Question Answering

Hung-yi Lee

李宏毅

Question Answering

- word
- span in source
- correct choice
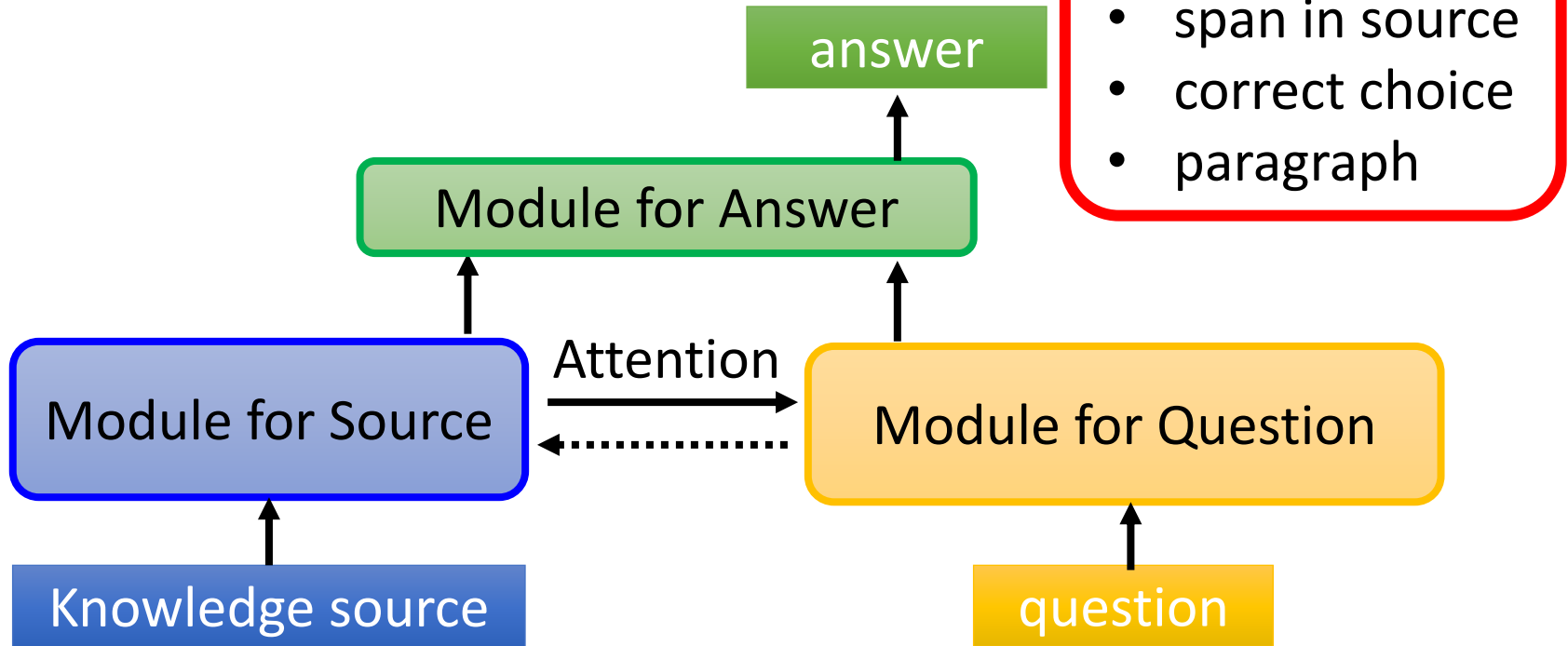- paragraph

