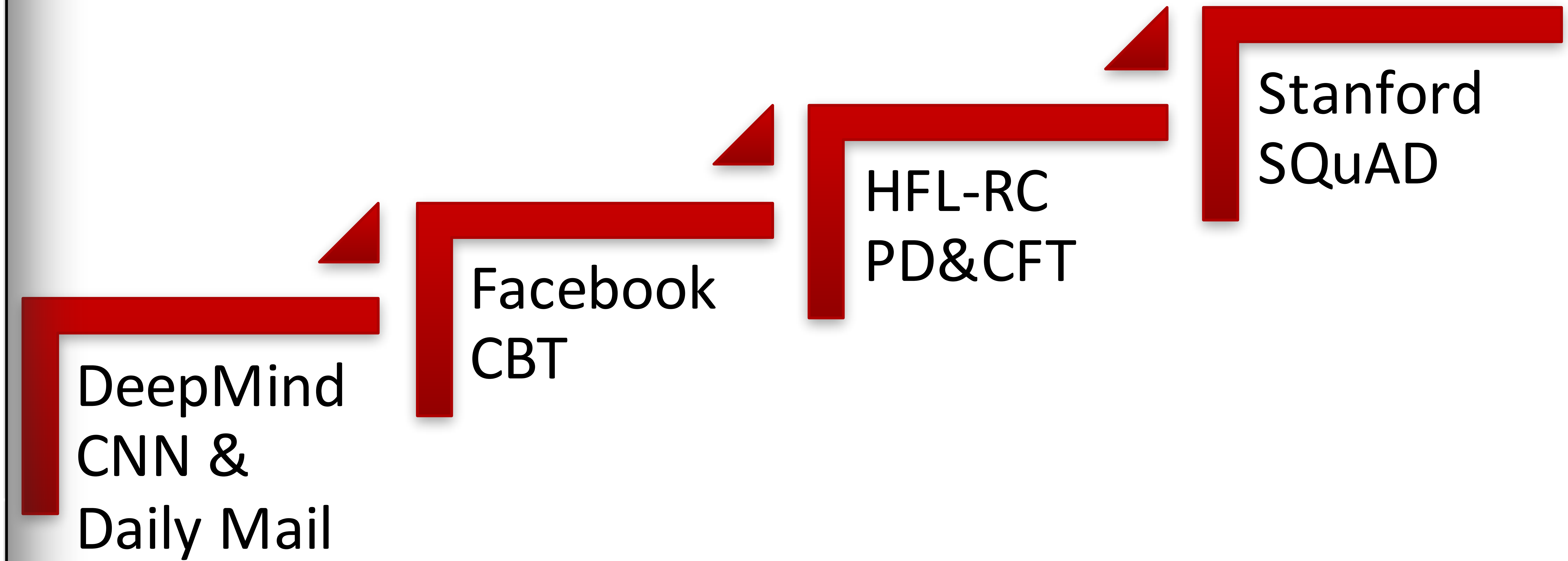


Progress of RC Datasets

- MCTest (Richardson et al., EMNLP2013)

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



PROGRESS OF RC DATASETS

- DeepMind CNN/DailyMail (Hermann et al., NIPS2015)

	Original Version	Anonymised Version
Context	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." …	the *ent381* producer allegedly struck by *ent212* will not press charges against the " *ent153* " host , his lawyer said friday . *ent212* , who hosted one of the most - watched television shows in the world , was dropped by the *ent381* wednesday after an internal investigation by the *ent180* broadcaster found he had subjected producer *ent193* " to an unprovoked physical and verbal attack . " …
Query	Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.	producer **X** will not press charges against *ent212* , his lawyer says .
Answer	Oisin Tymon	*ent193*

Stanford
SQuAD

Progress of RC Datasets

- Facebook CBT (Hill et al., ICLR2016)

"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 two young rascals of his own to send soon. Esther felt relieved. She thought that Mr. Baxter had exaggerated matters a little.

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 , hoped she was getting on well , and said he had two young rascals of his own to send soon .
20 Esther felt relieved .

q: She thought that Mr. ________ had exaggerated matters a little .

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

a: Baxter

Progress of RC Datasets

- HFL-RC: PD&CFT (Cui et al., COLING2016)

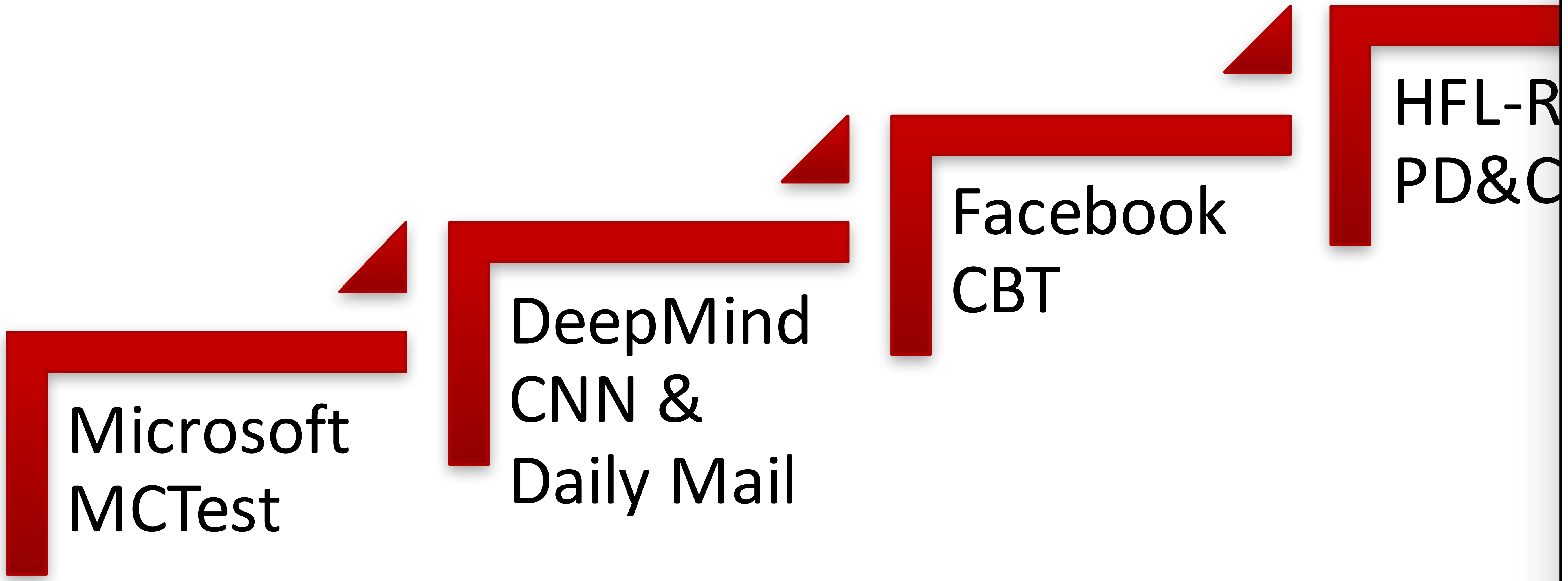
<table>
<tr><td>Document</td><td>1 ||| 人民网 1月 1日 讯 据 《纽约 时报》 报道， 美国 华尔街 股市 在 2013年 的 最后 一 天 继续 上涨， 和 全球 股市 一样， 都 以 最高 纪录 或 接近 最高 纪录 结束 本年 的 交易。
2 ||| 《纽约 时报》 报道 说， 标普 500 指数 今年 上升 29.6%， 为 1997年 以来 的 最 大 涨幅 ；
3 ||| 道琼斯 工业 平均 指数 上升 26.5%， 为 1996年 以来 的 最 大 涨幅 ；
4 ||| 纳斯达克 上涨 38.3%。
5 ||| 就 12月 31日 来说， 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。
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。
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。
8 ||| 这 一 年 里， 投资 美国 股市 的 明智 做法 是 追 着 “傻钱” 跑。
9 ||| 所谓 的 “傻钱” X， 其实 就 是 买 入 并 持有 美国 股票 这样 的 普通 组合。
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。</td></tr>
<tr><td>Query</td><td>所谓 的 “傻钱” X， 其实 就 是 买 入 并 持有 美国 股票 这样 的 普通 组合。</td></tr>
<tr><td>Answer</td><td>策略</td></tr>
</table>

Microsoft MCTest

Deep CNN & Daily Mail

Progress of RC Datasets

- Stanford SQuAD (Rajpurkar et al., EMNLP2016)



In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

OUTLINE

- Introduction to Machine Reading Comprehension (MRC)
- **(Almost) Recent Progress of MRC**
  - **Cloze-style MRC**
  - Complex MRC
- Future Prospects
- Conclusion

CLOZE-STYLE RC

- Definition of cloze-style RC
  - 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)

Original Version

Context
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." …

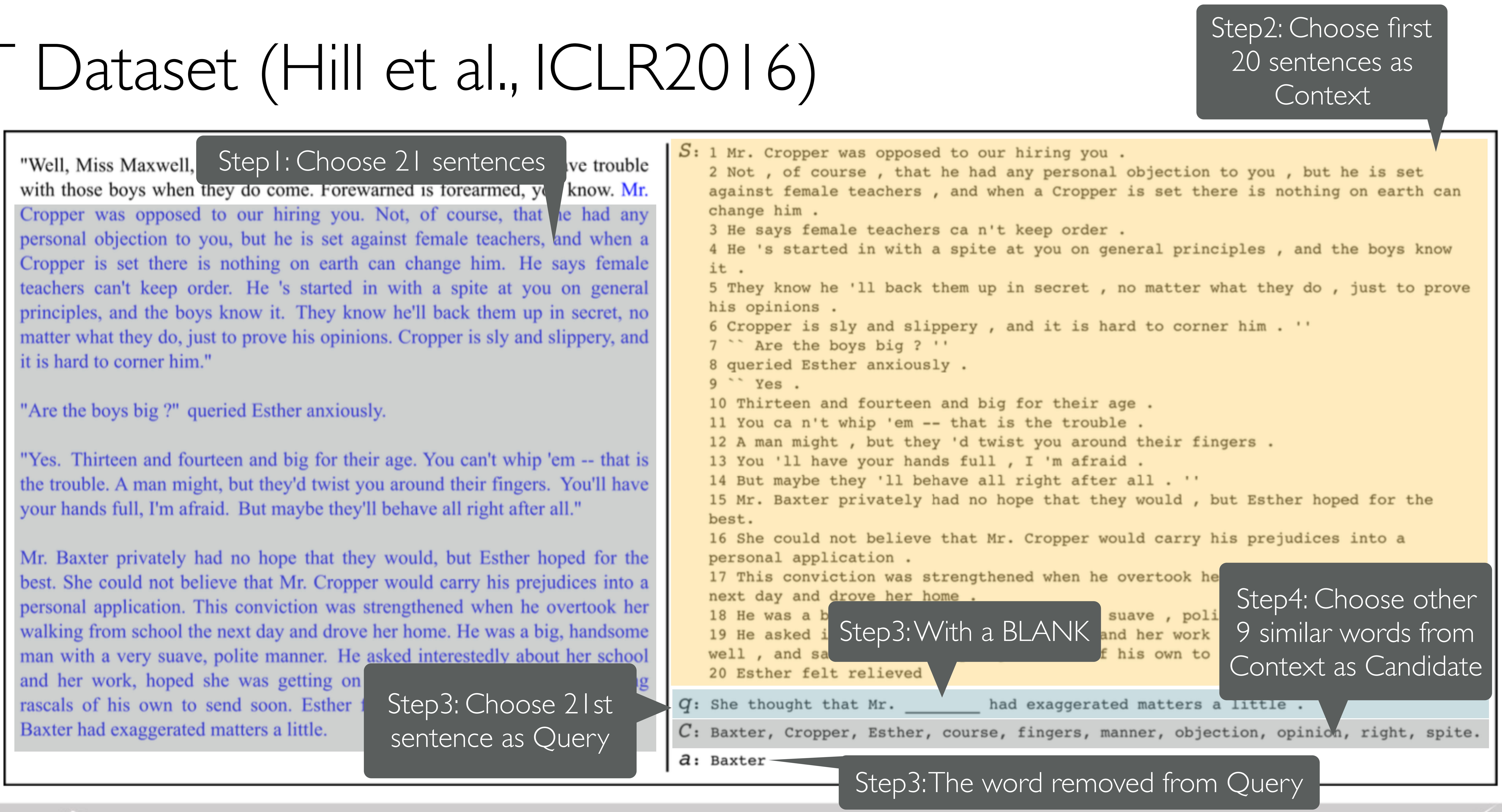
Query
Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.

