

# Machine Reading Comprehension

Yizhong Wang

2017.5.29

# Outline

1. Overview of Reading Comprehension
2. Relevant Tasks and Datasets
3. Key Techniques and Cutting-Edge Models
4. Conclusion: Challenges and Directions

# What is Reading Comprehension ?



the community to our exhibit. Our parents are invited to go with us. I have three paintings I am working on now. I hope one of them will be chosen to be in the exhibit. I like having a goal to work toward.

## ANSWER THE QUESTIONS:

1. What is one of the favorite classes of this student?

- a. math
- b. reading
- c. art
- d. music

2. The author describes Mrs. Hilbert as a \_\_\_\_\_.

- a. nice lady
- b. wonderful artist
- c. sweet person
- d. beautiful person

3. What did Mrs. Hilbert teach the class to do?

- a. bake chocolate cake
- b. write great stories
- c. mix paint
- d. do long division

4. Why does it help this student to see the works of other artists?

- a. to learn to draw
- b. mix paint
- c. to think about what she wants to paint
- d. none of these

5. Where will the art exhibit be?

- a. at the library
- b. at one of the banks
- c. at the school
- d. at one of the churches

6. What does this student hope will happen with her paintings?

- a. one of them will be sold
- b. one of them will be chosen for the exhibit
- c. one of them will be of her younger sister
- d. one of them will be lost

Given a **passage**, you are supposed to answer several **questions** according to the passage.

$$\operatorname{argmax}_a P(a | c, q)$$

c : a context document

q : a query relating to that document

a : the answer to that query

# Why is it important?

What did Zuckerberg talk at Harvard recently

All News Videos Images Shopping More Settings Tools

About 2,780,000 results (1.00 seconds)

Top stories



 Mark Zuckerberg's speech as written for Harvard's Class of 2017  Harvard Gazette · 1 day ago	 Facebook's Zuckerberg talks automation and immigration in Harvard commencement speech  CNNMoney · 1 day ago	 Make a difference, Zuckerberg tells Harvard graduates  Harvard Gazette · 1 day ago
--	--	--

→ More for What did Zuckerberg talk at Harvard recently

Mark Zuckerberg's speech as written for Harvard's Class of 2017 ...

[news.harvard.edu/.../mark-zuckerbergs-speech-as-written-for-harvards-class-of-2017/](http://news.harvard.edu/.../mark-zuckerbergs-speech-as-written-for-harvards-class-of-2017/) ▾

2 days ago - **Mark Zuckerberg**'s Commencement address at **Harvard** ... I couldn't figure out why no one would **talk** to me — except one guy, KX Jin, he just ...

Make a difference, Zuckerberg tells Harvard graduates | Harvard Gazette

[news.harvard.edu/gazette/.../make-a-difference-zuckerberg-tells-harvard-graduates/](http://news.harvard.edu/gazette/.../make-a-difference-zuckerberg-tells-harvard-graduates/) ▾

2 days ago - "Today I want to **talk** about purpose," he said. "But I'm not here to ... **Mark Zuckerberg** Commencement Address | Harvard Commencement 2017 ...

- New form of Search
- Can AI give us a definite answer just as like we are chatting?
- An AI-complete Problem

# Current Task Types

According to the type of the answer, tasks can be categorized into the following:

- Document-based QA – selecting one sentence
- Cloze test – selecting one word or entity (No explicit question)
- Selecting a span of continuous text
- Generating answers based on supporting information
- Choice question – selecting answer from multiple candidate answers

# Dataset 1: CNN and Daily Mail

- Passages are news articles from CNN and Daily Mail websites;
- Questions are constructed by turning bullet points into Cloze style questions by replacing one entity with placeholder;
- Answer is the missing entity.

## Passage

( @entity4 ) if you feel a ripple in the force today , it may be the news that the official @entity6 is getting its first gay character . according to the sci-fi website @entity9 , the upcoming novel " @entity11 " will feature a capable but flawed @entity13 official named @entity14 who " also happens to be a lesbian . " the character is the first gay figure in the official @entity6 -- the movies , television shows , comics and books approved by @entity6 franchise owner @entity22 -- according to @entity24 , editor of " @entity6 " books at @entity28 imprint @entity26 .

## Question

characters in " @placeholder " movies have gradually become more diverse

## Answer

@entity6

# Dataset 2: SQuAD

- Passages are paragraphs from Wikipedia;
- Questions are posed by paid crowdworkers;
- The answer to every question is a segment of text, or span, from the corresponding reading passage.

