Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification

Yizhong Wang[1]   Kai Liu[2]   Jing Liu[2]   Wei He[2]

Yajuan Lyu[2]   Hua Wu[2]   Sujian Li[1]   Haifeng Wang[2]

[1] MOE Key Laboratory of Computational Linguistics, Peking University
[2] Baidu Inc.

ACL, July 17, 2018

Outline

❑ **Background / Motivation**
- Machine Reading Comprehension (MRC)
- Why Multi-Passage MRC is Challenging?

❑ **Model Architecture**
- Answer Boundary Prediction
- Answer Content Modeling
- Cross-Passage Answer Verification
- Joint Training and Prediction

❑ **Experiments**
- Results on MS-MARCO and DuReader
- Ablation Study
- Quantitative Analysis

❑ **Conclusion**

Machine Reading Comprehension (MRC)

Passage: … Tesla later approached Morgan to ask for more funds to build a more powerful transmitter. When asked where all the money had gone, Tesla responded by saying that he was affected by the Panic of 1901, which he (Morgan) had caused Morgan was shocked by the reminder of his part in the stock market …

Question: On what did Tesla blame for the loss of the initial money?

[from SQuAD v1.1[1]]

Machine Reading Comprehension (MRC)

Passage: … Tesla later approached Morgan to ask for more funds to build a more powerful transmitter. **When asked where all the money had gone, Tesla responded by saying that he was affected by the Panic of 1901**, which he (Morgan) had caused Morgan was shocked by the reminder of his part in the stock market …

Question: On what did Tesla blame for the loss of the initial money?

Answer: Panic of 1901

[from SQuAD v1.1[1]]

Machine Reading Comprehension (MRC)

Passage: … Tesla later approached Morgan to ask for more funds to build a more powerful transmitter. **When asked where all the money had gone, Tesla responded by saying that he was affected by the Panic of 1901**, which he (Morgan) had caused Morgan was shocked by the reminder of his part in the stock market …

Question: On what did Tesla blame for the loss of the initial money?

Answer: Panic of 1901

[from SQuAD v1.1[1]]

Single-passage MRC

Machine Reading Comprehension (MRC)

Passage: … Tesla later approached Morgan to ask for more funds to build a more powerful transmitter. **When asked where all the money had gone, Tesla responded by saying that he was affected by the Panic of 1901**, which he (Morgan) had caused Morgan was shocked by the reminder of his part in the stock market …

Question: On what did Tesla blame for the loss of the initial money?

Answer: Panic of 1901

[from SQuAD v1.1[1]]

Single-passage MRC

- **Different types:** cloze test, entity extraction, span extraction, multiple-choice …

- **Various models:** Match-LSTM[2], BiDAF[3], R-Net[4], QANet[5] …

- **Very impressive performance**

Reading the Web to Answer Questions?

Applying MRC to the Web

- Search engine is employed.

- Multiple passages are retrieved.

Applying MRC to the Web

- Search engine is employed.

- Multiple passages are retrieved.

- **All of them seem relevant.**

Applying MRC to the Web

- Search engine is employed.

- Multiple passages are retrieved.

- All of them seem relevant.

- **But they give different answers!**

Applying MRC to the Web

how many teams will be in the 2022 world cup?

2022 FIFA World Cup - Wikipedia
https://en.wikipedia.org/wiki/2022_FIFA_World_Cup ▾
In the end, there were five bids for the 2022 FIFA World Cup Australia, Japan, Qatar, South Korea and the United States.

Teams: 32 (from 5 or 6 confederations)   **Dates**: 21 November – 18 December
Host country: Qatar   **Venue(s)**: 8 or 12 (in 5 or 8 host cities)

2018 FIFA World Cup qualification - Wikipedia
https://en.wikipedia.org/wiki/2018_FIFA_World_Cup_qualification ▾
United Arab Emirates Ahmed Khalil (16 goals each). ← 2014 · 2022 →. The 2018 FIFA World Cup qualification process was a series of tournaments organised by the six FIFA confederations to decide 31 of the 32 teams which will play in the 2018 FIFA World Cup, with .... the suspension of their football association by FIFA on 30 May 2015.

Teams: 210 (from 6 confederations)   **Goals scored**: 2,454 (2.81 per match)
Matches played: 872

Qataris considering a 48-team option for 2022 World Cup - The ...
https://www.washingtonpost.com/...team...2022-world-cup/.../64dae0e6-8124-11e8-b3b...
5 days ago - The organizers of the 2022 World Cup in Qatar are open to talks about a ... power it apparently gained is key to any progress on the tournament expansion. ... 32 nations from which 16 winners would join 16 seeded teams in a ...

- Search engine is employed.

- Multiple passages are retrieved.

- All of them seem relevant.

- **But they give different answers!**

⇩

Key challenge :

Much more misleading candidates

An Example from MS-MARCO[6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

1) A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture. While the answer given is . . .

2) . . . The mixed economy is a balance between socialism and capitalism. As a result, some institutions are owned and maintained by . . .

3) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. Culture on the . . .

4) . . . A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species. A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. A pure culture is a culture consisting of only one strain. You can obtain a pure culture by picking . . .

6) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. A pure culture . . .

An Example from MS-MARCO [6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

■ Correct

1) A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture. While the answer given is . . .

2) . . . The mixed economy is a balance between socialism and capitalism. As a result, some institutions are owned and maintained by . . .

3) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** Culture on the . . .

4) . . . A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species. A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. A pure culture is a culture consisting of only one strain. You can obtain a pure culture by picking . . .

6) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** A pure culture . . .

An Example from MS-MARCO [6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

■ **Partially Correct**

1) A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture. While the answer given is . . .

2) . . . The mixed economy is a balance between socialism and capitalism. As a result, some institutions are owned and maintained by . . .

3) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. Culture on the . . .

4) . . . **A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species.** A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. **A pure culture is a culture consisting of only one strain.** You can obtain a pure culture by picking . . .

6) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. A pure culture . . .

An Example from MS-MARCO [6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

■ **Incorrect**

1) **A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture.** While the answer given is . . .

2) . . . **The mixed economy is a balance between socialism and capitalism.** As a result, some institutions are owned and maintained by . . .

3) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. Culture on the . . .

4) . . . A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species. A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. A pure culture is a culture consisting of only one strain. You can obtain a pure culture by picking . . .

6) A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies. A pure culture . . .

An Example from MS-MARCO [6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

■ Incorrect   ■ Partially Correct   ■ Correct

1) **A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture.** While the answer given is . . .

2) . . . **The mixed economy is a balance between socialism and capitalism.** As a result, some institutions are owned and maintained by . . .

Different

3) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** Culture on the . . .

4) . . . **A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species.** A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. **A pure culture is a culture consisting of only one strain.** You can obtain a pure culture by picking . . .

6) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** A pure culture . . .

Similar or same

An Example from MS-MARCO [6] Dataset

Question: What is the difference between a mixed and pure culture?