Answer
Oisin Tymon

*Example is chosen from the CNN dataset (Hermann et al., 2015)

CLOZE-STYLE RC

- CBT Dataset (Hill et al., ICLR2016)

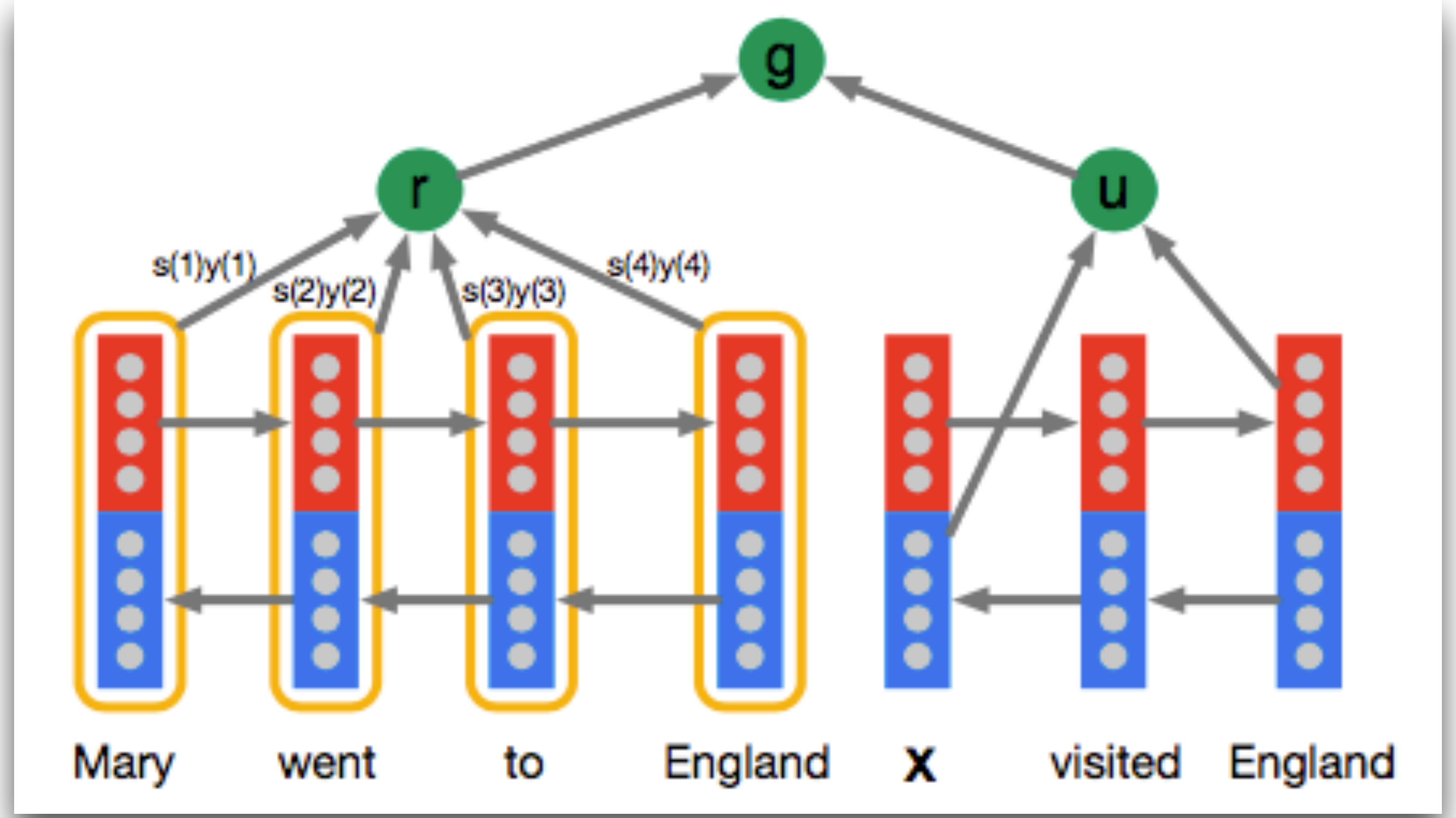


RELATED WORKS

- **Predictions on full vocabulary**
  - Attentive Reader (Hermann et al., 2015)
  - Stanford AR (Chen et al., 2016)
- **Pointer-wise predictions (Vinyals et al., 2015)**
  - Attention Sum Reader (Kadlec et al., 2016)
  - Consensus Attention Reader (Cui et al., 2016)
  - Gated-attention Reader (Dhingra et al., 2017)

Attentive Reader

- Teaching Machines to Read and Comprehend (Hermann et al., NIPS2015)
- Propose **attention-based neural networks** for reading comprehension

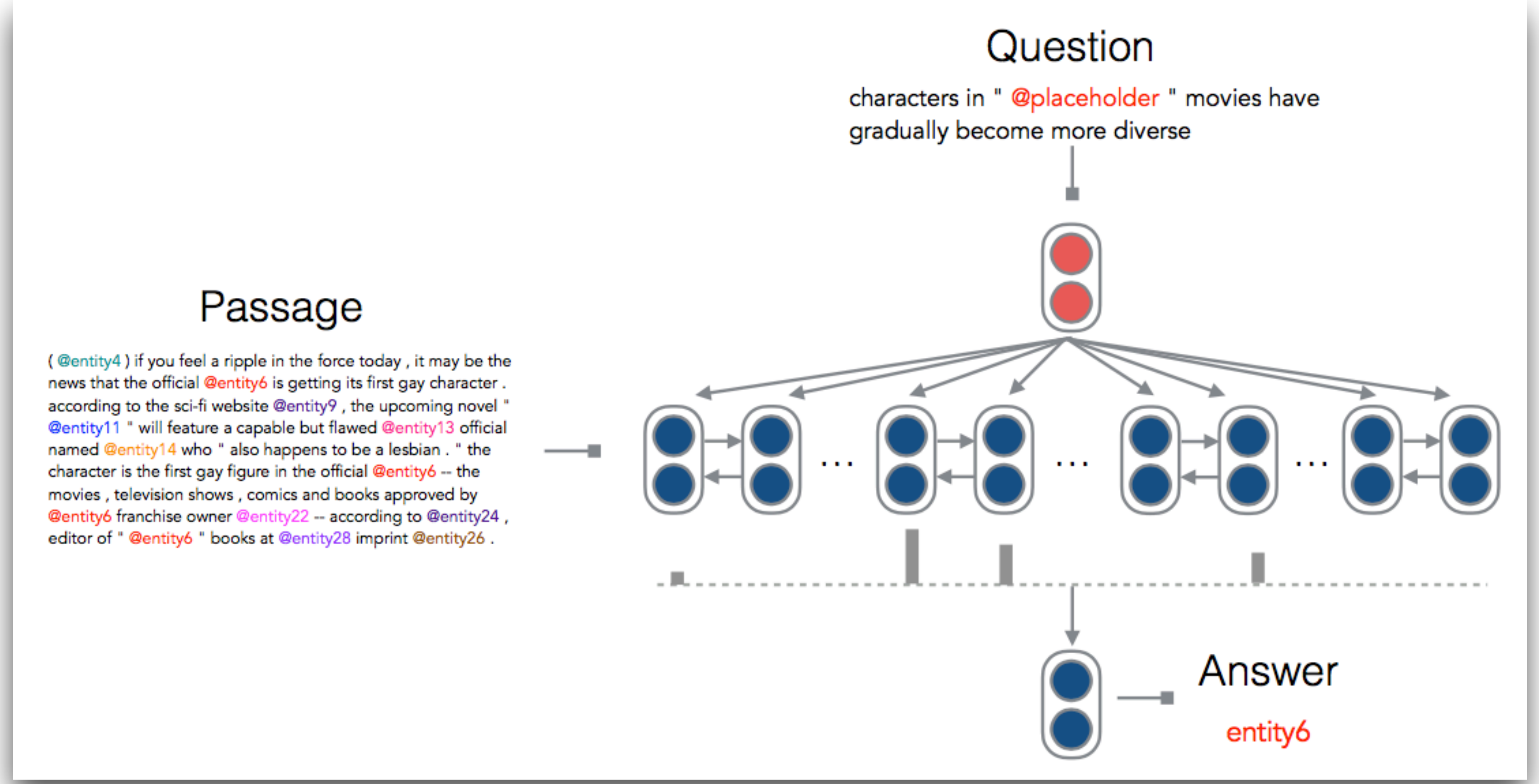


$$m(t) = \tanh\left(W_{ym} y_d(t) + W_{um} u\right),$$
$$s(t) \propto \exp\left(\mathrm{w}_{ms}^{\intercal} m(t)\right),$$
$$r = y_d s,$$

$$g^{\mathrm{AR}}(d, q) = \tanh\left(W_{rg} r + W_{ug} u\right).$$

STANFORD AR

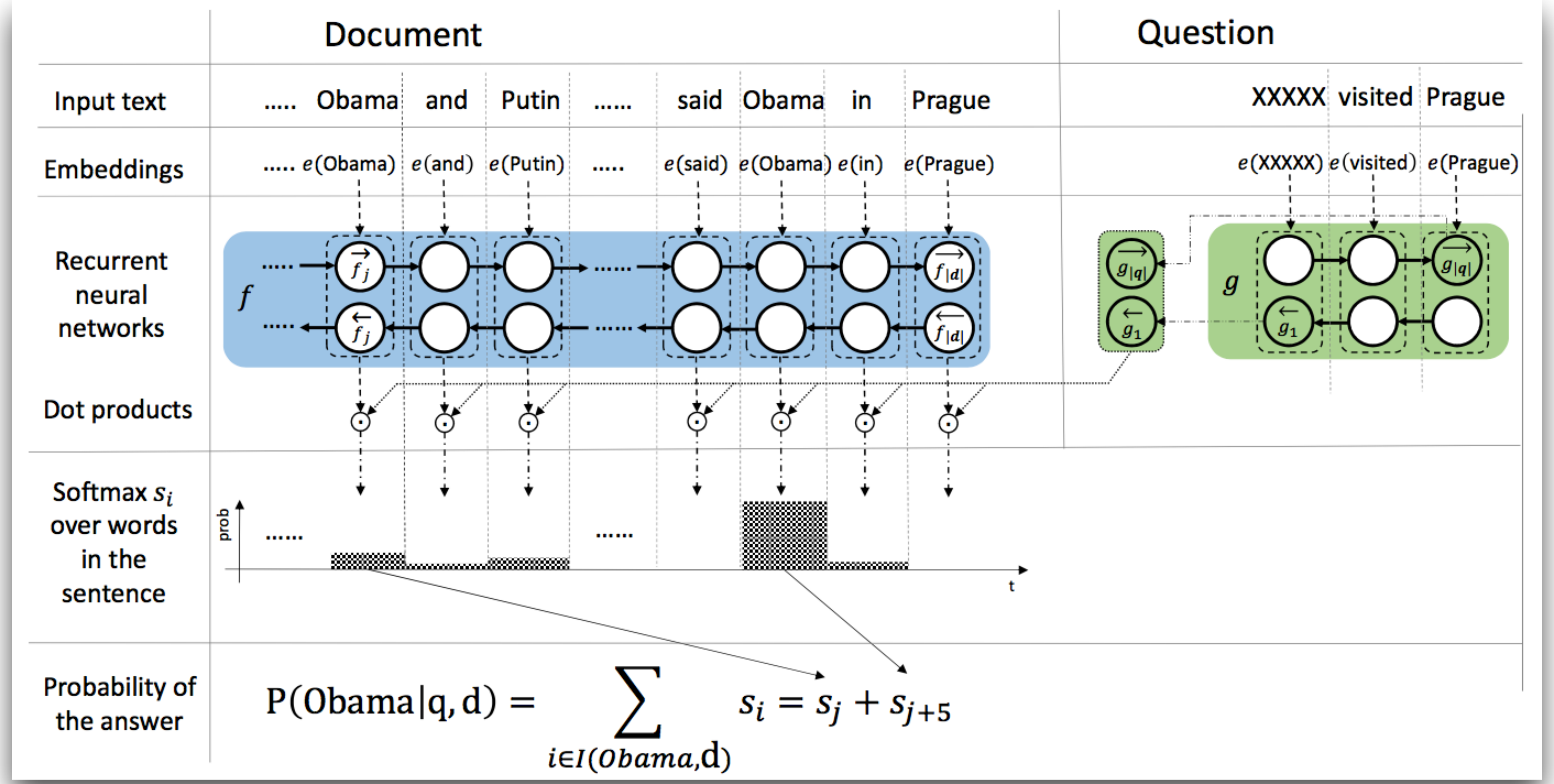
- A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task (Chen et al., ACL2016)
- Nothing special in NN model, but provides **valuable insights in the CNN/DailyMail datasets**