answer

Module for Answer

Module for Source → Attention → Module for Question

Module for Source ← ············· Module for Question

Knowledge source
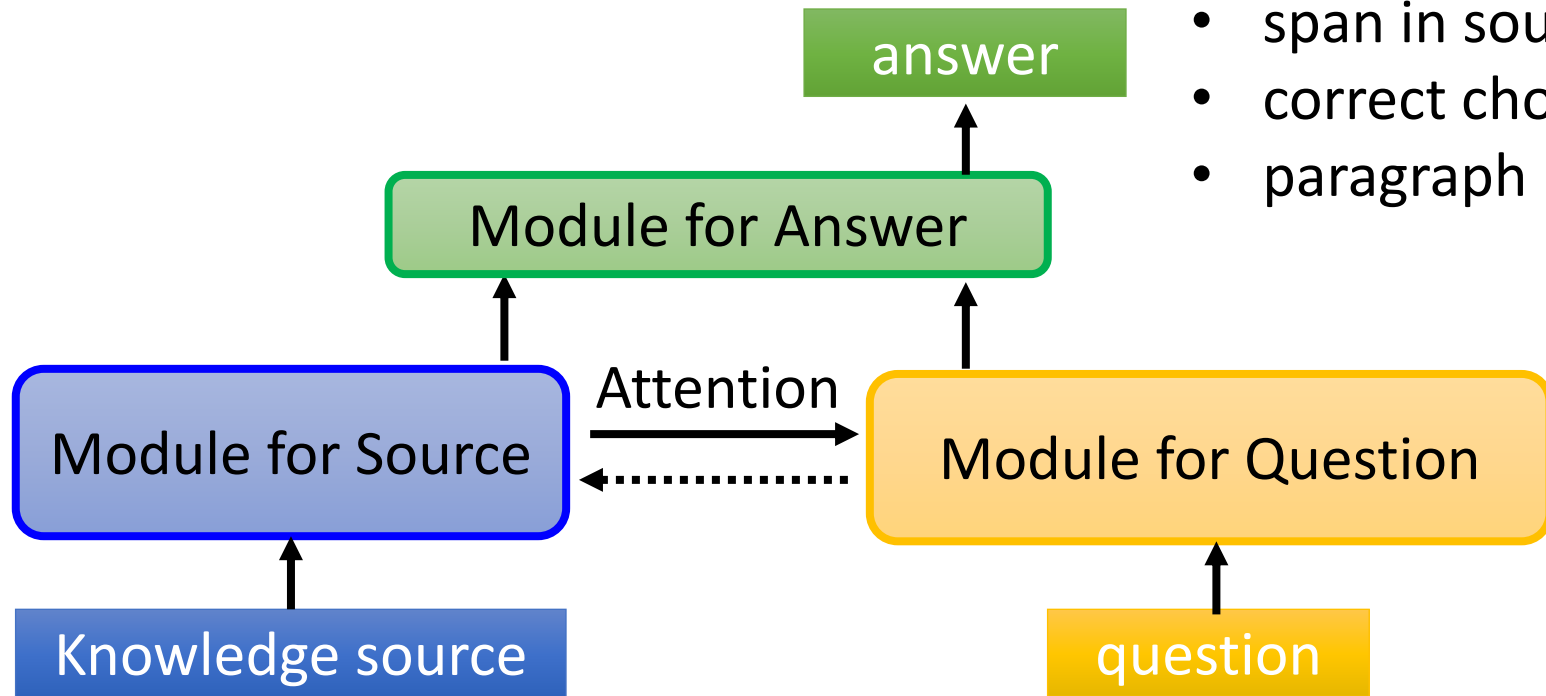
question

Who is the U.S. president?

Is Trump older than than Obama?

Who should pay for the date, and why?

Question Answering

answer

- word
- span in source
- correct choice
- paragraph

Module for Answer

Module for Source → Attention → Module for Question

Knowledge source

question

Who is the U.S. president?

Is Trump older than than Obama?

Who should pay for the date, and why?

Answer is simply a word

- bAbI
  - MNIST of QA
  - 20 types of questions
  - Synthesized

Whether a system can answer questions via chaining facts, simple induction, deduction, etc.

[Weston, et al., arXiv'15]