Oxygen is a chemical element with symbol O and atomic number 8. It is a member of the chalcogen group on the periodic table and is a highly reactive nonmetal and oxidizing agent that readily forms compounds (notably oxides) with most elements. By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O<sub>2</sub>. Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere. However, monitoring of atmospheric oxygen levels show a global downward trend, because of fossil-fuel burning. Oxygen is the most abundant element by mass in the Earth's crust as part of oxide compounds such as silicon dioxide, making up almost half of the crust's mass.

What are the three most abundant elements of the universe by mass?

Ground Truth Answers: By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium | oxygen is the third-most abundant element in the universe, after hydrogen and helium | oxygen is the third-most abundant element in the universe, after hydrogen and helium | oxygen is the third-most abundant element in the universe, after hydrogen and helium | By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium.

Prediction: oxygen

What is the atomic number of the element oxygen?

Ground Truth Answers: 8 8 8 8 8

Prediction: 8

# Dataset 3: MS MARCO

- Questions are real queries issued through Bing or Cortana;
- Documents are related web pages;
- Answers are human-generated with the supporting information only from the relevant passages.

Q: will i qualify for osap if i'm new in canada

**Candidate passages**

Click passages to select or unselect them  
<http://www.settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/how-do-i-apply-for-the-ontario-student-assistance-program-osap/>

Ontario.ca Francais. Francais in order to apply online for funding consideration from The Ontario Student Assistance (PROGRAM), osap you must first register as a new user to this, website  
Source: <https://osap.gov.on.ca/OSAPSecurityWeb/public/agreement.xhtml>

Visit the OSAP website for application deadlines. To get OSAP, you have to be eligible. You can apply using an online form, or you can print off the application forms. If you submit a paper application, you must pay an application fee. The online application is free.  
Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/how-do-i-apply-for-the-ontario-student-assistance-program-osap/>

Visit the OSAP website for application deadlines. To get OSAP, you have to be eligible. You can apply using an online form, or you can print off the application forms. If you submit a paper application, you must pay an application fee.  
Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/how-do-i-apply-for-the-ontario-student-assistance-program-osap/>

To be eligible to apply for financial assistance from the Ontario Student Assistance Program (OSAP), you must be a: 1 Canadian citizen; 2 Permanent resident; or. 3 Protected person/convention refugee with a Protected Persons Status Document (PPSD).  
Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/who-is-eligible-for-the-ontario-student-assistance-program-osap/>

You will not be eligible for a Canada-Ontario Integrated Student Loan, but can apply for a part-time loan through the Canada Student Loans Program. There are also grants, bursaries and scholarships available for both full-time and part-time students.  
Source: <http://www.campusaccess.com/financial-aid/osap.html>

Selected passages

Visit the OSAP website for application deadlines. To get OSAP, you have to be eligible. You can apply using an online form, or you can print off the application forms. If you submit a paper application, you must pay an application fee. The online application is free.  
Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/how-do-i-apply-for-the-ontario-student-assistance-program-osap/>

To be eligible to apply for financial assistance from the Ontario Student Assistance Program (OSAP), you must be a: 1 Canadian citizen; 2 Permanent resident; or. 3 Protected person/convention refugee with a Protected Persons Status Document (PPSD).  
Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/who-is-eligible-for-the-ontario-student-assistance-program-osap/>

You will not be eligible for a Canada-Ontario Integrated Student Loan, but can apply for a part-time loan through the Canada Student Loans Program. There are also grants, bursaries and scholarships available for both full-time and part-time students.  
Source: <http://www.campusaccess.com/financial-aid/osap.html>

Summarize the answer given by the selected passages:  
No, You won't qualify.

Submit answer      Can't summarize

# Dataset 4: RACE

- Questions, answers and associated passages are collected from the English exams for middle- and high-school Chinese students.
- Both the questions and candidate answers in RACE are designed by experts and are not restricted to be the text spans in the original passage.

**Passage:**

In a small village in England about 150 years ago, a mail coach was standing on the street. It didn't come to that village often. People had to pay a lot to get a letter. The person who sent the letter didn't have to pay the postage, while the receiver had to. "Here's a letter for Miss Alice Brown," said the mailman.  
"I'm Alice Brown," a girl of about 18 said in a low voice.