1) CNN/DailyMail dataset is noisy
2) Current NN models have almost reached CEILING performance
3) Requires less reasoning and inference

Attention Sum Reader

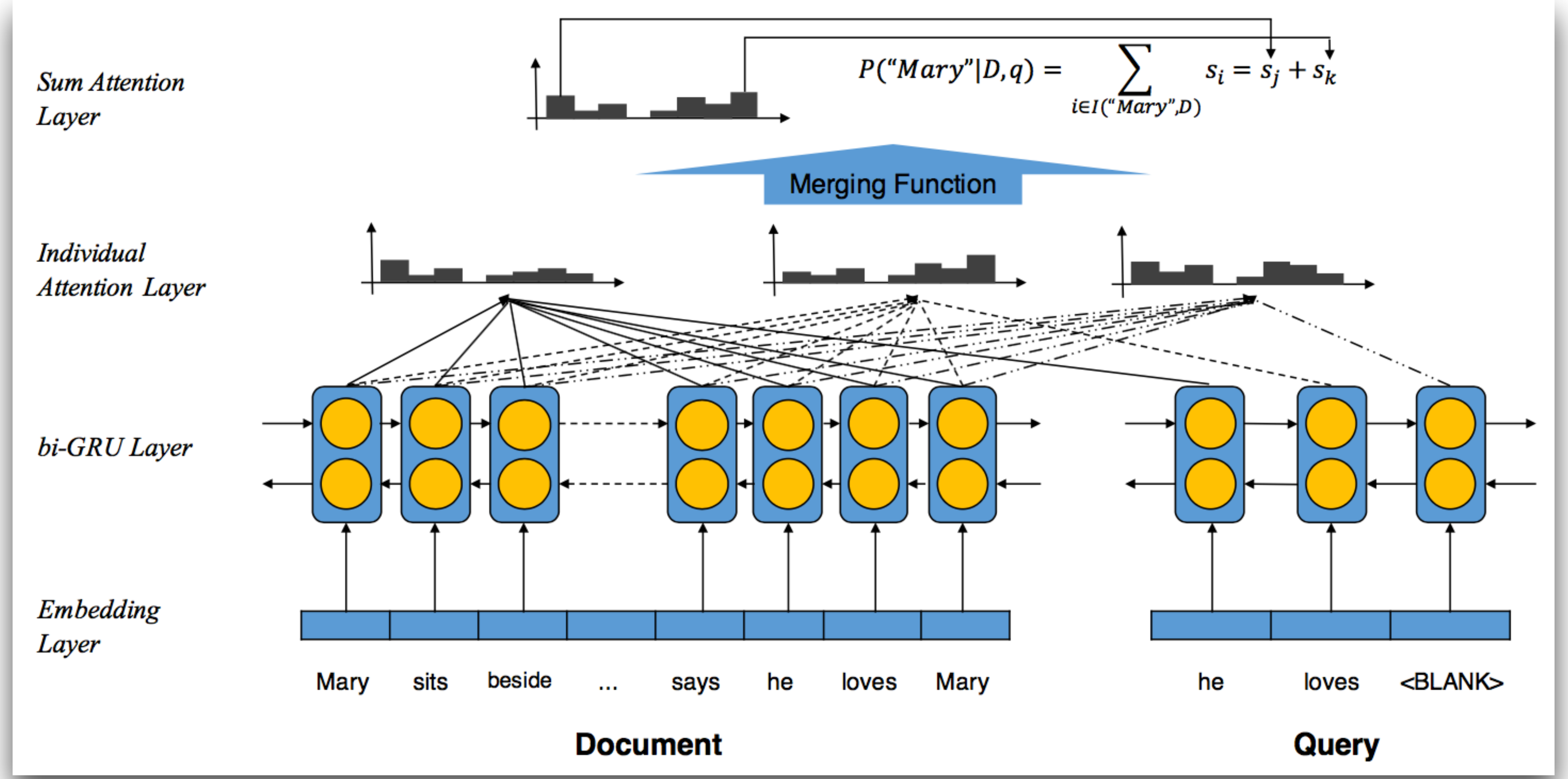
- Text Understanding with the Attention Sum Reader Network (Kadlec et al., ACL2016)
- Propose to **utilize and improve Pointer Network** (Vinyals et al., 2015) in RC



Sum Attention Mechanism

CAS READER

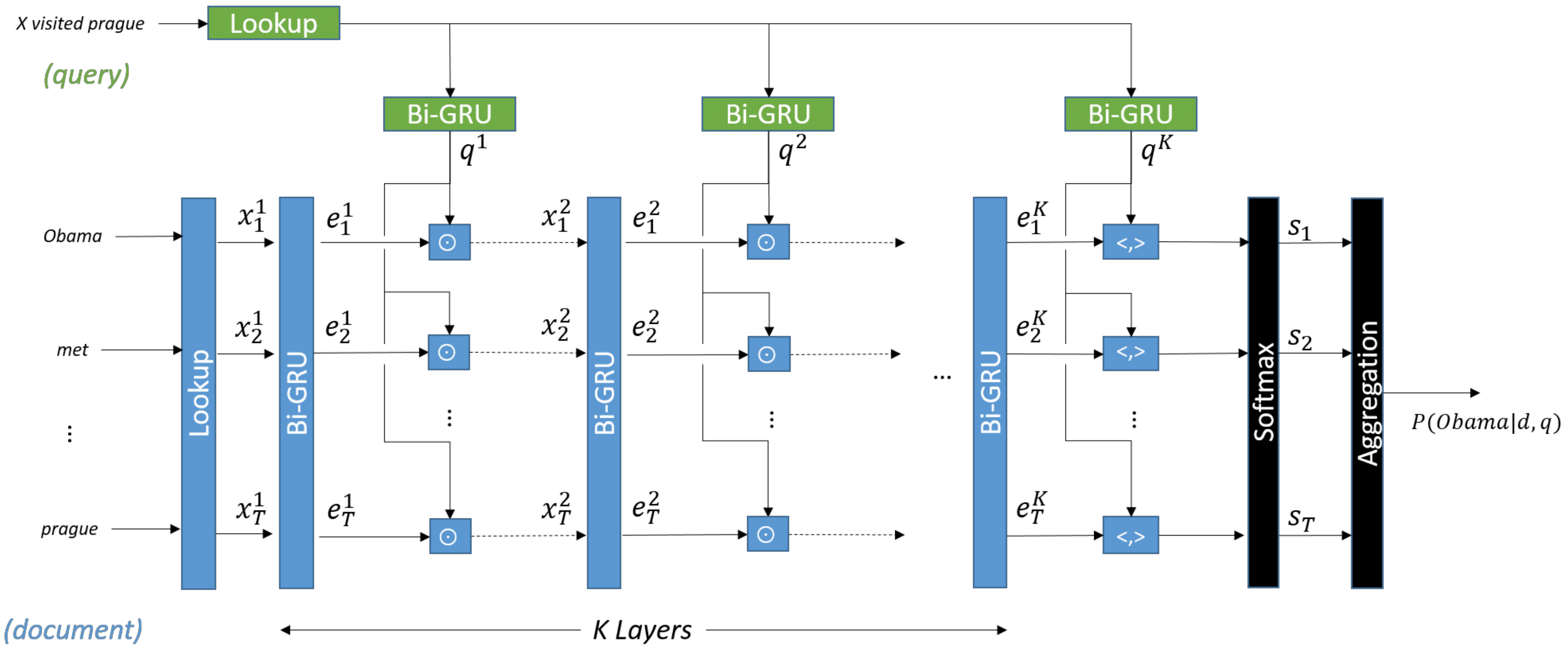
- Consensus Attention-based Neural Networks for Chinese Reading Comprehension (Cui et al., COLING2016)
- Propose to consider different timesteps of query and **generate multiple doc-level attentions**



Gated-Attention Reader

- Gated-Attention Reader for Text Comprehension (Dhingra et al., ACL2017)