Task 5: Three Argument Relations

Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 10: Indefinite Knowledge

John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

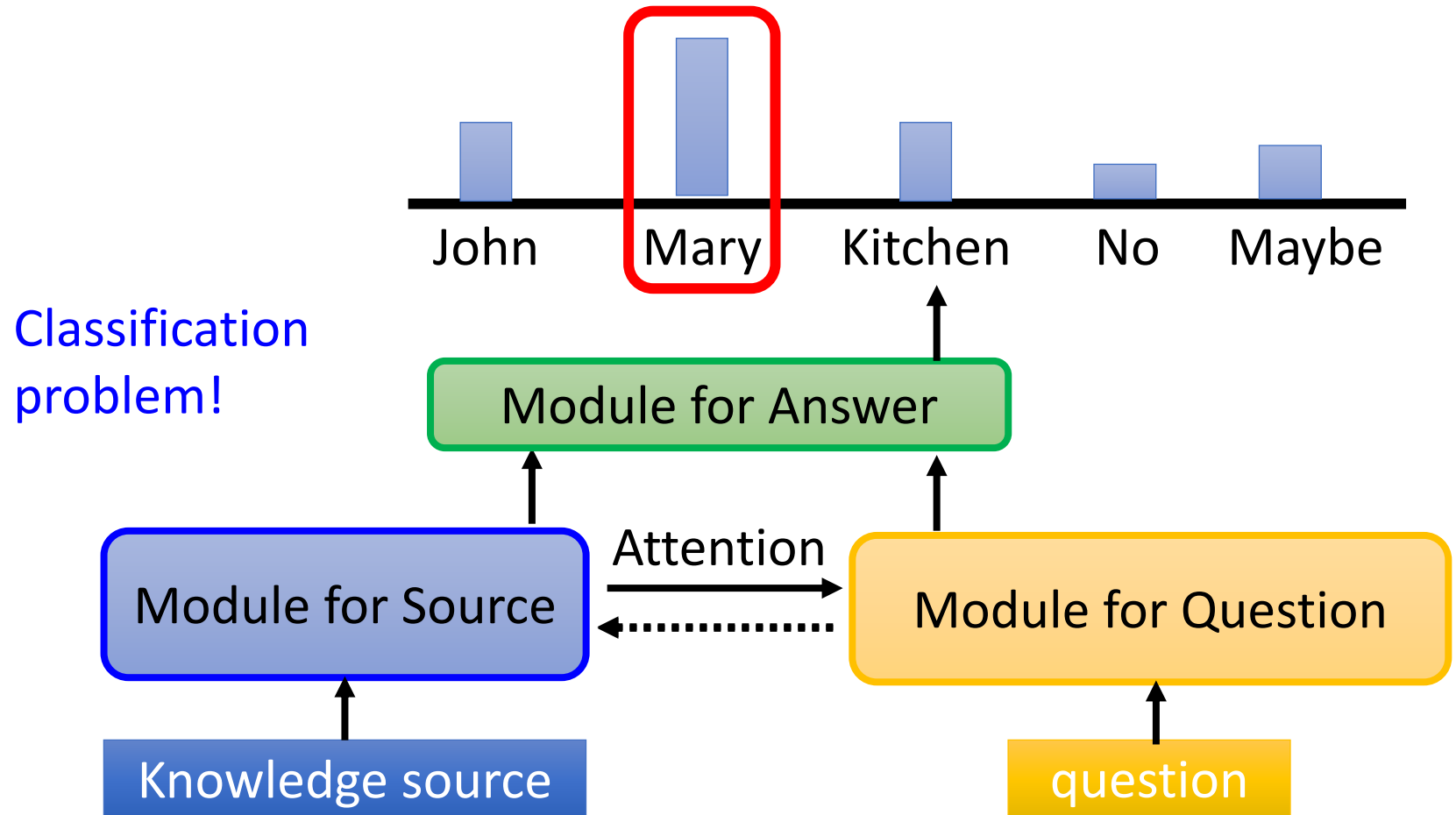**Task 15: Basic Deduction**

Sheep are afraid of wolves.
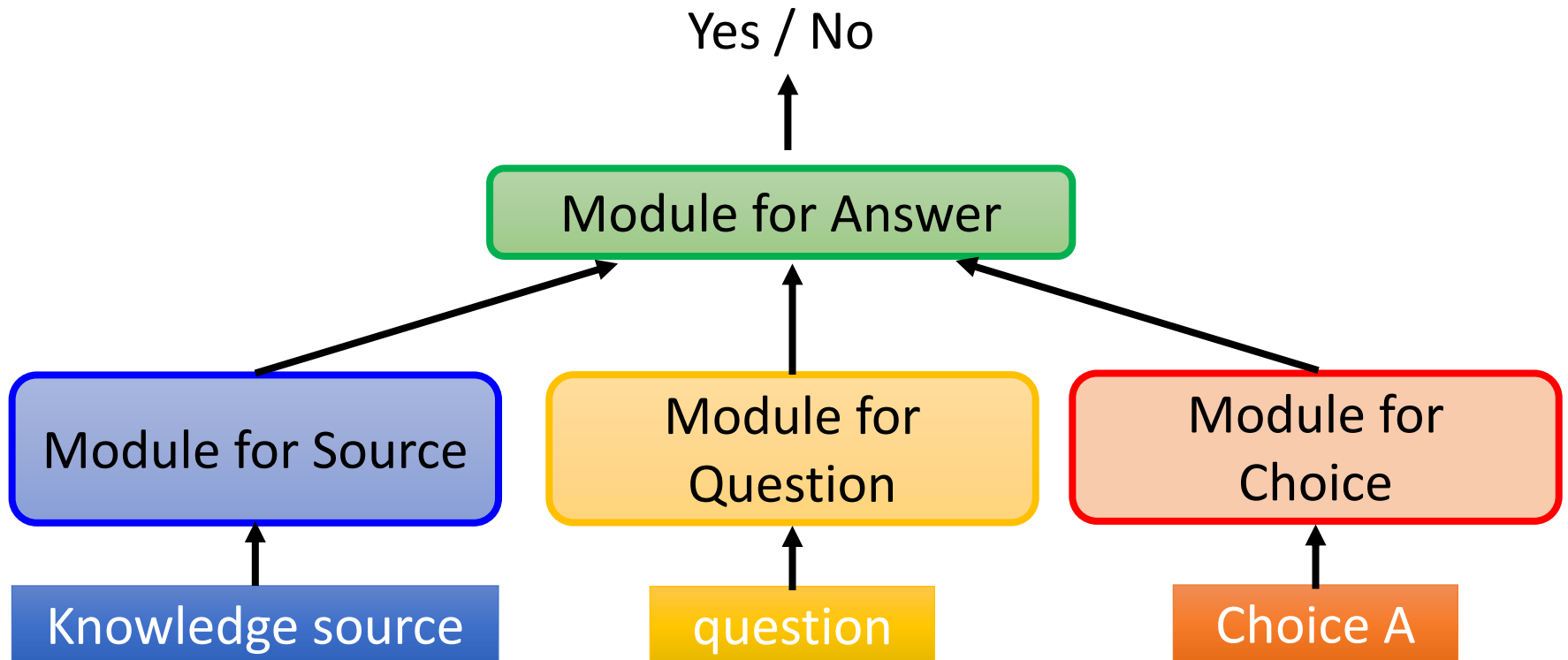Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.
What is Gertrude afraid of? A:wolves

Answer is simply a word

Classification problem!

Multiple Choices

(Ignoring attention here for simplicity)