**Questions:**

- 1): The first postage stamp was made \_ .  
A. in England B. in America C. by Alice D. in 1910
- 2): The girl handed the letter back to the mailman because \_ .  
A. she didn't know whose letter it was  
B. she had no money to pay the postage  
C. she received the letter but she didn't want to open it  
D. she had already known what was written in the letter
- 3): We can know from Alice's words that \_ .  
A. Tom had told her what the signs meant before leaving  
B. Alice was clever and could guess the meaning of the signs  
C. Alice had put the signs on the envelope herself  
D. Tom had put the signs as Alice had told him to

- 4): The idea of using stamps was thought of by \_ .  
A. the government  
B. Sir Rowland Hill  
C. Alice Brown  
D. Tom
- 5): From the passage we know the high postage made \_ .  
A. people never send each other letters  
B. lovers almost lose every touch with each other  
C. people try their best to avoid paying it  
D. receivers refuse to pay the coming letters

**Answer:** ADABC

# Dataset Summary

Dataset	Query Type	Answer Type	# Queries	# Documents	State-of-the-art
1. CNN / DM	Cloze	Fill in entity	1.4M	93K (CNN), 220K (DM)	Acc 77.6 (CNN) 79.2 (DM) <b>Reached Ceiling!</b>
2. SQuAD	Crowd-sourced	Span of words	100K	536 Articles, 23K Paragraphs	EM 76.9 / F1 84.0 Human:82.3 / 91. 2
3. MS MARCO	User logs	Human generated	100K	1M Passages	Rouge-L 42.89 Human: 47
4. RACE	Designed by experts	Choices (Designed by experts)	97K+	27K+ Passages	Acc 43.3 Human: 93.3

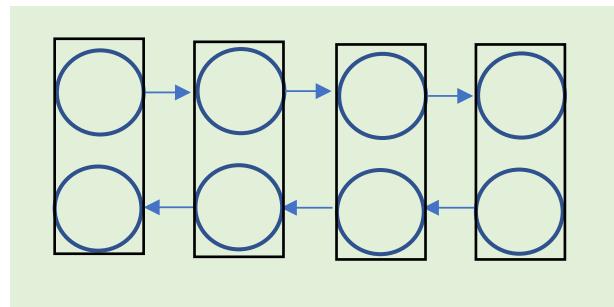
- 1/2/4 are classification tasks.
- 3 is a generative task, but the best current model is a classification model.

# Key Techniques

Task: Selecting answer from the given passage to answer the question.

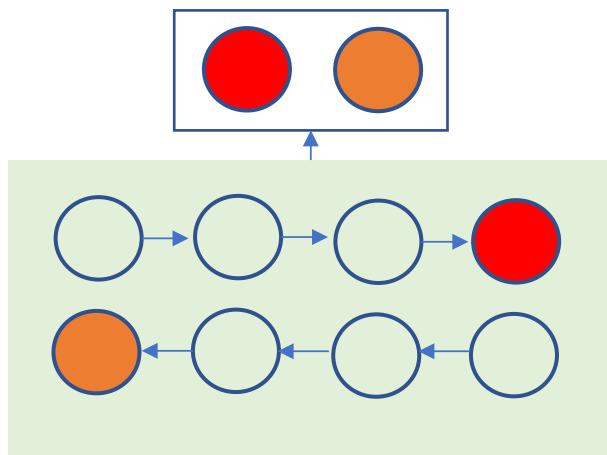
- Question and Passage Encoding
- Attention:
  - Question-aware passage encoding
  - Passage-aware question encoding
  - Self attention
  - Multi-hop attention
- Pointer Network - Select the best answer