- Propose to use **multiple hops** for refining attended representations

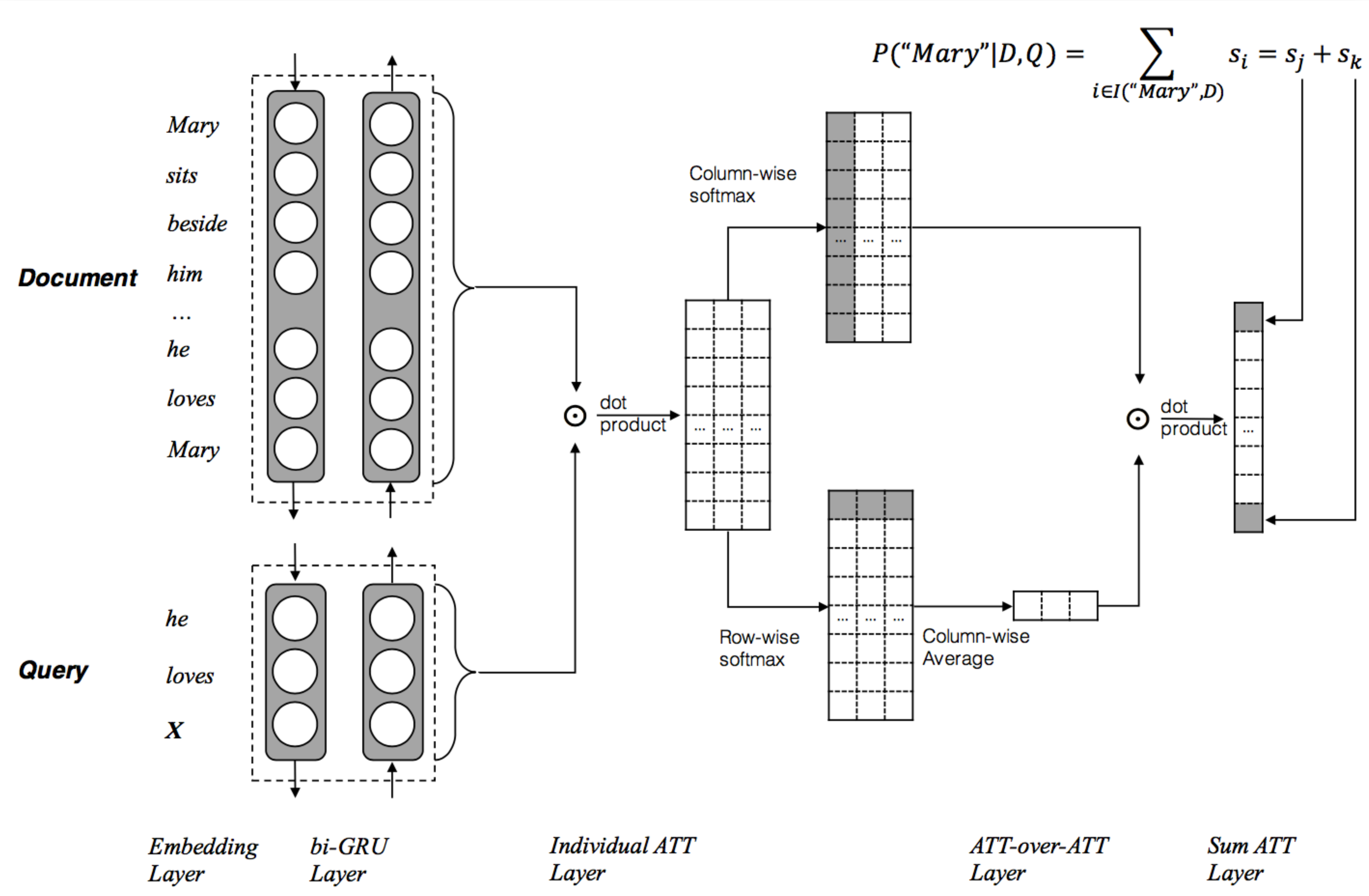


AoA Reader

- **Attention-over-Attention Neural Networks for Reading Comprehension (Cui et al., ACL2017)**
- Primarily motivated by AS Reader (Kadlec et al., 2016) and CAS Reader (Cui et al., 2016)
  - Introduce matching matrix for indicating doc-query relationships
  - Mutual attention: doc-to-query and query-to-doc
  - Instead of using heuristics to combine individual attentions, we place another attention to dynamically assign weights to the individual ones
- Some of the ideas in our work has already been adopted in the follow-up works not only in cloze-style RC but also other types of RC (such as SQuAD).

AoA Reader

- Model architecture at a glance



AoA Reader

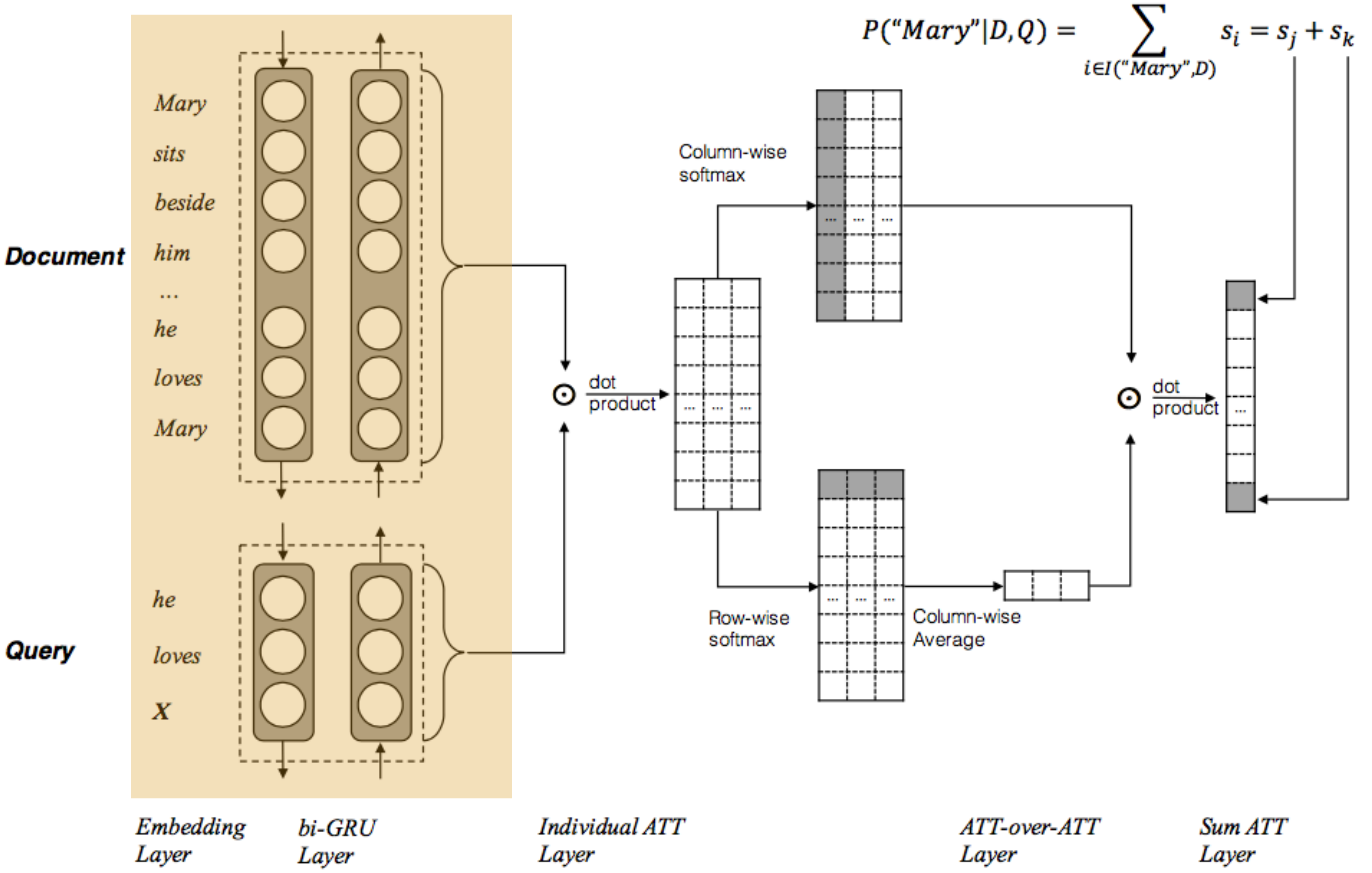
- **Contextual Embedding**
  - Transform document and query into contextual representations using word-embeddings and bi-GRU units

$$e(x) = W_e \cdot x,\ where\ \ x \in \mathcal{D}, \mathcal{Q} \qquad (1)$$

$$\overrightarrow{h_s(x)} = \overrightarrow{GRU}(e(x)) \qquad (2)$$

$$\overleftarrow{h_s(x)} = \overleftarrow{GRU}(e(x)) \qquad (3)$$

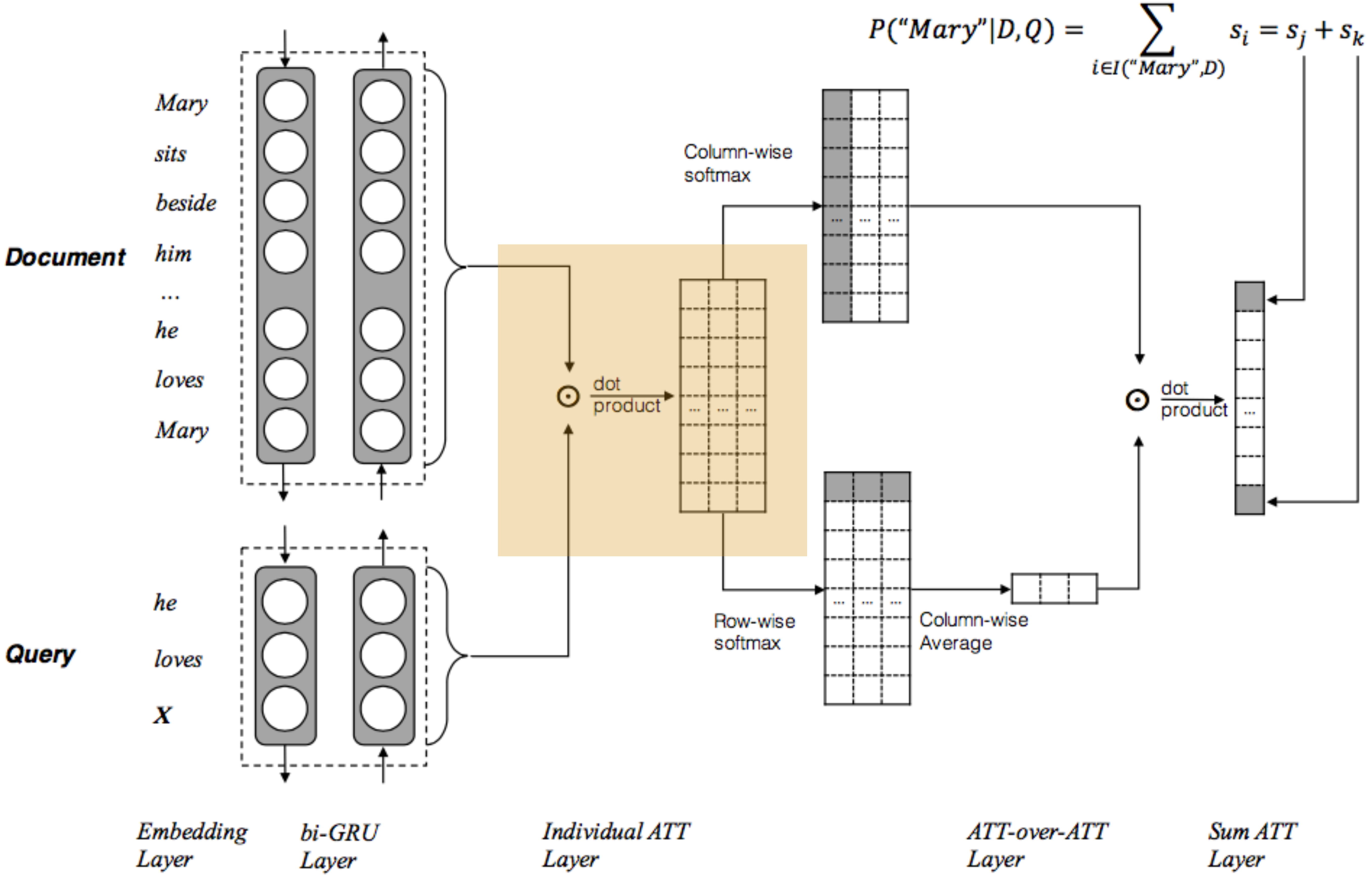
$$h_s(x) = [\overrightarrow{h_s(x)}; \overleftarrow{h_s(x)}] \qquad (4)$$