Knowledge source	question	Choice A	Choice B
		Choice C	Choice D

Yes / No

Module for Answer

Module for Source

Module for Question

Module for Choice

Knowledge source

question

Choice A

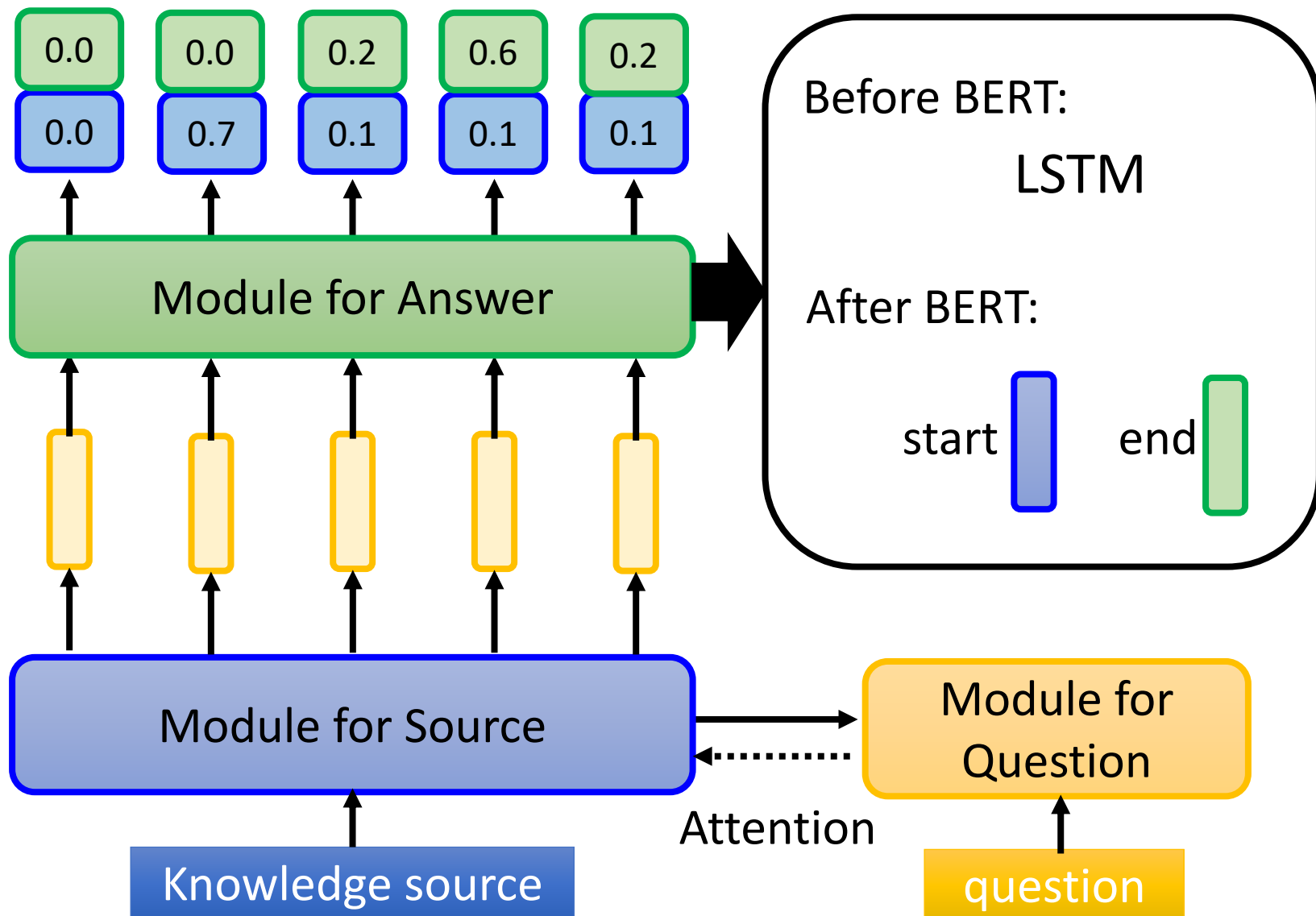# Span in Source / Extraction-based

- SQuAD [Rajpurkar, et al., EMNLP'16], DRCD [Shao, et al., arXiv'18]

End Score	0.0	0.0	0.0	0.2	0.6	0.2	0.0
Start Score	0.1	0.1	0.7	0.1	0.0	0.0	0.0
Source	w_1	w_2	w_3	w_4	w_5	w_6	w_7

The answer is w_3 w_4 w_5

Free Answer Generation

- MS MARCO [Bajaj, et al., NIPS'16], DuReader [He, et al., ACL workshop'18]

All words in the answer are in the passage but from multiple text spans.

Q: who did odysseus see in the underworld

P₁: The souls that Odysseus saw in the Underworld On seeing **Achilles**' soul, said Odysseus: Achilles, the most fortunate man that ever was or will be honored as though you were a god and now you are a mighty prince among the dead.

P₂: Odysseus talked to his mother Anticlea, who died of grief when he did not return home after the Trojan War. Odysseus was also surprised to see **Elphenor**, the youngest member of his crew, in the Underworld.

A: Elphenor and Achilles.

All words in the answer are in the passage and question.

Q: what do **producers need to make food**

P: Plants are producers. Producers are living things that can make their own food using **air, light, soil, and water**. Plants use a process called photosynthesis to make food.

A: Producers need air, light, soil, and water to make food.

Part of words in the answer are not found in the passage or question.

Q: why conversion observed in body

P: Conversion disorder **symptoms** may appear suddenly after a stressful event or trauma, whether physical or psychological. Signs and symptoms that affect movement function may include: 1 Weakness or paralysis. 2 Abnormal movement, such as tremors or difficulty walking. 3 Loss of balance.

A: Due to symptoms in the body

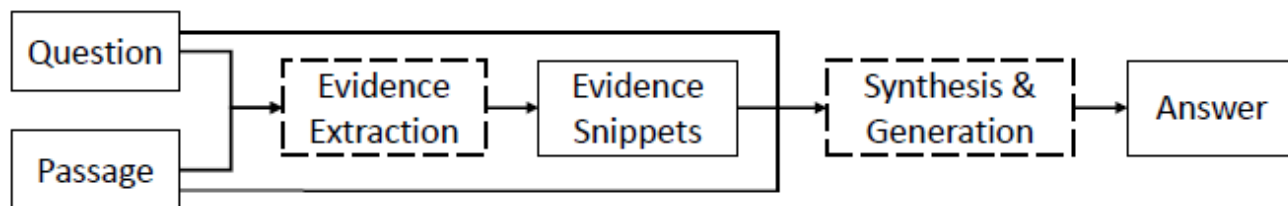**All Words in the answer are not found in the passages or question.**