Passages:

■ Incorrect   ■ Partially Correct   ■ Correct

1) A culture is a society's total way of living and a society is a group that live in a defined territory and participate in common culture. While the answer given is . . .

2) . . . The mixed economy is a balance between socialism and capitalism. As a result, some institutions are owned and maintained by . . .

} Different

3) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** Culture on the . . .

4) . . . **A pure culture comprises a single species or strains. A mixed culture is taken from a source and may contain multiple strains or species.** A contaminated . . .

5) . . . It will be at that time when we can truly obtain a pure culture. **A pure culture is a culture consisting of only one strain.** You can obtain a pure culture by picking . . .

6) **A pure culture is one in which only one kind of microbial species is found whereas in mixed culture two or more microbial species formed colonies.** A pure culture . . .

Verify

√ → Correct Answer

Overview of Our Model
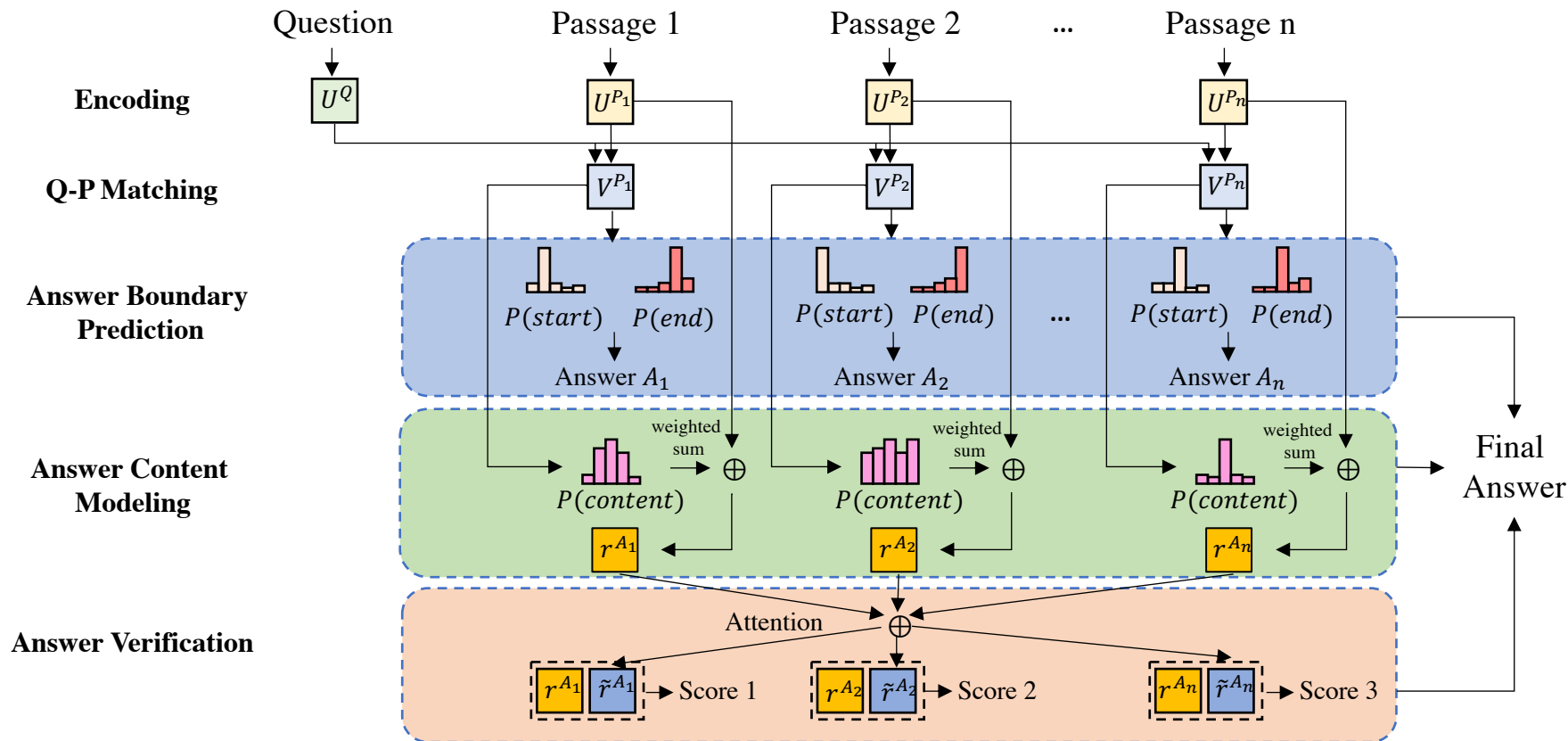
Overview of Our Model
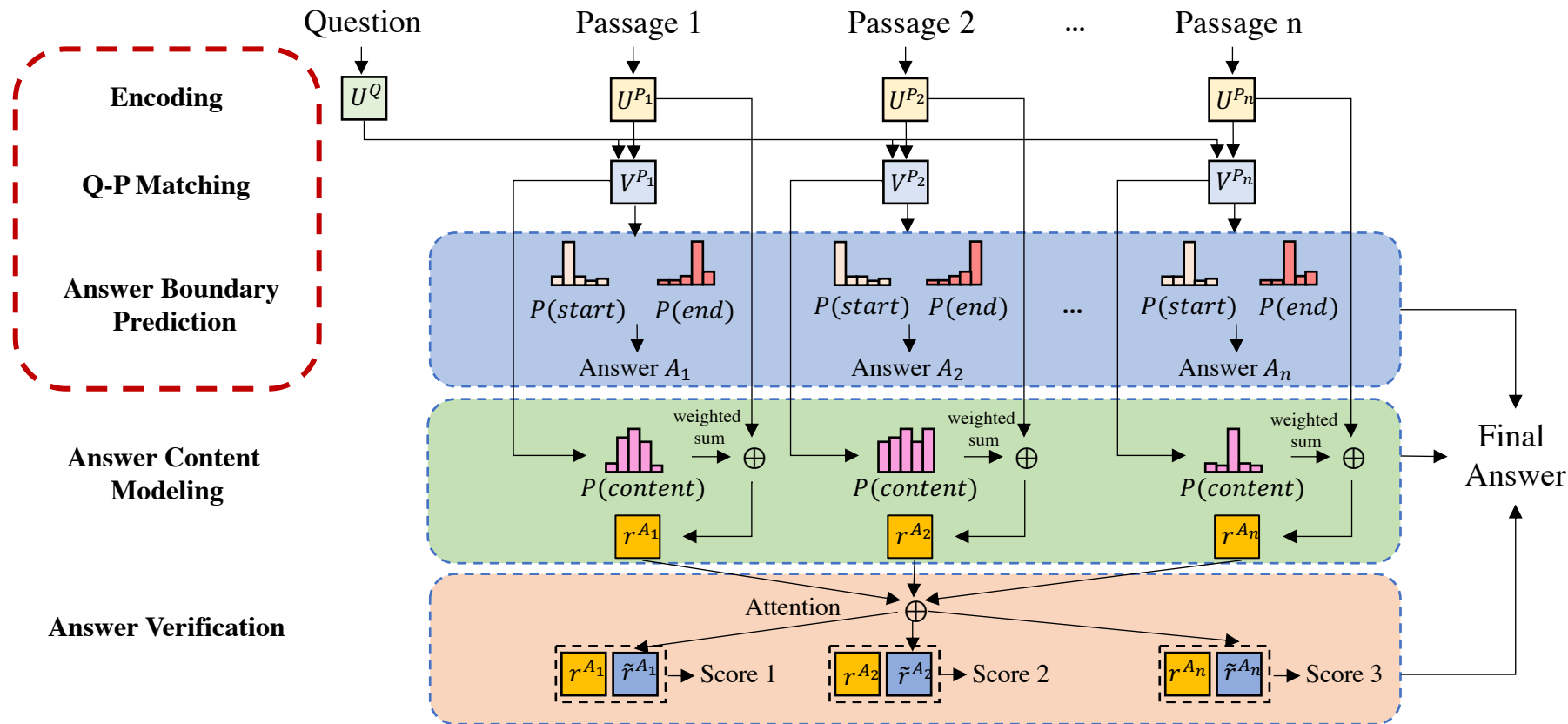
Overview of Our Model
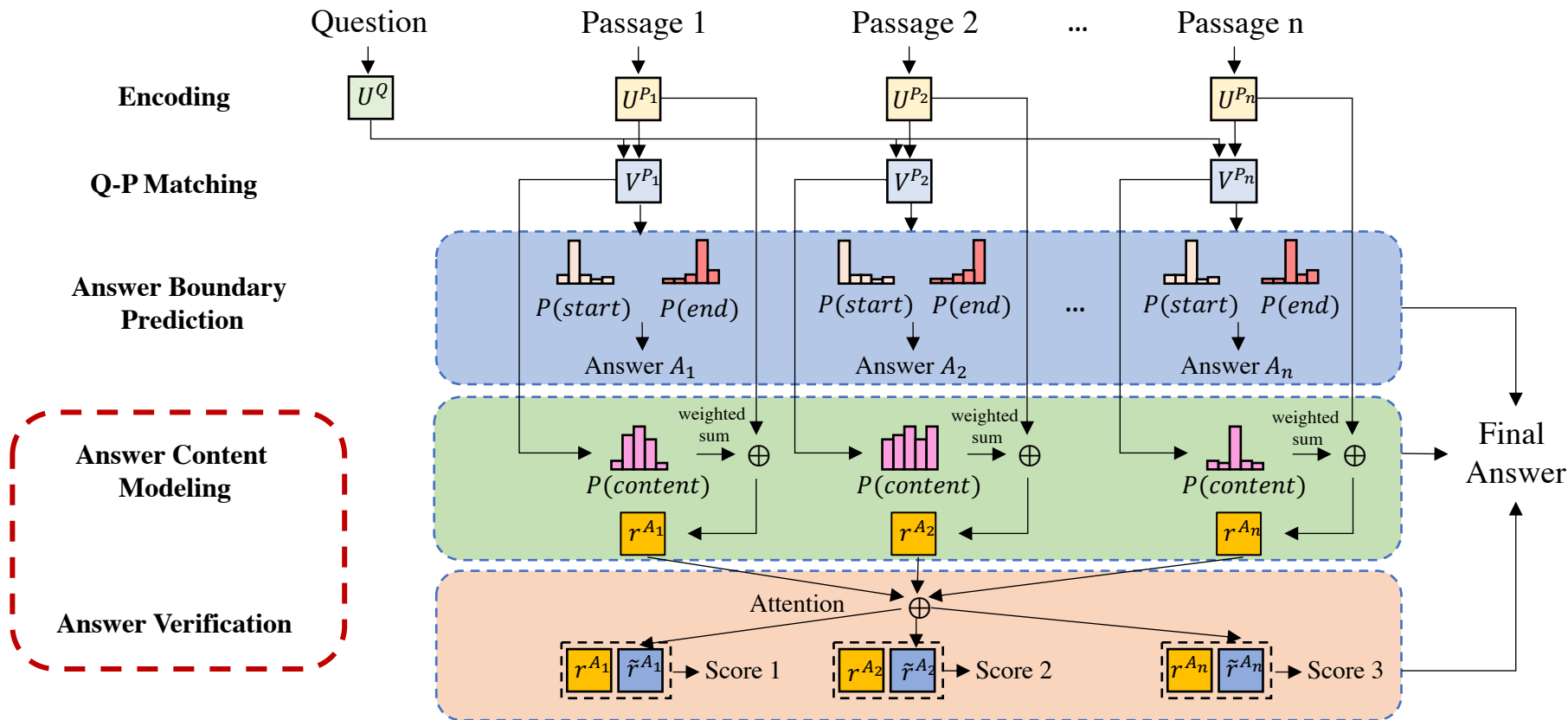
Overview of Our Model
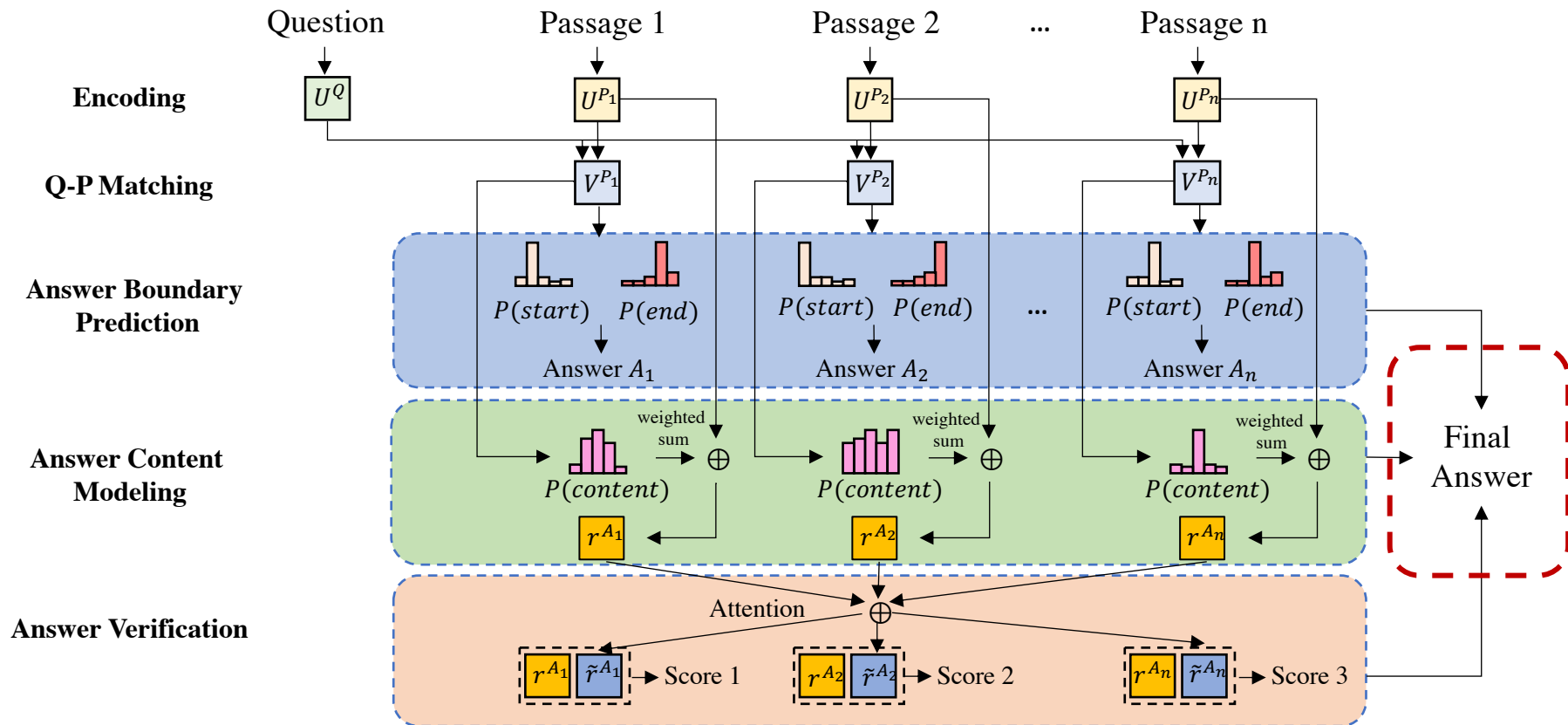
Input