AoA Reader

- **Pair-wise Matching Score**
  - Calculate similarity between each document word and query word
  - For simplicity, we just calculate dot product between document and query word

$$M(i,j) = h_{doc}(i)^T \cdot h_{query}(j) \qquad (5)$$

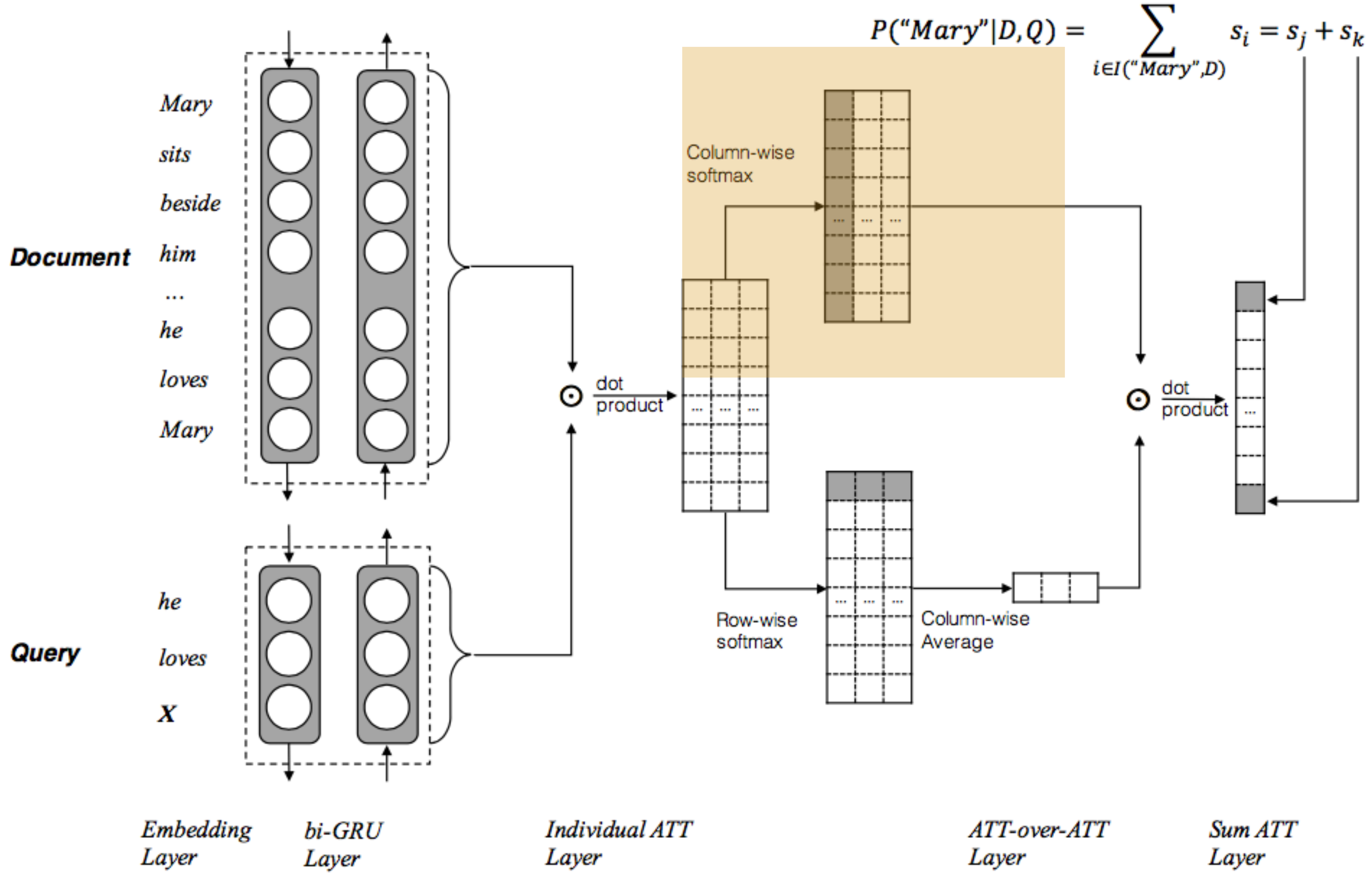


AoA Reader

- **Individual Attentions**
  - Calculate document-level attention with respect to each query word

$$\alpha(t) = softmax(M(1,t), ..., M(|\mathcal{D}|,t)) \quad (6)$$

$$\alpha = [\alpha(1), \alpha(2), ..., \alpha(|\mathcal{Q}|)] \quad (7)$$



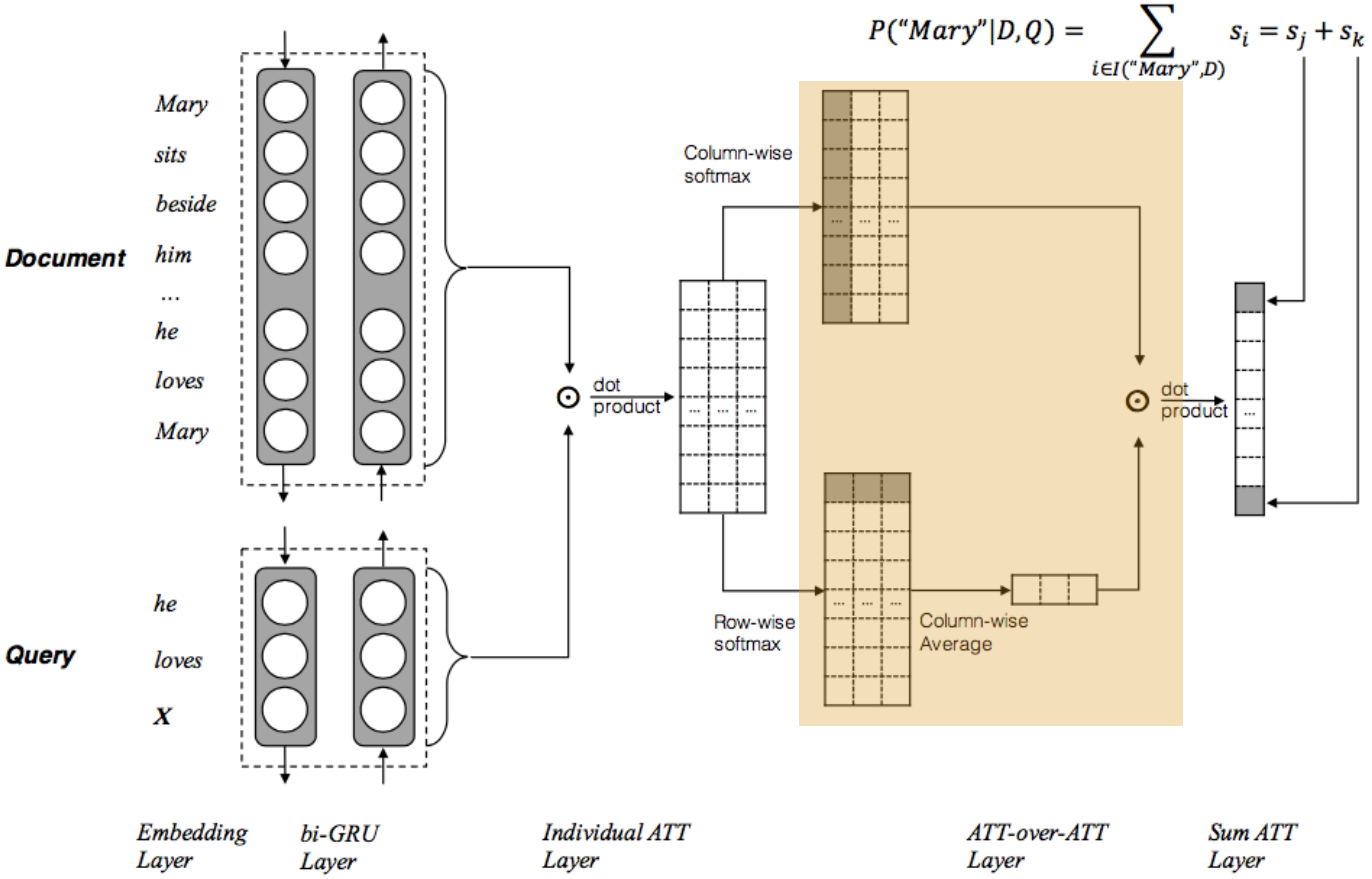
AoA Reader

- **Attention-over-Attention**
  - Dynamically assign weights to individual doc-level attentions
  - Get "attended attention"

$$\beta(t) = softmax(M(t,1), ..., M(t, |\mathcal{Q}|)) \quad (8)$$

$$\beta = \frac{1}{n} \sum_{t=1}^{|\mathcal{D}|} \beta(t) \quad (9)$$

$$s = \alpha^T \beta \quad (10)$$

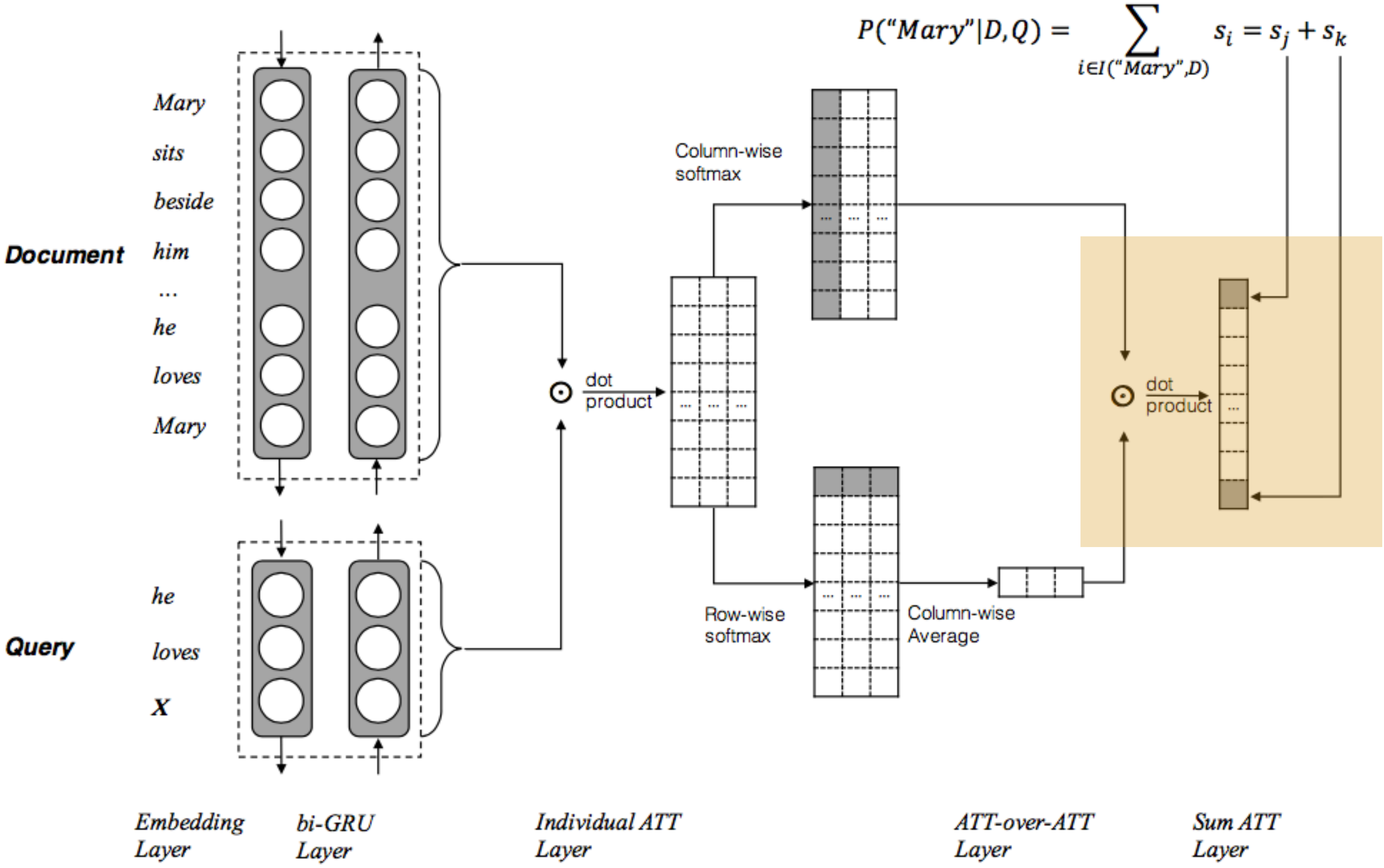


AoA Reader

- **Final Predictions**
  - Adopt Pointer Network (Vinyals et al., 2015) for predictions
  - Apply **sum-attention mechanism** (Kadlec et al., 2016) to get the final probability of the answer

$$P(w|\mathcal{D},\mathcal{Q}) = \sum_{i \in I(w,\mathcal{D})} s_i, \quad w \in V \qquad (11)$$

$$\mathcal{L} = \sum_i \log(p(x)) \quad , x \in \mathcal{A} \qquad (12)$$



AoA Reader

- **An intuitive example**

- Let say this is a story about `Tom bought a diamond ring for his beloved girl friend…`