Q: is there an age limit for learning speech

P: Age is not a detriment to language learning, and by all accounts, learning a second (or third etc) language actually keeps the older language learners mind active. People of all ages can benefit from learning languages.
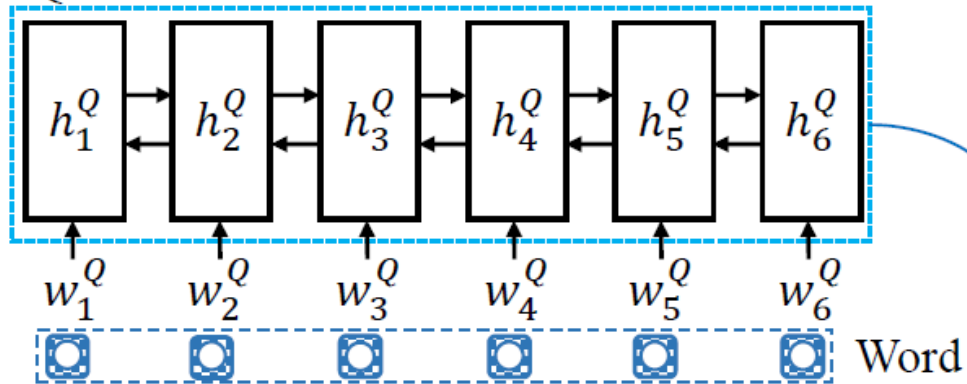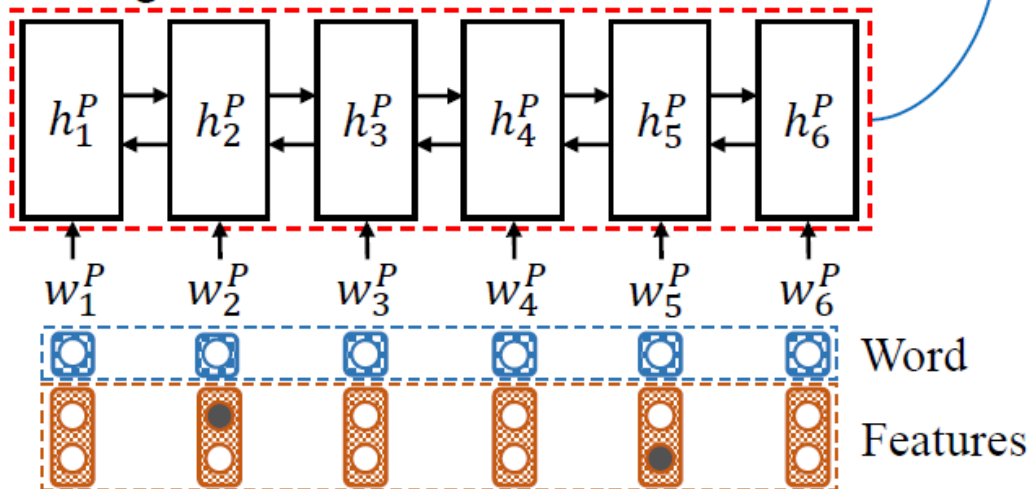
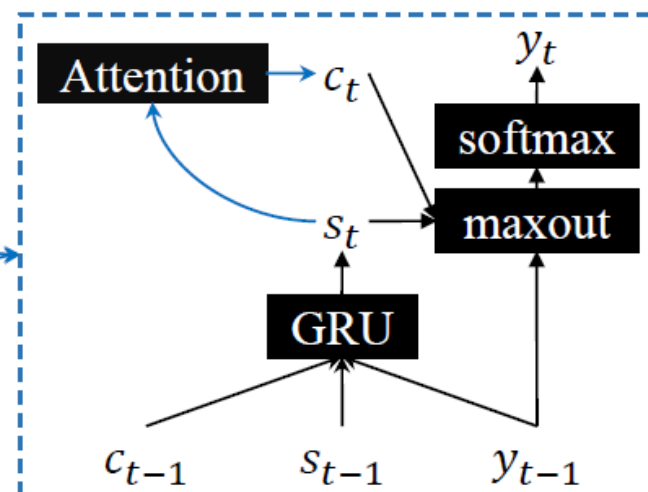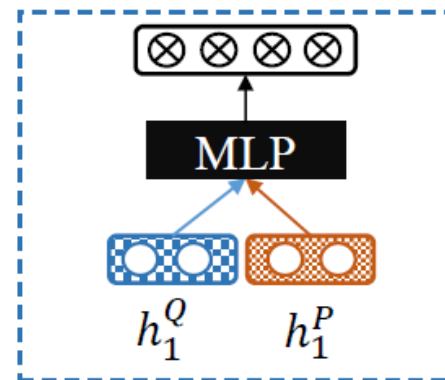*A*: No

S-net

[Tan, et al., AAAI'18]

Question → Evidence Extraction → Evidence Snippets → Synthesis & Generation → Answer

Passage →

Question Encoder

h_1^Q h_2^Q h_3^Q h_4^Q h_5^Q h_6^Q

w_1^Q w_2^Q w_3^Q w_4^Q w_5^Q w_6^Q

Word

Passage Encoder

h_1^P h_2^P h_3^P h_4^P h_5^P h_6^P

w_1^P w_2^P w_3^P w_4^P w_5^P w_6^P

Word

Features

Decoder

Attention → c_t

y_t

softmax

s_t → maxout

GRU

c_{t-1} s_{t-1} y_{t-1}

Decoder Initialization

$\otimes \otimes \otimes \otimes$

MLP

h_1^Q h_1^P

並不是所有問題都有答案 ……

沈默！

這就是正確答案。

你說的沒錯，「這答案完全因人而異」。

所以也就沒有所謂的正確答案！可是評審卻又說我們只能回答①或②。

也就是，不能回答。

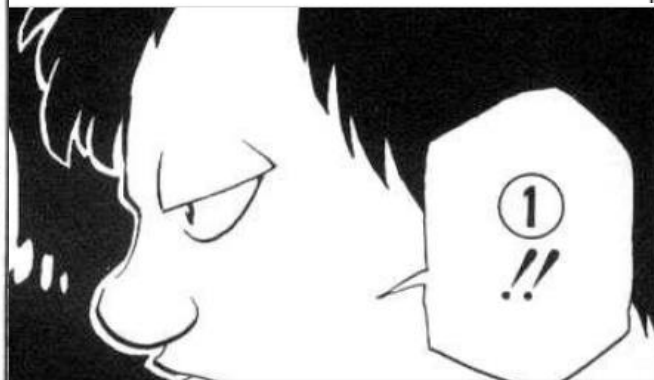只有沈默才是正確。

可是，剛剛那個人⋯

評審沒說他答對。

只是要他過去。

你的壽命就

78
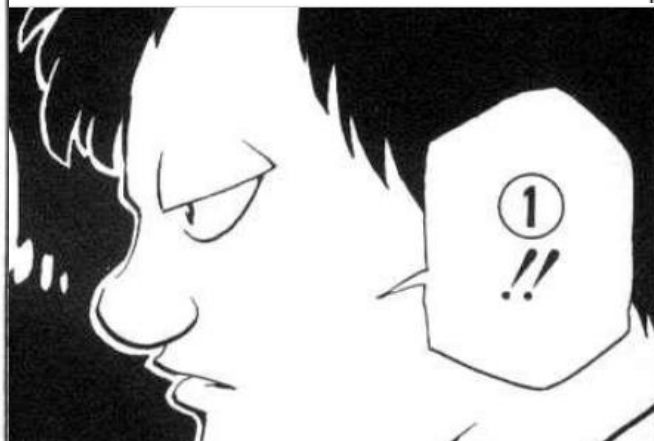
並不是所有問題都有答案 ......