Question   Passage 1   Passage 2   ...   Passage n

Question and Passage Encoding

Question   Passage 1   Passage 2   ...   Passage n

U^Q U^{P_1} U^{P_2} U^{P_n}
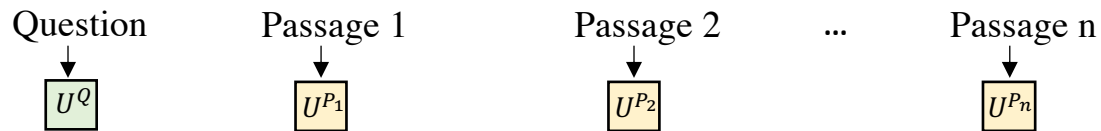
- Encoding with Bi-LSTM:

$$\mathbf{u}_t^Q = \mathrm{BiLSTM}_Q(\mathbf{u}_{t-1}^Q, [\mathbf{e}_t^Q, \mathbf{c}_t^Q])$$

$$\mathbf{u}_t^{P_i} = \mathrm{BiLSTM}_P(\mathbf{u}_{t-1}^{P_i}, [\mathbf{e}_t^{P_i}, \mathbf{c}_t^{P_i}])$$

Question-Passage Matching
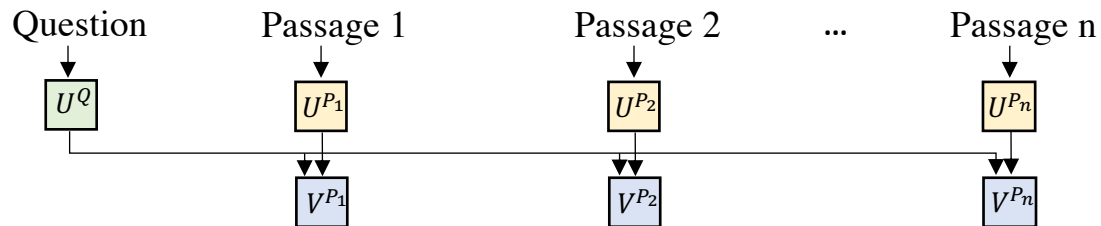
- Bi-directional Attention Flow
  (Seo et al., 2016)

- Dot attention matrix:

$$\mathbf{S}_{t,k} = \mathbf{u}_t^{Q\mathsf{T}} \cdot \mathbf{u}_k^{P_i}$$

Answer Boundary Prediction
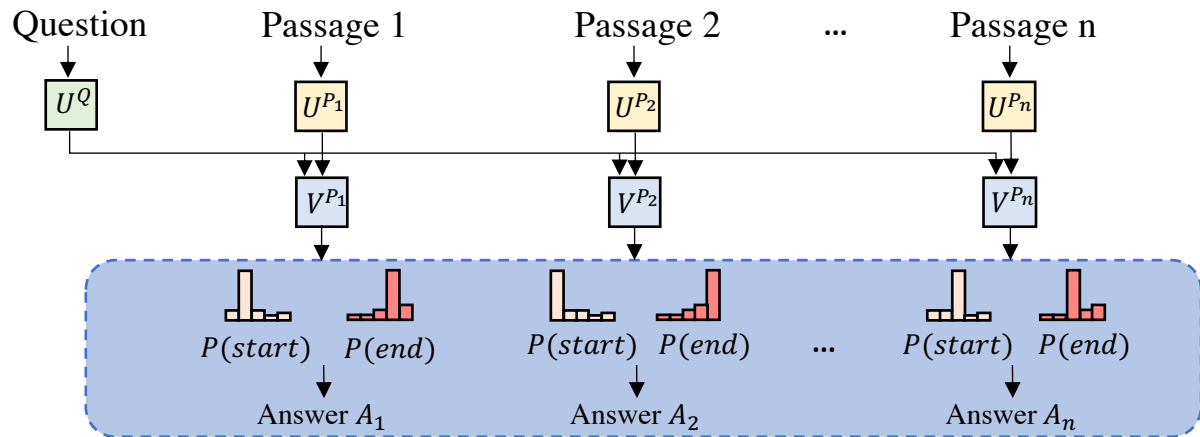
- Start and end pointer:

$$g_k^t = \mathbf{w}_1^{a\top} \tanh(\mathbf{W}_2^a [\mathbf{v}_k^P, \mathbf{h}_{t-1}^a])$$