# Question and Passage Encoding



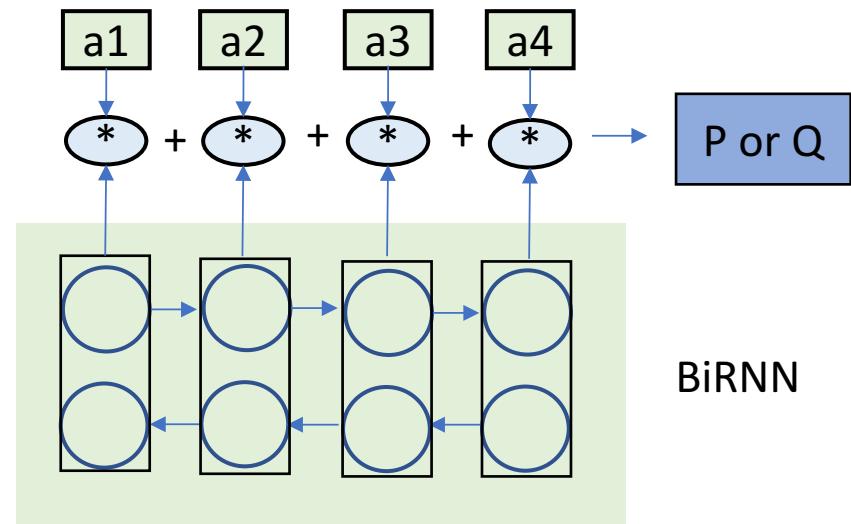
BiRNN

Passage or Question



BiRNN

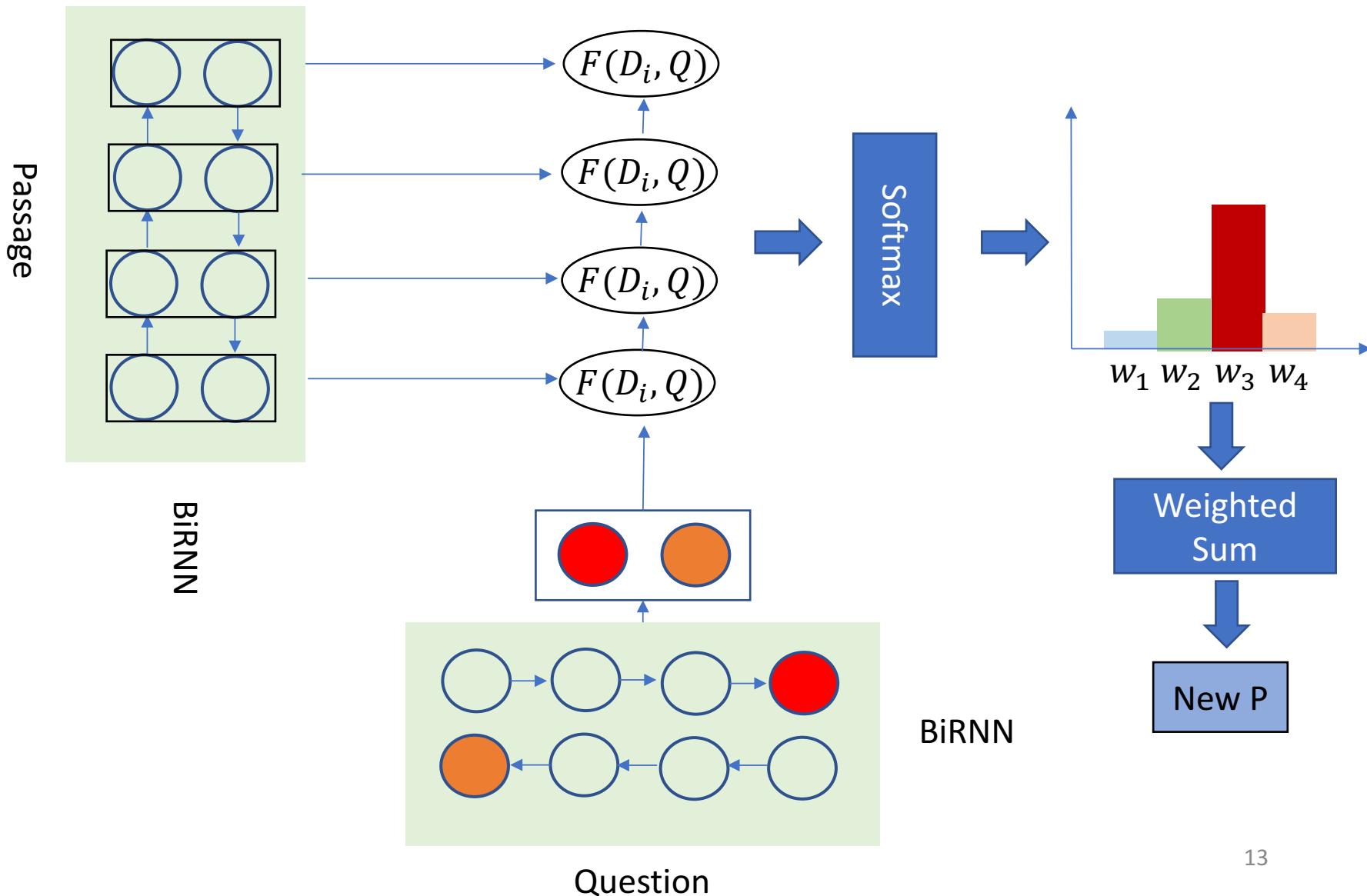
Passage or Question



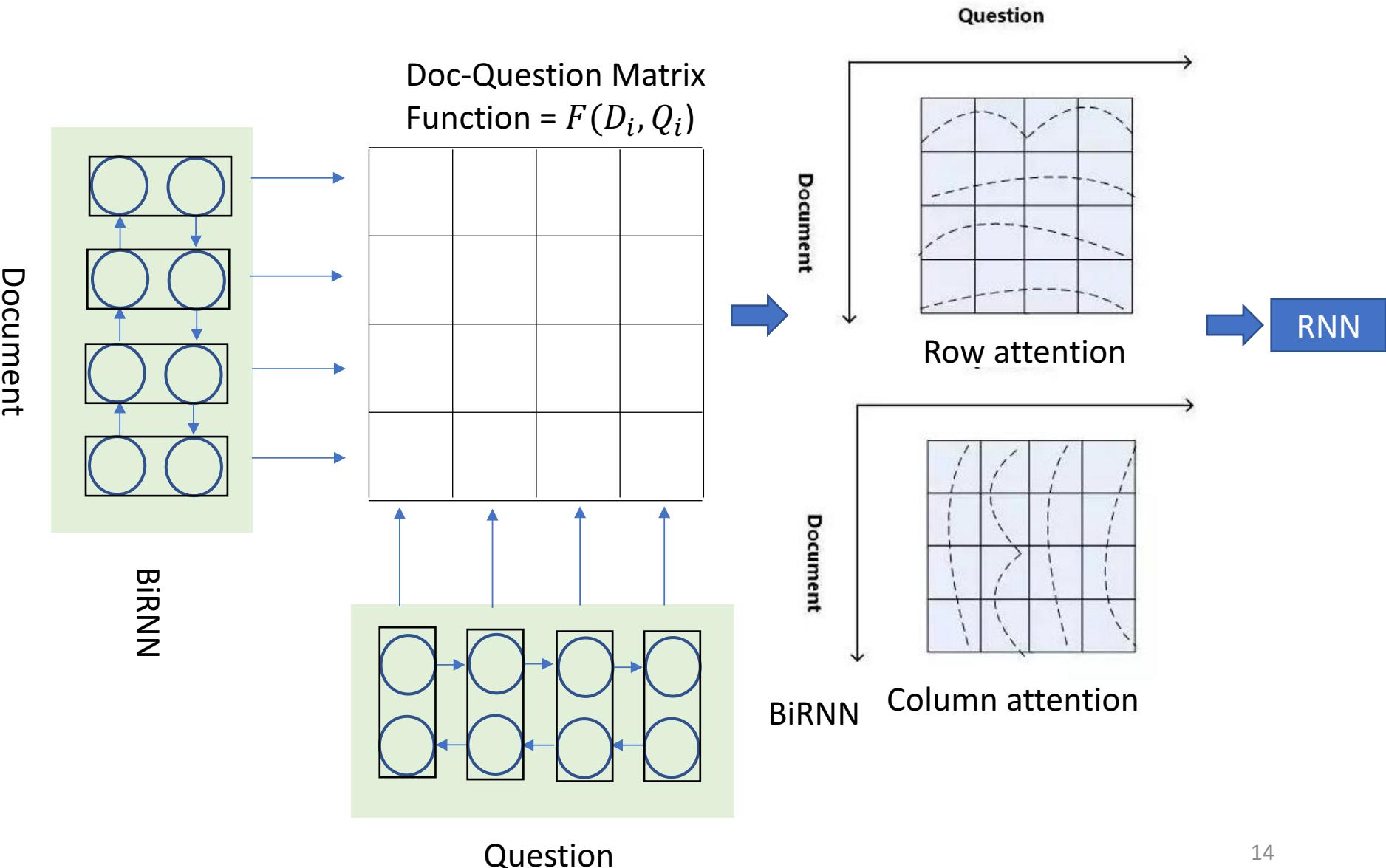
BiRNN

Passage or Question

# Attention : One Dimensional



# Attention: Two Dimensional



# Attention: Self Attention

- Question-aware passage representation has very limited knowledge of its context.
- Passage context is necessary to infer the answer.

Sam walks into the kitchen.

Sam picks up an apple.

Sam walks into the bedroom.

Sam drops the apple.