Know what you don't know

• SQuAD 2.0 [Rajpurkar, et al., ACL'18]

Article: Endangered Species Act
Paragraph: " ... *Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*"

Question 1: "*Which laws faced significant opposition?*"
Plausible Answer: *later laws*
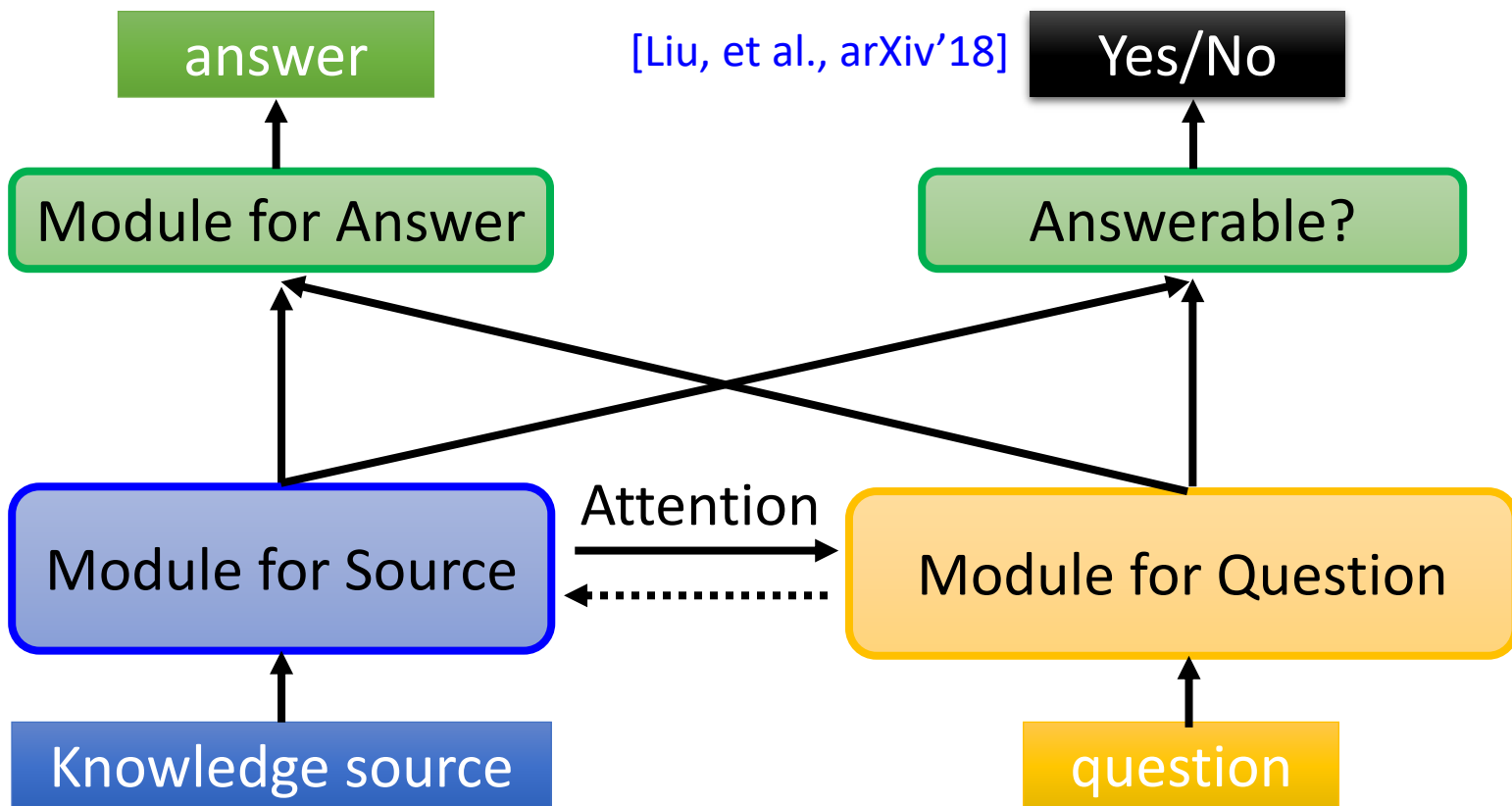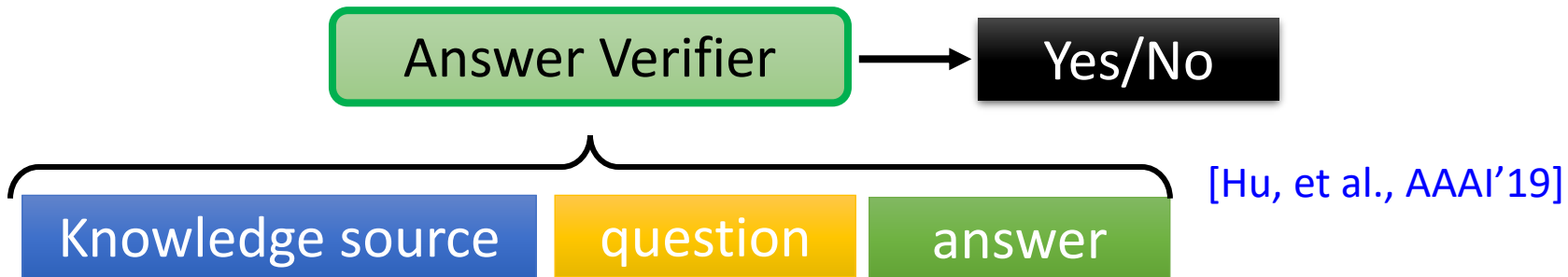
Question 2: "*What was the name of the 1937 treaty?*"
Plausible Answer: *Bald Eagle Protection Act*

Know what you don't know

End Score	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.6
Start Score	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.7
Source	w_1	w_2	w_3	w_4	w_5	w_6	w_7	Null

This is used in the original BERT paper by considering [CLS] as Null.