$$\alpha_k^t = \exp(g_k^t) / \sum\nolimits_{j=1}^{|\mathbf{P}|} \exp(g_j^t)$$

$$\mathbf{c}_t = \sum\nolimits_{k=1}^{|\mathbf{P}|} \alpha_k^t \mathbf{v}_k^P$$

$$\mathbf{h}_t^a = \mathrm{LSTM}(\mathbf{h}_{t-1}^a, \mathbf{c}_t)$$

Answer Content Modeling
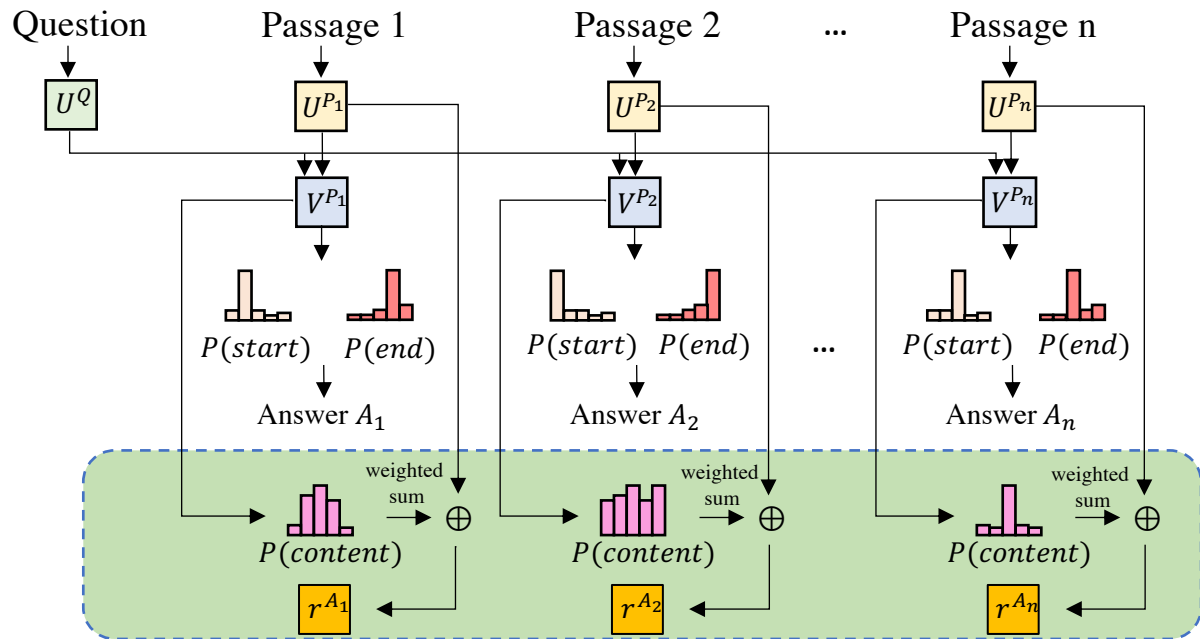
- Content score for each word:

$$p_k^c = \text{sigmoid}(\mathbf{w}_1^{c\top}\text{ReLU}(\mathbf{W}_2^c\mathbf{v}_k^{P_i}))$$

- Representation for A^i:

$$\mathbf{r}^{A_i} = \frac{1}{|\mathbf{P}_i|}\sum_{k=1}^{|\mathbf{P}_i|} p_k^c[\mathbf{e}_k^{P_i}, \mathbf{c}_k^{P_i}]$$

Cross-Passage Answer Verification
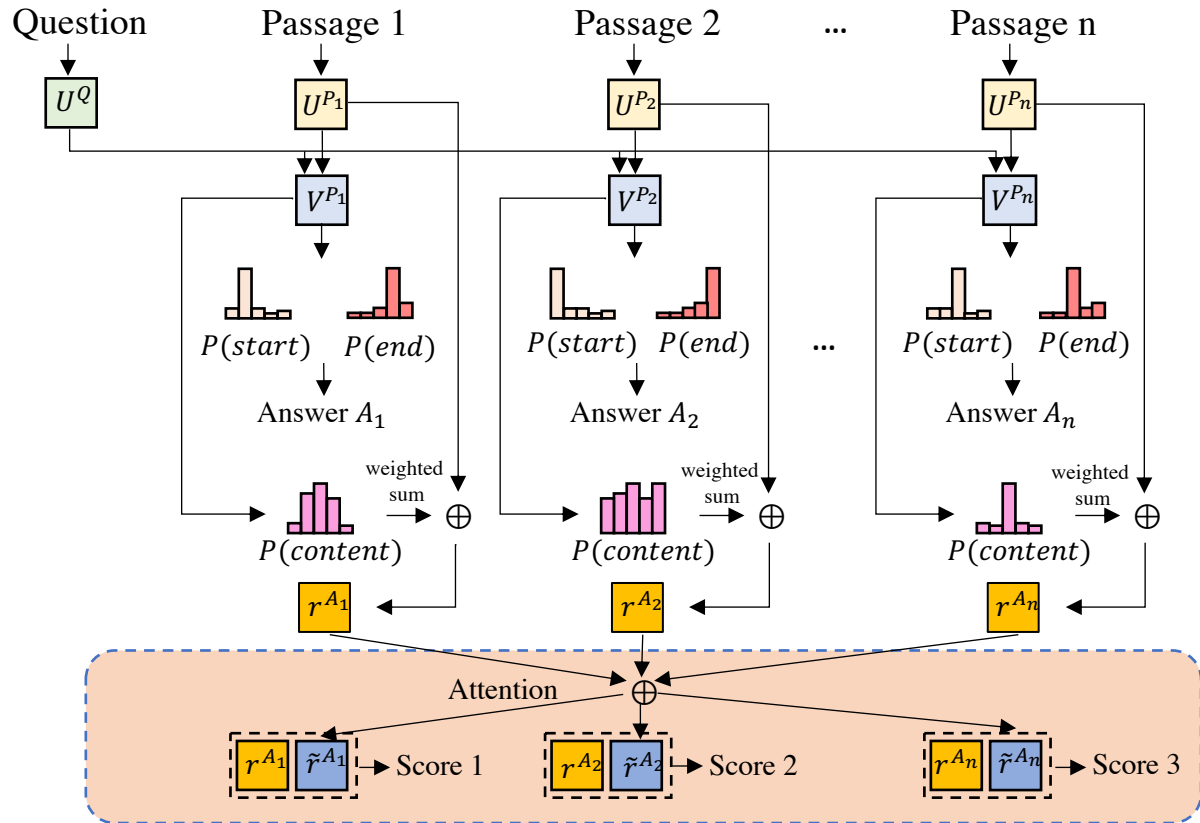
- Ans-to-ans Attention:

$$s_{i,j} = \begin{cases} 0, & \text{if } i = j, \\ \mathbf{r}^{A_i \mathsf{T}} \cdot \mathbf{r}^{A_j}, & \text{otherwise} \end{cases}$$

$$\alpha_{i,j} = \exp(s_{i,j}) / \sum_{k=1}^{n} \exp(s_{i,k})$$

$$\tilde{\mathbf{r}}^{A_i} = \sum_{j=1}^{n} \alpha_{i,j} \mathbf{r}^{A_j}$$

- Verification score:

$$g_i^v = \mathbf{w}^{v\mathsf{T}} [\mathbf{r}^{A_i}, \tilde{\mathbf{r}}^{A_i}, \mathbf{r}^{A_i} \odot \tilde{\mathbf{r}}^{A_i}]$$

$$p_i^v = \exp(g_i^v) / \sum_{j=1}^{n} \exp(g_j^v)$$

Joint Training and Prediction

- **Three objectives:**

  - Finding the boundary of the answer

  - Predicting whether each word should be included in the answer

  - Selecting the best answer from all the candidates

- **Training Loss:**

$$\mathcal{L}_{\text{joint}} = \mathcal{L}_{boundary} + \beta_1 \mathcal{L}_{content} + \beta_2 \mathcal{L}_{verify}$$