	Tom	loves	<blank>	.
Query-level Attention	0.5	0.3	0.15	0.05
Candidate Answers	Mary = 0.6 diamond = 0.3 beside = 0.1	Mary = 0.3 diamond = 0.5 beside = 0.2	Mary = 0.4 diamond = 0.4 beside = 0.2	Mary = 0.2 diamond = 0.4 beside = 0.4
Average Score (CAS Reader)	Mary = (0.6+0.3+0.4+0.2) / 4 = 0.375 **diamond = (0.3+0.5+0.4+0.4) / 4 = 0.400** beside = (0.1+0.2+0.2+0.4) / 4 = 0.225			
Weighted Score (AoA Reader)	**Mary = 0.6*0.5+0.3*0.3+0.4*0.15+0.2*0.05 = 0.460** diamond = 0.3*0.5+0.5*0.3+0.4*0.15+0.4*0.05 = 0.380 beside = 0.1*0.5+0.2*0.3+0.2*0.15+0.4*0.05 = 0.160			

AoA Reader

- **N-best re-ranking strategy for cloze-style RC**
  - Mimic the process of double-checking, in terms of fluency, grammatical correctness etc.
  - Main idea: Re-fill the candidate answer into the blank of query to form a complete sentence and using additional features to score the sentences

Generate candidate answers
N-best decoding

↓

Re-fill the candidate into query

↓

Scoring with additional features
(such as N-gram LM)

↓

Feature weight tuning
KBMIRA (Cherry and Foster, 2012)

↓

Re-scoring and Re-ranking

AoA Reader

- **Single model performance**
  - Significantly outperform previous works
  - Re-ranking strategy substantially improve performance
  - Introducing attention-over-attention mechanism instead of using heuristic merging function (Cui et al., 2016) may bring significant improvements
- **Ensemble performance**
  - We use greedy ensemble approach of 4 models trained on different random seed
  - Significant improvements over state-of-the-art systems

	CNN News		CBTest NE		CBTest CN	
	Valid	Test	Valid	Test	Valid	Test
Deep LSTM Reader (Hermann et al., 2015)	55.0	57.0	-	-	-	-
Attentive Reader (Hermann et al., 2015)	61.6	63.0	-	-	-	-
Human (context+query) (Hill et al., 2015)	-	-	-	*81.6*	-	*81.6*
MemNN (window + self-sup.) (Hill et al., 2015)	63.4	66.8	70.4	66.6	64.2	63.0
AS Reader (Kadlec et al., 2016)	68.6	69.5	73.8	68.6	68.8	63.4
CAS Reader (Cui et al., 2016)	68.2	70.0	74.2	69.2	68.2	65.7
Stanford AR (Chen et al., 2016)	72.4	72.4	-	-	-	-
GA Reader (Dhingra et al., 2016)	73.0	73.8	74.9	69.0	69.0	63.9
Iterative Attention (Sordoni et al., 2016)	72.6	73.3	75.2	68.6	72.1	69.2
EpiReader (Trischler et al., 2016)	**73.4**	74.0	75.3	69.7	71.5	67.4
AoA Reader	73.1	**74.4**	**77.8**	**72.0**	**72.2**	**69.4**
AoA Reader + Reranking	-	-	**79.6**	**74.0**	**75.7**	**73.1**
MemNN (Ensemble)	66.2	69.4	-	-	-	-
AS Reader (Ensemble)	73.9	75.4	74.5	70.6	71.1	68.9
GA Reader (Ensemble)	76.4	77.4	75.5	71.9	72.1	69.4
EpiReader (Ensemble)	-	-	76.6	71.8	73.6	70.6
Iterative Attention (Ensemble)	74.5	75.7	76.9	72.0	74.1	**71.0**
AoA Reader (Ensemble)	-	-	**78.9**	**74.5**	**74.7**	70.8
AoA Reader (Ensemble + Reranking)	-	-	**80.3**	**75.6**	**77.0**	**74.1**

TAKEAWAYS - 1

- **What are the gooooooooood things in cloze-style RC?**
  - Pointer Network is especially useful in this task, as the answer is assumed to be existed in the document, just directly PICK the right answer from document
  - A simple **DOT product** is capable of attention calculation
  - **Mutual attention mechanism** could bring additional information, using both doc-to-query and query-to-doc attentions
  - Re-ranking strategy with **traditional N-gram LMs** could substantially improve cloze-style RC performance due to its nature

OUTLINE

- Introduction to Machine Reading Comprehension (MRC)
- **(Almost) Recent Progress of MRC**
  - Cloze-style MRC
  - **Complex MRC**
- Future Prospects
- Conclusion

Complex MRC

- **Complex MRC (informal definition)**
- Including following types
  - Span extraction from passage (phrase, sentences etc.)
  - Generate answer directly
  - Choose correct answer from candidate choices (A/B/C/D)
- Span extraction is the most popular type in recent studies

SQuAD

- SQuAD: 100,000+ Questions for Machine Comprehension of Text (Rajpurkar et al., EMNLP2016)
- Dataset Features
  - More Difficult: word-level answers → words, phrases or even sentences
  - High Quality: automatically generated data → human-annotated data
  - Much Bigger: 100K+ questions, bigger than previous human-annotated RC datasets

SQuAD

The Stanford Question Answering Dataset

SQuAD

- **Sample of SQuAD**
  - **Document**: Passages from Wikipedia pages, segment into several small paragraphs
  - **Query**: Human-annotated, including various query types (what/when/where/who/how/why etc.)
  - **Answer**: Continuous segments (text spans) in the passage, which has a larger search space, and much harder to answer than cloze-style RC

Oxygen

The Stanford Question Answering Dataset

In the meantime, on August 1, 1774, an experiment conducted by the British clergyman Joseph Priestley focused sunlight on mercuric oxide (HgO) inside a glass tube, which liberated a gas he named "dephlogisticated air". He noted that candles burned brighter in the gas and that a mouse was more active and lived longer while breathing it. After breathing the gas himself, he wrote: "The feeling of it to my lungs was not sensibly different from that of common air, but I fancied that my breast felt peculiarly light and easy for some time afterwards." Priestley published his findings in 1775 in a paper titled "An Account of Further Discoveries in Air" which was included in the second volume of his book titled Experiments and Observations on Different Kinds of Air. Because he published his findings first, Priestley is usually given priority in the discovery.

Why is Priestley usually given credit for being first to discover oxygen?
Ground Truth Answers: published his findings first | he published his findings first | he published his findings first | he published his findings first | Because he published his findings first

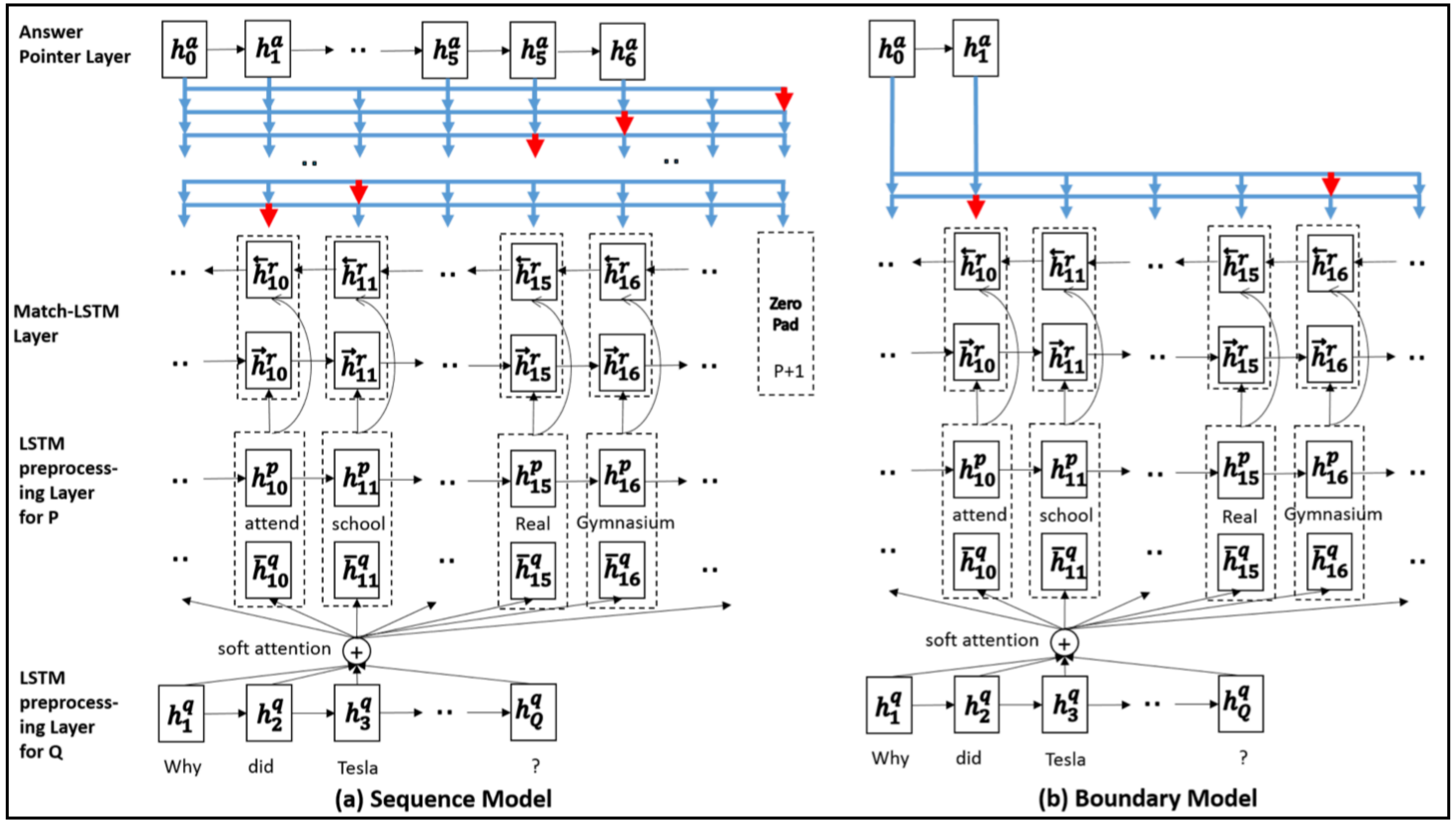
RELATED WORKS

- A large amount of researchers are investigating SQuAD after its release. Tons of models are proposed.