[Devlin, et al., NAACL'19]
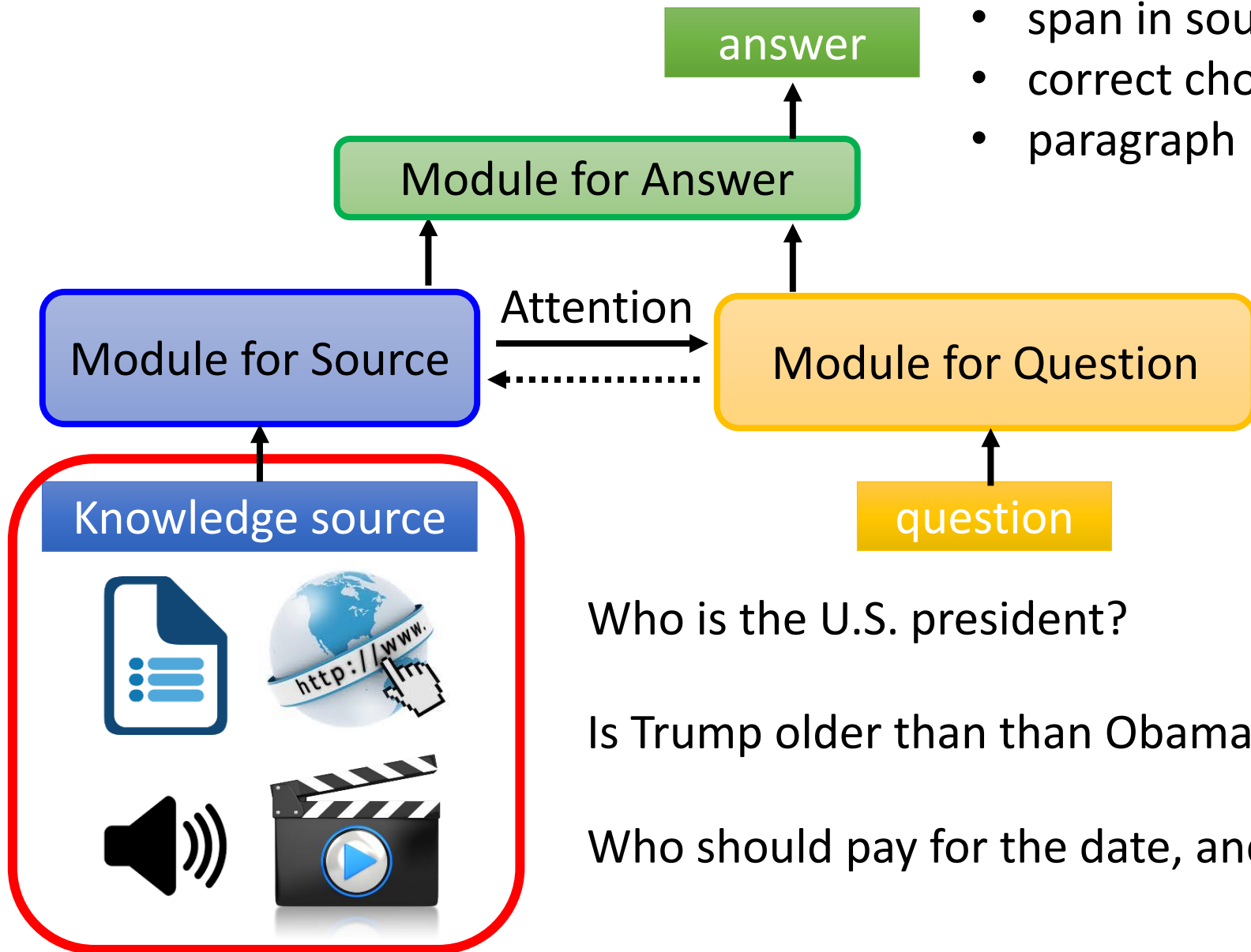
Answer Verifier → Yes/No

Knowledge source | question | answer

[Hu, et al., AAAI'19]

answer [Liu, et al., arXiv'18] Yes/No

Module for Answer → answer

Answerable? → Yes/No

Module for Source — Attention → Module for Question

Knowledge source → Module for Source

question → Module for Question

Question Answering

- word
- span in source
- correct choice
- paragraph

answer

Module for Answer

Module for Source ← Attention → Module for Question

Knowledge source

question

Who is the U.S. president?

Is Trump older than than Obama?

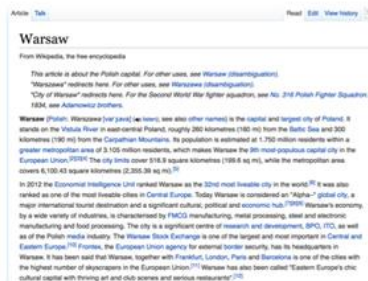Who should pay for the date, and why?

Internet   Not all the documents are relevant

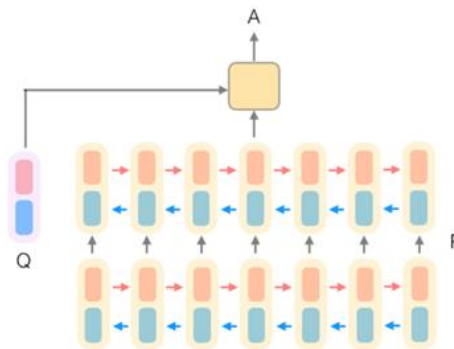Q: How many of Warsaw's inhabitants spoke Polish in 1933?