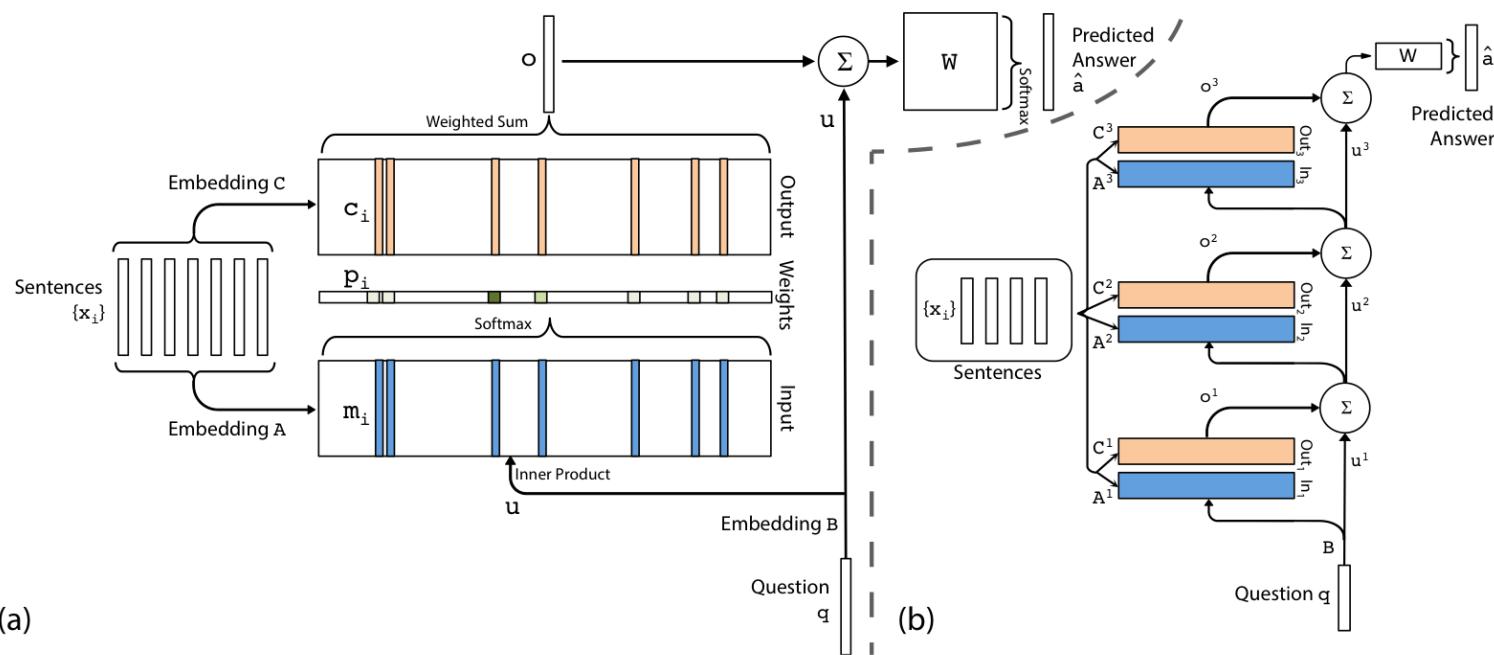
Q: Where is the apple?

A. Bedroom

- We could make the passage representation attend to itself to dynamically collects evidence from the whole passage.

# Attention: Multi-hop Attention

- Multi-hop attention is first applied in Memory Networks.
- The single attention model (as one layer) is repeated several times. The representation of the passage or question is updated in each hop.



# Pointer Net for Answer Selection

This answer pointer layer uses  $H^r$  as input.  $\mathbf{H}^r \in \mathbb{R}^{2l \times P}$

1. Sequence Model: select tokens separately from the passage.

$$\tilde{\mathbf{H}}^r \in \mathbb{R}^{2l \times (P+1)}$$

$$p(\mathbf{a}|\mathbf{H}^r) = \prod_k p(a_k|a_1, a_2, \dots, a_{k-1}, \mathbf{H}^r)$$

$$p(a_k = j|a_1, a_2, \dots, a_{k-1}, \mathbf{H}^r) = \beta_{k,j} \quad (1 \leq j \leq P+1)$$

$$\beta_k = \text{softmax}(\mathbf{v}^\top \mathbf{F}_k + c \otimes \mathbf{e}_{(P+1)})$$

$$\mathbf{F}_k = \tanh(\mathbf{V}\tilde{\mathbf{H}}^r + (\mathbf{W}^a \mathbf{h}_{k-1}^a + \mathbf{b}^a) \otimes \mathbf{e}_{(P+1)})$$

$$\mathbf{h}_k^a = \overrightarrow{LSTM}(\tilde{\mathbf{H}}^r \beta_k^\top, \mathbf{h}_{k-1}^a)$$