- **Prediction:**

$$\text{Score} = S_{boundary} \times S_{content} \times S_{verify}$$

Experiments Setup

- Datasets: MS-MARCO[6] and DuReader[7]:

	Language	Search Engine	Size	Questions with Multi Annotated Answers	Questions with Multi Answer Spans
MS-MARCO	English	Bing	100K+	9.93%	40.00%
DuReader	Chinese	Baidu	200K+	67.28%	56.38%

Experiments Setup

- Datasets: MS-MARCO[6] and DuReader[7]:

	Language	Search Engine	Size	Questions with Multi Annotated Answers	Questions with Multi Answer Spans
MS-MARCO	English	Bing	100K+	9.93%	40.00%
DuReader	Chinese	Baidu	200K+	67.28%	56.38%

Experiments Setup

- Datasets: MS-MARCO[6] and DuReader[7]:

	Language	Search Engine	Size	Questions with Multi Annotated Answers	Questions with Multi Answer Spans
MS-MARCO	English	Bing	100K+	9.93%	40.00%
DuReader	Chinese	Baidu	200K+	67.28%	56.38%

- Hyper-parameters (tuned on the dev set):

Word Embedding	Character Embedding	Hidden Size	L2	Optimizer	Learning Rate	Batch Size	β_1	β_2
300-D Glove	30-D Random	150	3e-4	Adam	4e-4	32	0.5	0.5

Main Results

Model	ROUGE-L	BLEU-1
FastQA_Ext	33.67	33.93
Match-LSTM	37.33	40.72
ReasoNet	38.81	39.86
R-Net	42.89	42.22
S-Net	45.23	43.78
Our Model	**46.15**	**44.47**
S-Net (Ensemble)	46.65	44.78
Our Model (Ensemble)	**46.66**	**45.41**
Human	47	46

Tab 1. Performance on MS-MARCO test set

Model	ROUGE-L	BLEU-4
Match-LSTM	39.0	31.8
BiDAF	39.2	31.9
PR+BiDAF	41.8	37.6
Our Model	**44.2**	**41.0**
Human	57.4	56.1

Tab 2. Performance on DuReader test set

Ablation Study on MS-MARCO Dev Set

Model	ROUGE-L	Δ
Complete Model	**45.65**	**-**
- Answer Verification	44.38	-1.27
- Content Modeling	44.27	-1.38
- Joint Training	44.12	-1.53
-Yes/No Classification	41.87	-3.78
Boundary Baseline	38.95	-6.70

Quantitative Analysis: the Predicted Scores

Question: What is the difference between a mixed and pure culture	Scores		
Answer Candidates:	**Boundary**	**Content**	**Verification**
[1] A culture is a society's total way of living and a society is a group …	$\mathbf{1.0 \times 10^{-2}}$	$\mathbf{1.0 \times 10^{-1}}$	1.1×10^{-1}
[2] The mixed economy is a balance between socialism and capitalism.	1.0×10^{-4}	4.0×10^{-2}	3.2×10^{-2}
[3] A pure culture is one in which only one kind of microbial species is …	5.5×10^{-3}	7.7×10^{-2}	1.2×10^{-1}
[4] A pure culture comprises a single species or strains. A mixed …	2.7×10^{-3}	8.1×10^{-2}	1.3×10^{-1}
[5] A pure culture is a culture consisting of only one strain.	5.8×10^{-4}	7.9×10^{-2}	5.1×10^{-2}
[6] A pure culture is one in which only one kind of microbial species …	5.8×10^{-3}	9.1×10^{-2}	$\mathbf{2.7 \times 10^{-1}}$
……		……	

Quantitative Analysis: the Predicted Scores

Question: What is the difference between a mixed and pure culture	Scores		
Answer Candidates:	**Boundary**	**Content**	**Verification**
[1] A culture is a society's total way of living and a society is a group ...	1.0×10^{-2}	1.0×10^{-1}	1.1×10^{-1}
[2] The mixed economy is a balance between socialism and capitalism.	1.0×10^{-4}	4.0×10^{-2}	3.2×10^{-2}
[3] A pure culture is one in which only one kind of microbial species is ...	5.5×10^{-3}	7.7×10^{-2}	1.2×10^{-1}
[4] A pure culture comprises a single species or strains. A mixed ...	2.7×10^{-3}	8.1×10^{-2}	1.3×10^{-1}
[5] A pure culture is a culture consisting of only one strain.	5.8×10^{-4}	7.9×10^{-2}	5.1×10^{-2}
[6] A pure culture is one in which only one kind of microbial species ...	5.8×10^{-3}	9.1×10^{-2}	2.7×10^{-1}
......		.....	

Boundary / content / verification scores
are usually positively relevant

Quantitative Analysis: the Predicted Scores

Question: What is the difference between a mixed and pure culture	Scores		
Answer Candidates:	**Boundary**	**Content**	**Verification**
[1] A culture is a society's total way of living and a society is a group …	$\mathbf{1.0 \times 10^{-2}}$	$\mathbf{1.0 \times 10^{-1}}$	1.1×10^{-1}
[2] The mixed economy is a balance between socialism and capitalism.	1.0×10^{-4}	4.0×10^{-2}	3.2×10^{-2}
[3] A pure culture is one in which only one kind of microbial species is …	5.5×10^{-3}	7.7×10^{-2}	1.2×10^{-1}
[4] A pure culture comprises a single species or strains. A mixed …	2.7×10^{-3}	8.1×10^{-2}	1.3×10^{-1}
[5] A pure culture is a culture consisting of only one strain.	5.8×10^{-4}	7.9×10^{-2}	5.1×10^{-2}
[6] **A pure culture is one in which only one kind of microbial species …**	5.8×10^{-3}	9.1×10^{-2}	$\mathbf{2.7 \times 10^{-1}}$
……		……	

More commonality --> larger verification score

Quantitative Analysis: the Predicted Scores

| Question: What is the difference between a mixed and pure culture	Scores		
Answer Candidates:	Boundary	Content	Verification
[1] A culture is a society's total way of living and a society is a group …	1.0×10^{-2}	1.0×10^{-1}	1.1×10^{-1}
[2] The mixed economy is a balance between socialism and capitalism.	1.0×10^{-4}	4.0×10^{-2}	3.2×10^{-2}
[3] A pure culture is one in which only one kind of microbial species is …	5.5×10^{-3}	7.7×10^{-2}	1.2×10^{-1}
[4] A pure culture comprises a single species or strains. A mixed …	2.7×10^{-3}	8.1×10^{-2}	1.3×10^{-1}
[5] A pure culture is a culture consisting of only one strain.	5.8×10^{-4}	7.9×10^{-2}	5.1×10^{-2}
[6] A pure culture is one in which only one kind of microbial species …	5.8×10^{-3}	9.1×10^{-2}	2.7×10^{-1}
……		……	

Correct answer is selected by considering verification!