- **Representative Works**
  - Match-LSTM (Wang and Jiang, 2016)
  - Bi-directional Attention Flow (BiDAF) (Seo et al., 2016)
  - Dynamic Coattention Network (DCN) (Xiong et al., 2017)
  - r-net (Wang et al., 2017)

Match-LSTM

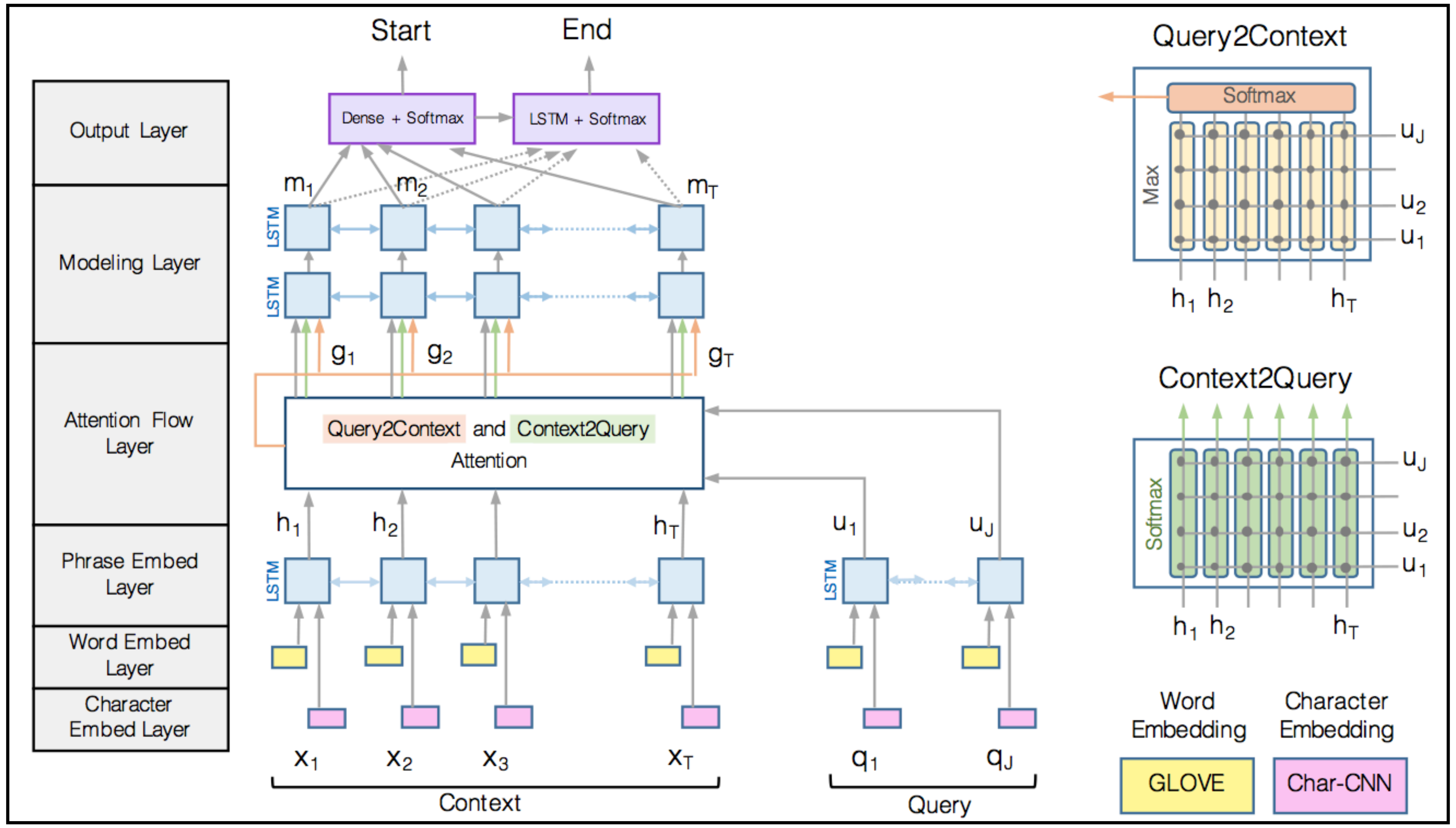
- Machine Comprehension using Match-LSTM and Answer Pointer (Wang and Jiang, 2016)
- Propose to use Pointer Network to **directly output start and end position in document**



BiDAF

- Bi-Directional Attention Flow for Machine Comprehension (Seo et al., 2016)

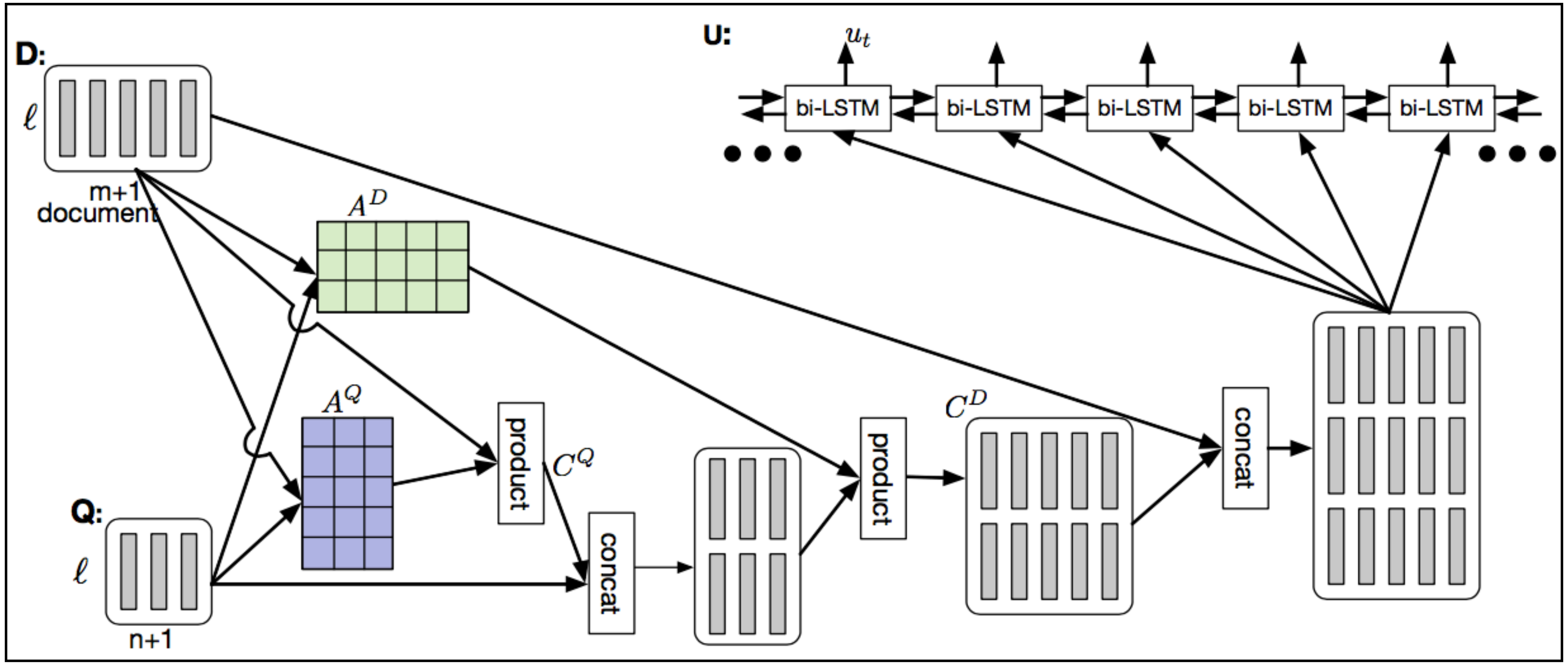
- Propose bi-directional attention, which has become a **<u>stereotype</u>** in SQuAD task



DCN

- **Dynamic Coattention Networks for Question Answering (Xiong et al., 2016)**

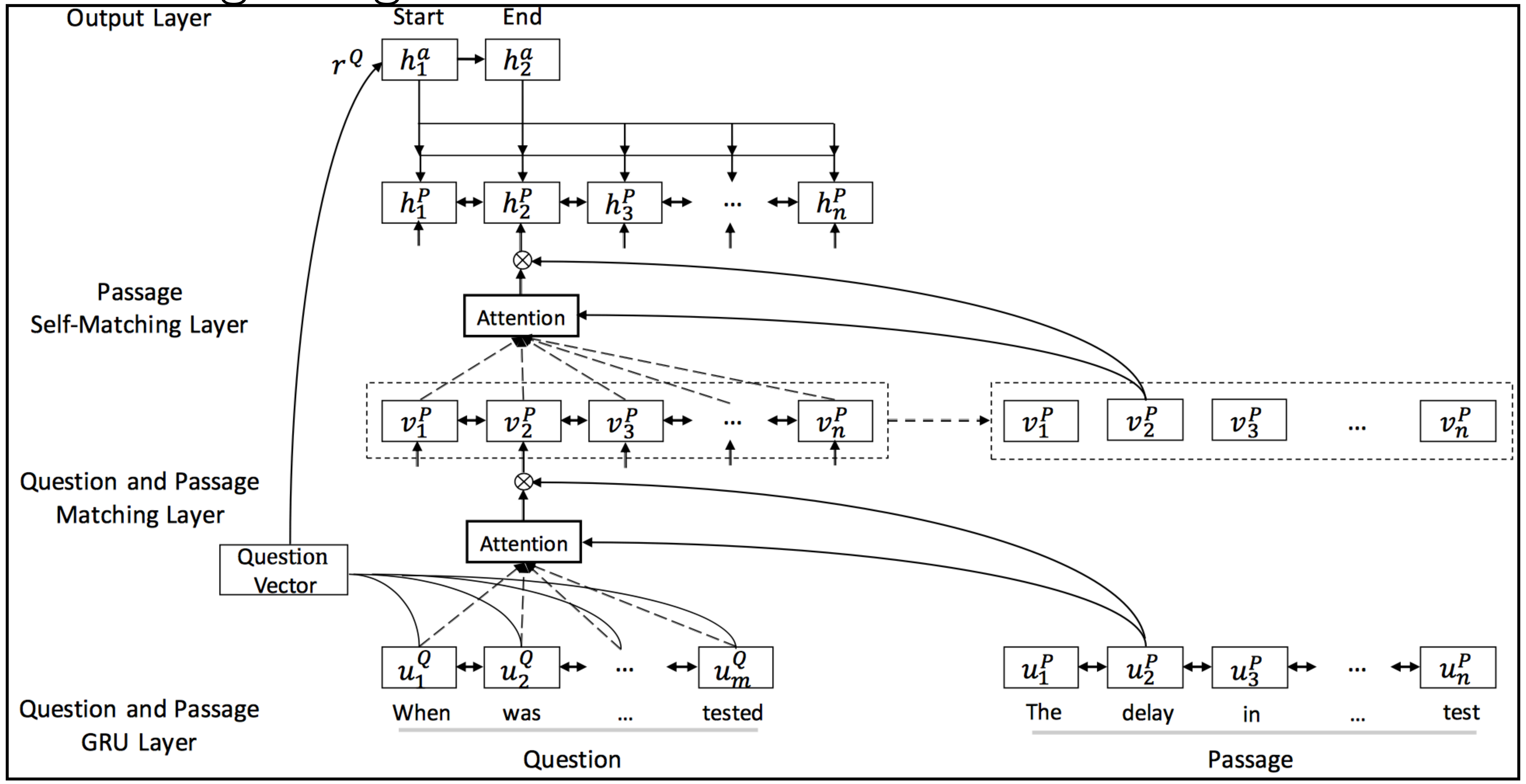
- Propose dynamic coattention model, iterative pointer mechanism



R-NET

- **Gated Self-Matching Networks for Reading Comprehension and Question Answering (Wang et al., 2017)**

- Propose to use self-matching and gated attention



OUR WORK

- Interactive AoA Reader: an improved version of AoA Reader (Cui et al., 2017)

- We have been working on SQuAD task for months, and get on the 1st place in the late July, 2017

Rank	Model	EM	F1
1 Jul 2017	Interactive AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	77.845	85.297
2 Jun 2017	r-net (ensemble) *Microsoft Research Asia* http://aka.ms/rnet	77.688	84.666
3 Jul 2017	r-net (single model) *Microsoft Research Asia* http://aka.ms/rnet	75.705	83.496
3 Jul 2017	smarnet (ensemble) *Eigen Technology & Zhejiang University*	75.989	83.475
4 Jul 2017	DCN+ (single model) *Salesforce Research*	74.866	82.806

*As of August 1, 2017. http://stanford-qa.com

Our Work

- As our work is not published, we cannot reveal the detailed architecture and algorithms
- But…we can tell you a little bit of the techniques that adopted (published techniques with modifications)
  - Char+Word level embeddings
  - Multiple hops for representation refining
  - Incorporating historical attentions
- And more? Stay tuned for our papers!