# Pointer Net for Answer Selection

This answer pointer layer uses  $H^r$  as input.  $\mathbf{H}^r \in \mathbb{R}^{2l \times P}$

2. Boundary Model: select span from the passage with predicting the start index and end index.

$$p(\mathbf{a}|\mathbf{H}^r) = p(a_s|\mathbf{H}^r)p(a_e|a_s, \mathbf{H}^r).$$

During prediction, we try to limit the length of the span and globally search the span with the highest probability computed by  $p(a_s) \times \bar{p}(a_e)$

# Currently Leading Model: R-Net

- Proposed in 2017, by Microsoft Research Asia.
- Using several effective techniques, including:
  - Character-level embeddings
  - Gated attention-based recurrent network
  - Self-matching attention layer
  - Modified Pointer Network to predict boundary
- Best Performance in both SQuAD and MS MARCO.

# Question and Passage Encoder

Consider a question  $Q = \{w_t^Q\}_{t=1}^m$  and a passage  $P = \{w_t^P\}_{t=1}^n$ .

- First, words are converted to their respective word-level embeddings and character-level embeddings.

$$\{e_t^Q\}_{t=1}^m \text{ and } \{e_t^P\}_{t=1}^n \quad \{c_t^Q\}_{t=1}^m \text{ and } \{c_t^P\}_{t=1}^n$$

- Then use a bi-directional RNN to produce new representations  $u_1^Q, \dots, u_m^Q$  and  $u_1^P, \dots, u_n^P$

$$u_t^Q = \text{BiRNN}_Q(u_{t-1}^Q, [e_t^Q, c_t^Q])$$

$$u_t^P = \text{BiRNN}_P(u_{t-1}^P, [e_t^P, c_t^P])$$

# Attention-based Recurrent Networks

- Each passage representation could dynamically incorporates aggregated matching information from the question.
- For each word in the passage, get an attention-pooling vector  $c_t$  of the whole question:

$$s_j^t = \mathbf{v}^T \tanh(W_u^Q u_j^Q + W_u^P u_t^P + W_v^P v_{t-1}^P)$$

$$a_i^t = \exp(s_i^t) / \sum_{j=1}^m \exp(s_j^t)$$

$$c_t = \sum_{i=1}^m a_i^t u_i^Q$$

- This matching information from the question is aggregated with the original information via an RNN.

$$v_t^P = \text{RNN}(v_{t-1}^P, [u_t^P, c_t])$$

# Add Another Gate

- Gated Attention-Based Recurrent Networks:

$$g_t = \text{sigmoid}(W_g[u_t^P, c_t])$$

$$[u_t^P, c_t]^* = g_t \odot [u_t^P, c_t]$$

$$v_t^P = \text{RNN}(v_{t-1}^P, [u_t^P, c_t]^*)$$

- The gate determines the importance of passage parts and attend to the ones relevant to the question

# Self-Matching Attention

- Collect evidence from the whole passage and encode it into the current passage word.
- Use similar attention mechanism.
- For each word in the passage, get an attention-pooling vector of the whole passage:

$$s_j^t = \mathbf{v}^T \tanh(W_v^P v_j^P + W_{\tilde{v}}^P v_t^P)$$

$$a_i^t = \exp(s_i^t) / \sum_{j=1}^n \exp(s_j^t)$$

$$c_t = \sum_{i=1}^n a_i^t v_i^P$$

- Aggregate with Gated Attention-based Recurrent Network:

$$h_t^P = \text{BiRNN}(h_{t-1}^P, [v_t^P, c_t])$$

# Output Layer

- Select the start position ( $p^1$ ) and end position ( $p^2$ ) from the passage via Pointer Network:

$$s_j^t = \mathbf{v}^T \tanh(W_h^P h_j^P + W_h^a h_{t-1}^a)$$

$$a_i^t = \exp(s_i^t) / \sum_{j=1}^n \exp(s_j^t)$$

$$p^t = \text{argmax}(a_1^t, \dots, a_n^t)$$

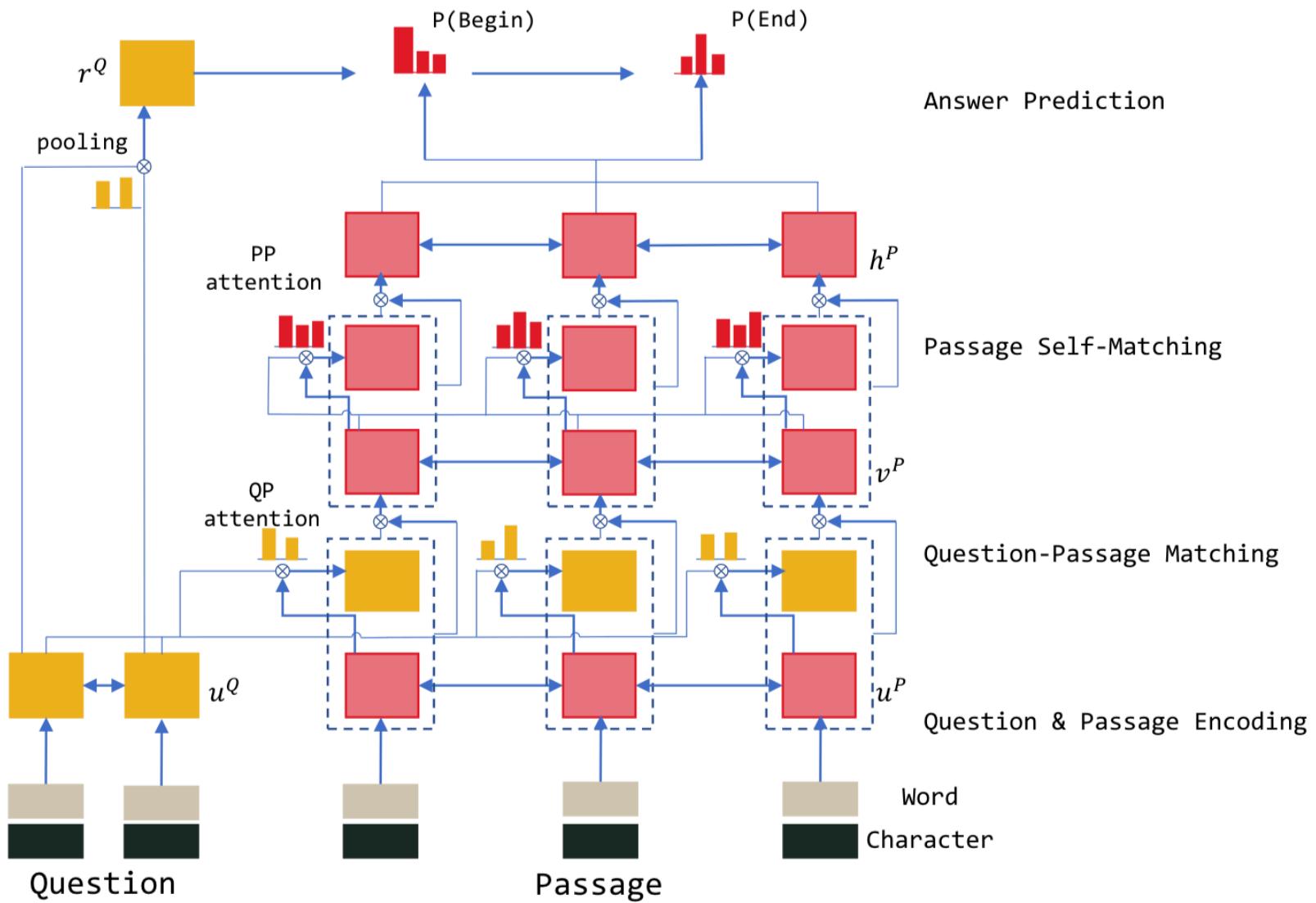
$$c_t = \sum_{i=1}^n a_i^t h_i^P$$

$$h_t^a = \text{RNN}(h_{t-1}^a, c_t)$$

- Use an attention-pooling vector of the question as the initial state of the answer recurrent network

$$r^Q = \text{att}(u^Q, V_r^Q)$$

# R-NET structure overview.



# Conclusion: Challenges

- A very hot NLP subfield, with many efforts on it.
- Performance on SQuAD is very hard to improve.
- Ideas are similar and attention mechanism has been well-exploited.
- With the models becoming more complex, training models could be much harder.
- Currently imperfect system lacks usage scenario.

# Conclusion: Possible Directions

- Large-scale datasets with more difficult questions or more practical values.
- Chinese datasets.
- Incorporating world knowledge into reading comprehension and question answering.
- Try different types of models to use more information, such as syntax.
- Better models designed for inference.

# Thanks!