Necessity of the Content Model
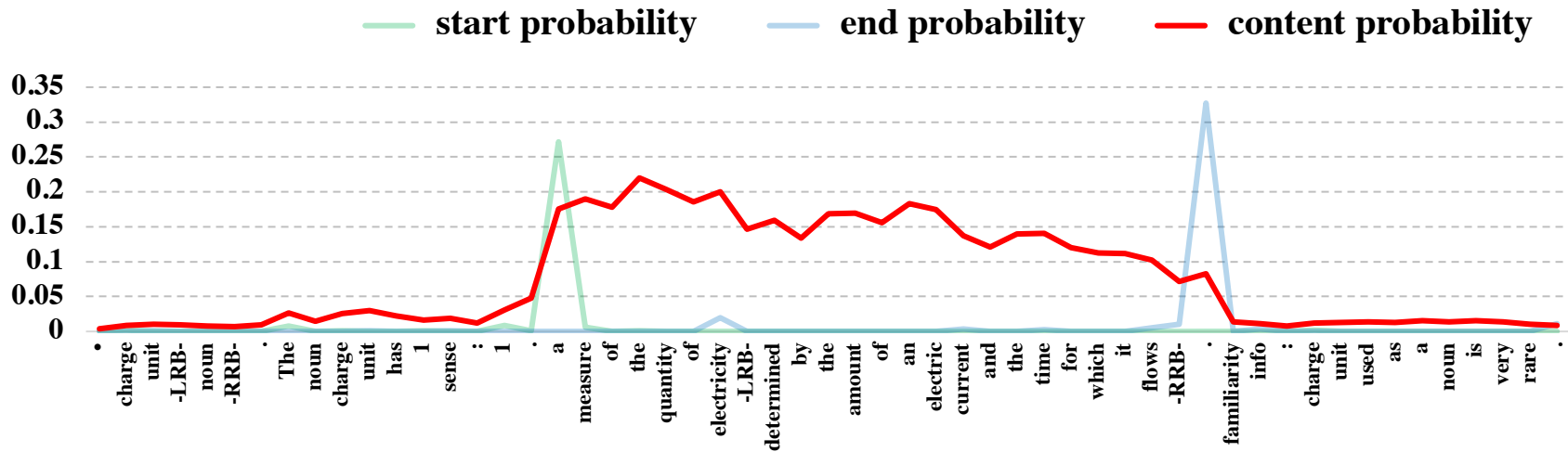
Necessity of the Content Model
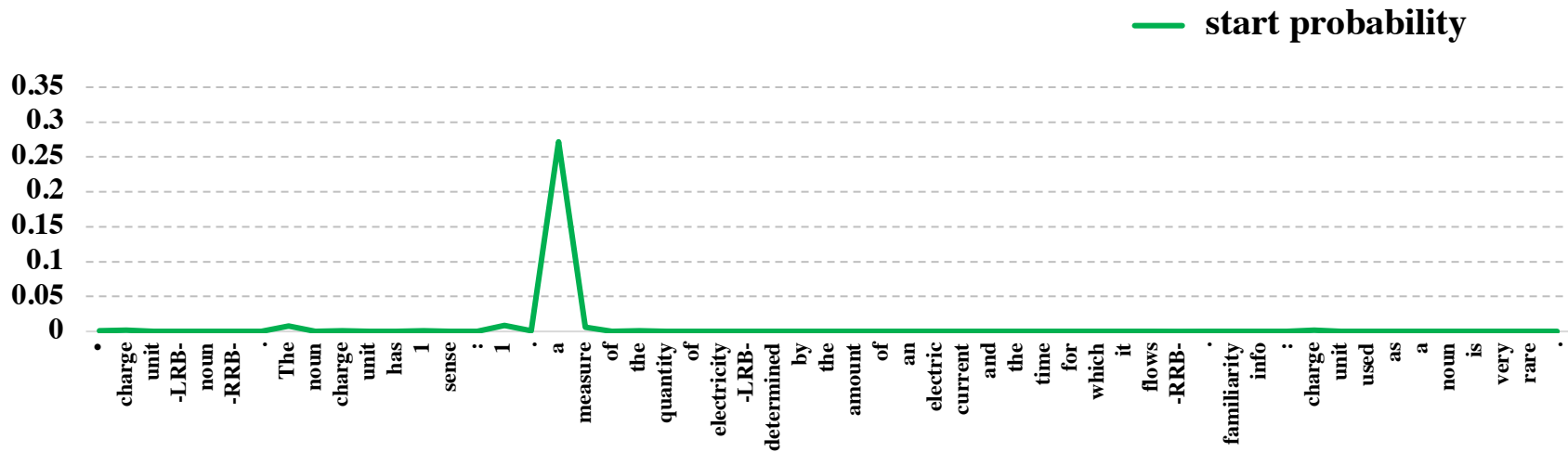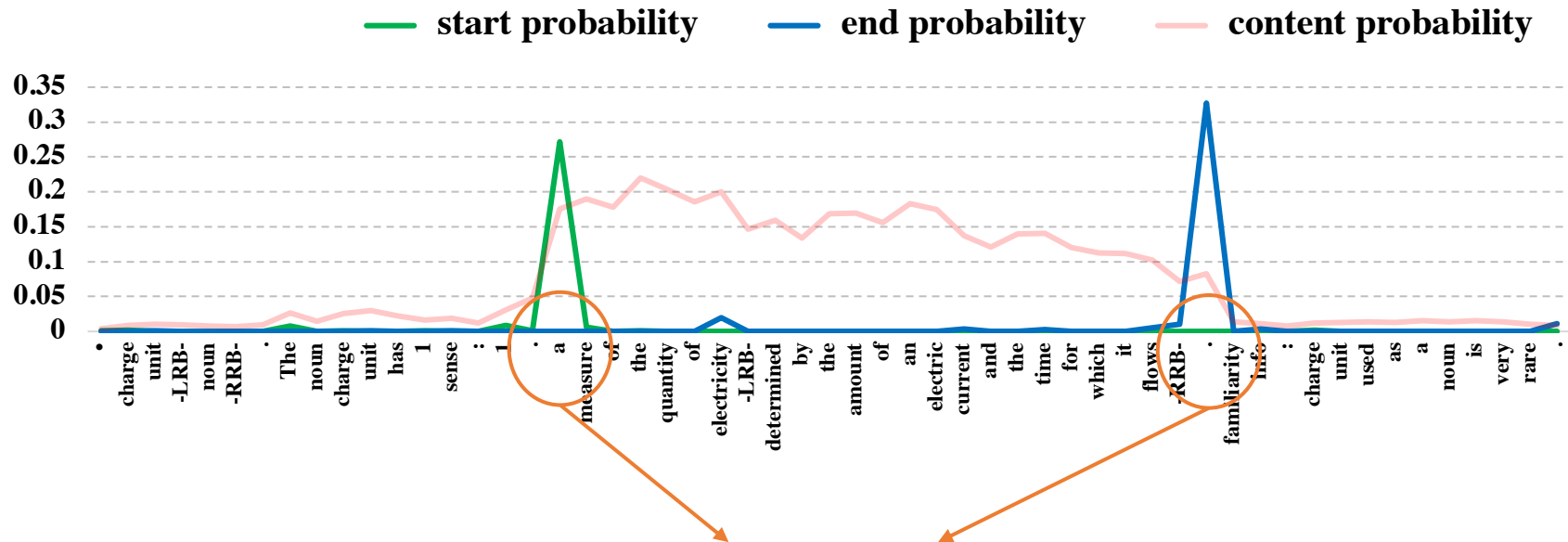
Necessity of the Content Model
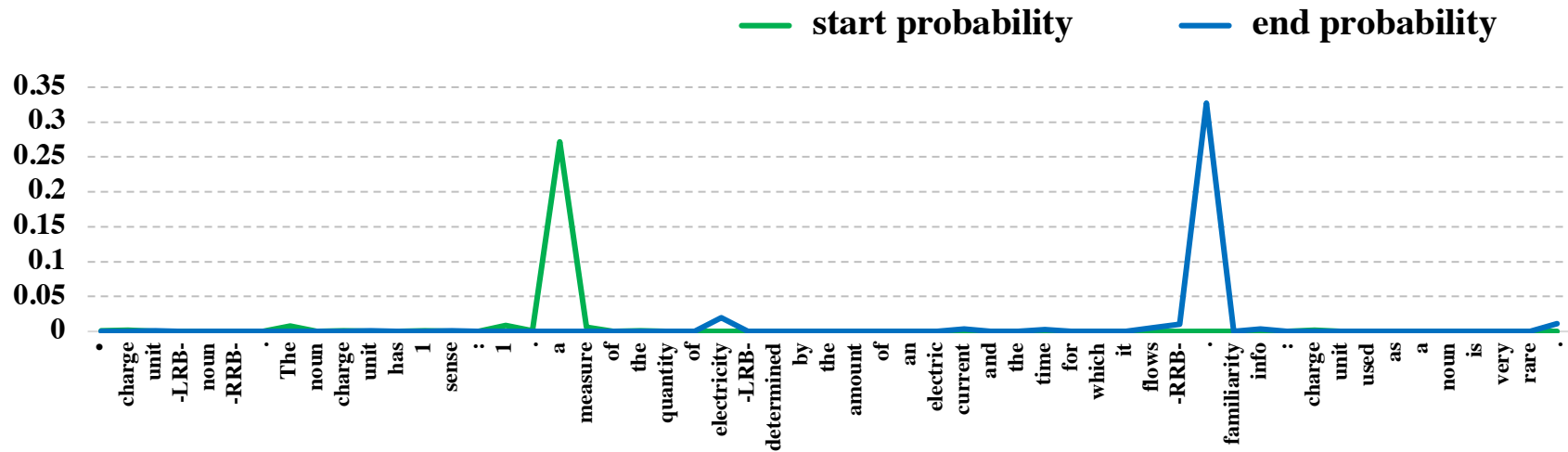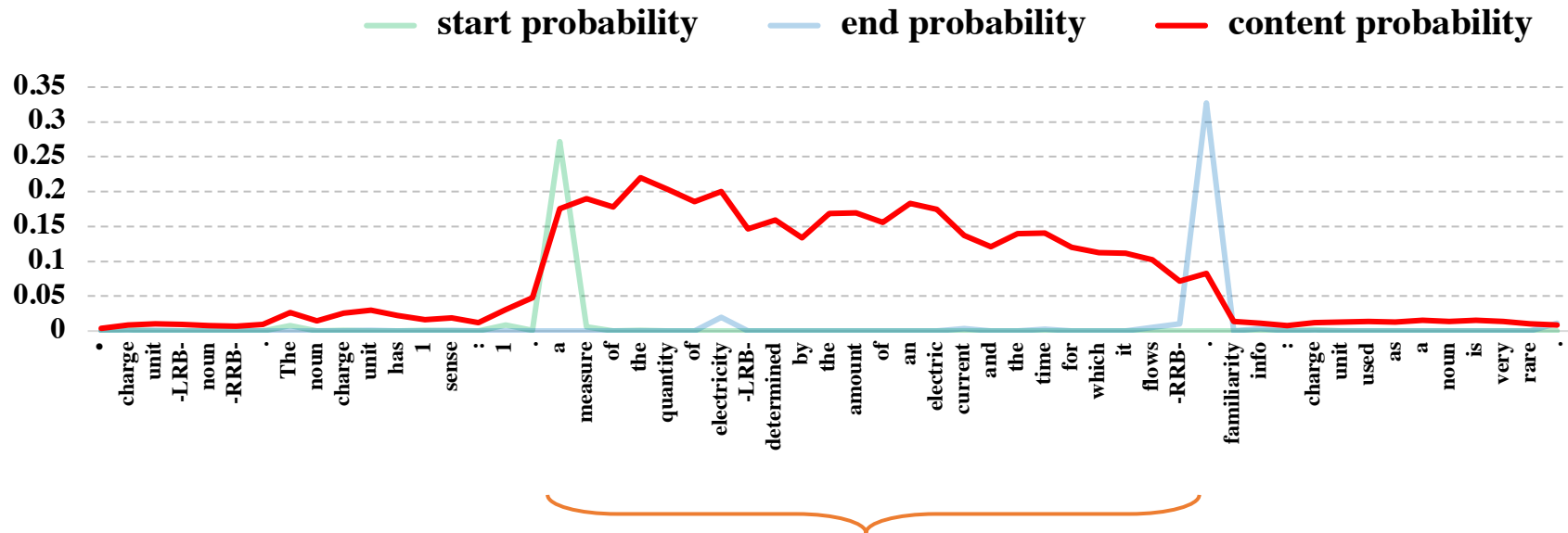
Visualization of the Probability Distribution

Necessity of the Content Model

When the answer is long, boundary words carry little information.

Necessity of the Content Model

Content words reflect the real semantics of this answer.

Conclusion

- **Multi-passage MRC:** much more misleading answers

- **End-to-end model for multi-passage MRC:**

  - Find the answer boundary

  - Model the answer content

  - Cross-passage answer verification

  - Joint training and prediction

- **SOTA performance on two datasets created from real-world web data:**

  - MS-MARCO (English)

  - DuReader (Chinese)

References

1) Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text.

2) Shuohang Wang and Jing Jiang. 2016. Machine comprehension using match-lstm and answer pointer.

3) Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2016. Bidirectional attention flow for machine comprehension.

4) Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. 2017. Gated self-matching net- works for reading comprehension and question answering.

5) Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. Qanet: Combining local convolution with global self-attention for reading comprehension.

6) Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A human generated machine reading comprehension dataset.

7) Wei He, Kai Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Xuan Liu, Tian Wu, and Haifeng Wang. 2017. Dureader: a chinese machine reading comprehen- sion dataset from real-world applications.

Thank you!

Q & A

Contact: yizhong@pku.edu.cn