TAKEAWAYS - II

- **What are the goooooooood things in SQuAD models?**
  - Old things still works
    - Pointer Network for directly predict start/end position in document
    - Mutual attention mechanism
  - What's new?
    - Word-level + Char-level embeddings
    - More complex attention calculation with multiple attended representations

OUTLINE

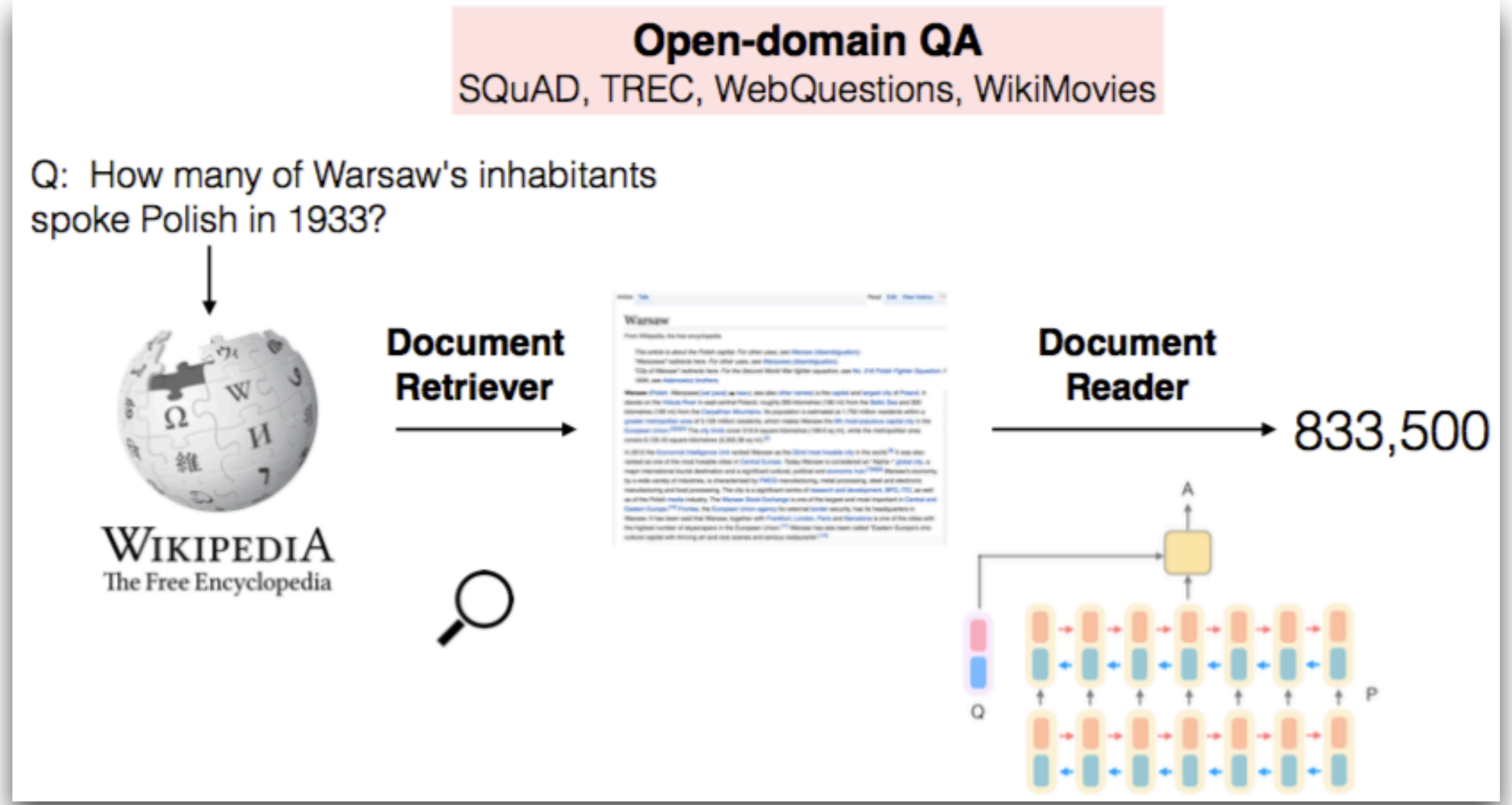
- Introduction to Machine Reading Comprehension (MRC)
- (Almost) Recent Progress of MRC
  - Cloze-style MRC
  - Span Extraction MRC
- **Future Prospects**
- Conclusion

Recent Highlights

- **Open-Domain MRC**
  - Reading Wikipedia to Answer Open-Domain Questions (Chen et al., ACL2017)
- **More complex MRC datasets**
  - RACE (Lai et al., EMNLP2017)
  - MS MARCO (Nguyen et al., 2016)
  - NewsQA (Trischeler et al., 2017)
  - TriviaQA (Joshi et al., ACL2017)

Open-Domain MRC

- Reading Wikipedia to Answer Open-Domain Questions (Chen et al., ACL2017)
  - Open-Domain MRC = Document Retriever + MRC



DATASET COMPARISON

- Comparisons on recent MRC datasets

Dataset	Size (# Query)	Question Source	Answer Type	Difficulty
SQuAD	~100K	Crowd-sourced	Passage span	Medium
RACE	~97K	Exams in China	Choice Selection (from A/B/C/D)	High
MS MARCO	~100K	User logs	Human generated	High
NewsQA	~120K	Crowd-sourced	Passage span	Relatively High
TriviaQA	~650K (~95K unique)	Auto-gathered / Human-annotated	Passage span	Relatively High

RACE

- RACE: Large-scale ReAding Comprehension Dataset From Examinations (Lai et al., EMNLP2017)
- Features
  - Needs more comprehensive understanding of context
  - Answer is no longer a span in document
  - Misleading choices among candidates
  - SOTA model in SQuAD failed to give excellent performance(70%+ → 40%)

Passage:

Is it important to have breakfast every day? A short time ago, a test was given in the United States. People of different ages, from 12 to 83, were asked to have a test. During the test, these people were given all kinds of breakfast, and sometimes they got no breakfast at all. Scientists wanted to see how well their bodies worked after eating different kinds of breakfast.

The results show that if a person eats a right breakfast, he or she will work better than if he or she has no breakfast. If a student has fruit, eggs, bread and milk before going to school, he or she will learn more quickly and listen more carefully in class. Some people think it will help you lose weight if you have no breakfast. But the result is opposite to what they think. This is because people become so hungry at noon that they eat too much for lunch. They will gain weight instead of losing it.

Question: What do the results show?

A) They show that breakfast has affected on work and studies.

B) The results show that breakfast has little to do with a person's work.

C) The results show that a person will work better if he only has fruit and milk.

D) They show that girl students should have less for breakfast.

MRC TRENDS

- **MRC Trends**
  - Automatically generated data → Human-annotated data
  - Single document → Multiple documents
  - Single sentence inference → Multiple sentences inference
  - Answer retrieval → Summarization/Reasoning/Inference
  - Model aspects: all in one is not all we need

OUTLINE

- Introduction to Machine Reading Comprehension (MRC)
- (Almost) Recent Progress of MRC
  - Cloze-style MRC
  - Span Extraction MRC
- Future Prospects
- **Conclusion**

CONCLUSION

- Current MRC ≈ More Complex **QA**
- Does MRC do reasoning?
  - Maybe not, but not totally negative
  - More studies should be done
- "All roads lead to Rome"——In the DL era, every model deserve a chance to win the game (if you use it in correct way)

♫ ~ ADVERTISEMENT TIME ~ ♫

CCL-CMRC2017

CCL阅读理解评测

- The 1st Evaluation Workshop on Chinese Machine Reading Comprehension (CMRC2017)
  - Hosted by CIPS, organized by Joint Laboratory of HIT and iFLYTEK(HFL)
  - Date: 2017/10/14 14:00 ~ 17:30
  - Co-locate with CCL2017 at Nanjing
  - Welcome to join us !

填空类问题（Cloze-style Question）

最终排名	参赛单位	单/多系统	开发集准确率	测试集准确率↓
1	6ESTATES PTE LTD	多系统	81.85%	81.90%
2	上海交通大学仿脑计算与机器智能研究中心自然语言组 Shanghai Jiao Tong University (SJTU BCMI-NLP)	多系统	78.35%	80.67%
3	南京云思创智信息科技有限公司	多系统	79.20%	80.27%

用户提问类问题（User-Query Question）

最终排名	参赛单位	单/多系统	开发集准确率	测试集准确率↓
1	华东师范大学 East China Normal University (ECNU)	多系统	90.45%	69.53%
2	山西大学三队 Shanxi University (SXU-3)	单系统	47.80%	49.07%
3	郑州大学 Zhengzhou University (ZZU)	单系统	31.10%	32.53%

REFERENCES

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 .
- 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).
- Colin Cherry and George Foster. 2012. Batch tuning strategies for statistical machine translation. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Montre´al, Canada, pages 427–436. http://www.aclweb.org/anthology/N12-1047.
- Kyunghyun Cho, Bart van Merrienboer, 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). Association for Computational Linguistics, pages 1724–1734. http://aclweb.org/anthology/D14-1179.
- François Chollet. 2015. Keras. https://github. com/fchollet/keras.
- Yiming Cui, Ting Liu, Zhipeng Chen, Shijin Wang, and Guoping Hu. 2016. Consensus attention-based neural networks for chinese reading comprehension. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan, pages 1777–1786.
- Bhuwan Dhingra, Hanxiao Liu, William W Cohen, and Ruslan Salakhutdinov. 2016. Gated-attention readers for text comprehension. arXiv preprint arXiv: 1606.01549 .

REFERENCES

- Karl Moritz Hermann, Tomas Kocisky, 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 .
- Franz Josef Och. 1999. An efficient method for determining bilingual word classes. In Ninth Conference of the European Chapter of the Association for Computational Linguistics. http://aclweb.org/anthology/E99-1010.
- Rudolf Kadlec, Martin Schmid, Ondrej 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 .
- 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 . In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL-2017).
- Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. ICML (3) 28:1310–1318.

REFERENCES

- 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 .
- Alessandro Sordoni, Phillip Bachman, and Yoshua Bengio. 2016. Iterative alternating neural attention for machine reading. arXiv preprint arXiv: 1606.02245
- Lai, Guokun, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. "RACE: Large-scale ReAding Comprehension Dataset From Examinations." arXiv preprint arXiv:1704.04683 (2017).
- 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.
- Andreas Stolcke. 2002. Srilm - an extensible language modeling toolkit. In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002). pages 901–904.
- 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. http://arxiv.org/abs/1605.02688.

REFERENCES

- Adam Trischler, Zheng Ye, Xingdi Yuan, and Kaheer Suleman. 2016. Natural language comprehension with the epireader. arXiv preprint arXiv:1606.02270 .
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In Advances in Neural Information Processing Systems. pages 2692–2700.
- Chen, Danqi, Adam Fisch, Jason Weston, and Antoine Bordes. "Reading Wikipedia to Answer Open-Domain Questions." arXiv preprint arXiv:1704.00051 (2017).
- Trischler, Adam, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. "NewsQA: A Machine Comprehension Dataset." arXiv preprint arXiv:1611.09830 (2016).
- Joshi, Mandar, Eunsol Choi, Daniel S. Weld, and Luke Zettlemoyer. "TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension." arXiv preprint arXiv:1705.03551 (2017).
- Nguyen, Tri, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. "Ms marco: A human generated machine reading comprehension dataset." arXiv preprint arXiv:1611.09268 (2016).
- 崔一鸣，刘挺，王士进，陈致鹏，马文涛，胡国平.《斯坦福SQuAD挑战赛的中国亮丽榜单》，中国计算机学会通讯2017年9月刊

Thank you！

Slides Download

Contact: me [at] ymcui [dot] com