DrQA

Document Retriever

Document Reader

833,500

[Chen, et al., ACL'17]

MS MARCO and DuReader also have several documents for each question.
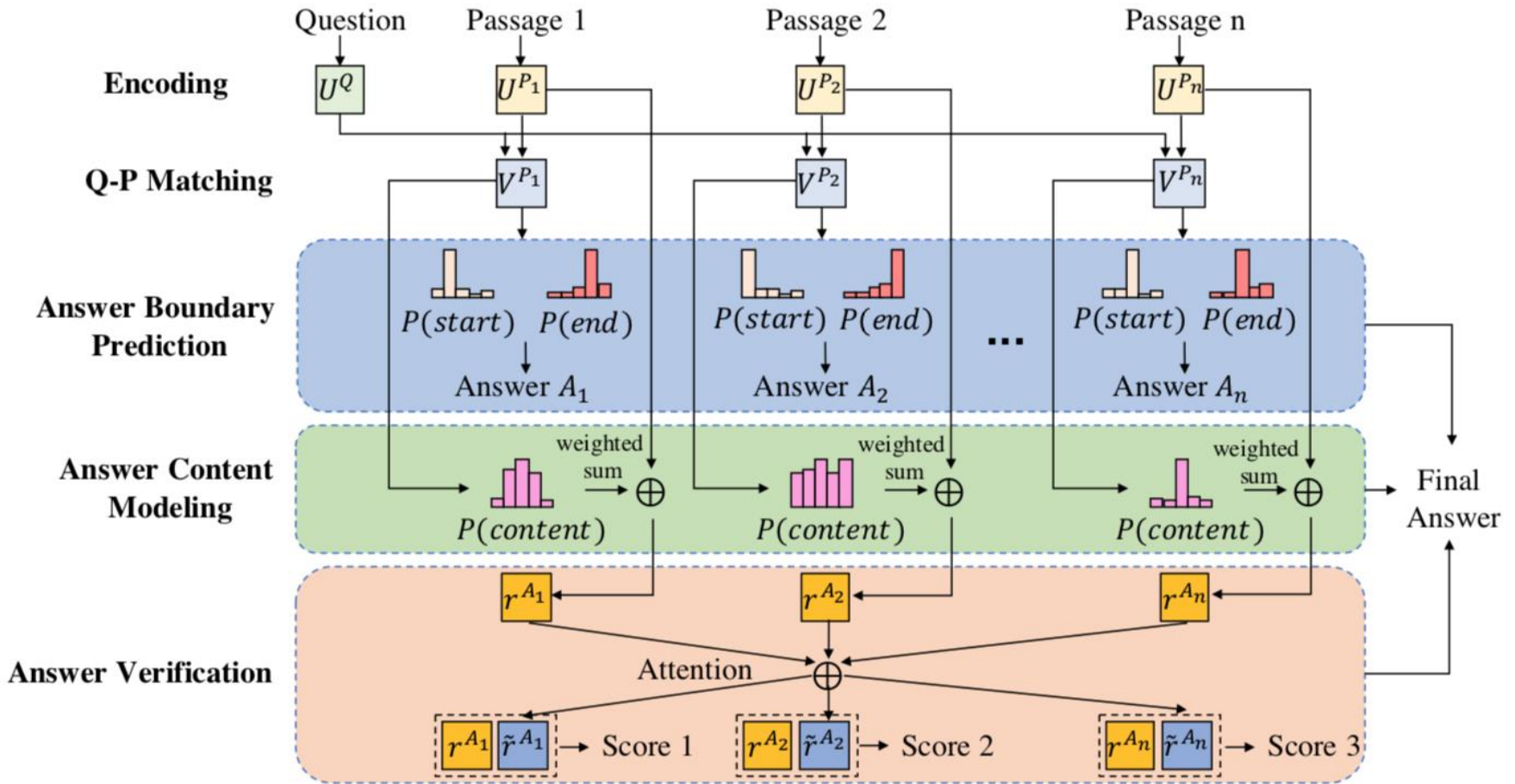
V-Net [Wang, et al., ACL'18]

Figure 1: Overview of our method for multi-passage machine reading comprehension

V-Net   [Wang, et al., ACL'18]

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 in essence true, societies originally form for the express purpose to enhance …

[2] … There has been resurgence in the economic system known as capitalism during the past two decades. 4. **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 other hand, is the lifestyle that the people in the country …

[4] Best Answer: **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 culture contains organisms that derived from some place …

[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 out a small portion of the mixed culture …

[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 is a culture consisting of only one strain. …

… … …

Reference Answer: 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.

Table 1: An example from MS-MARCO. The text in bold is the predicted answer candidate from each passage according to the boundary model. The candidate from [1] is chosen as the final answer by this model, while the correct answer is from [6] and can be verified by the answers from [3], [4], [5].

Visual
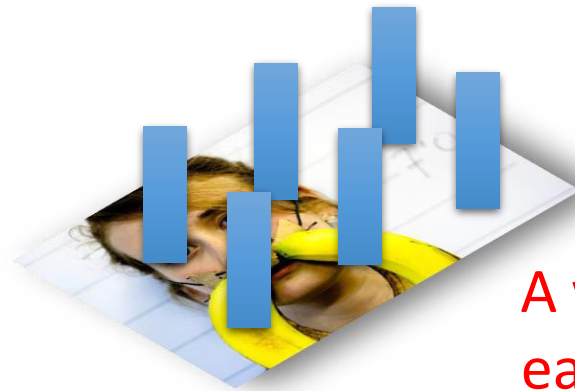
A vector for each region

- source: picture/video

- What is in the image?
- Are there any humans?
- What sport is being played?
- Who has the ball?
- How many players are in the image?
- Who are the teams?
- Is it raining?

source: http://visualqa.org/

Audio

- **TOEFL Listening Comprehension Test by Machine**

Audio Story: 🔊 (The original story is 5 min long.)

Question: " What is a possible origin of Venus' clouds? "

Choices:

(A) gases released as a result of volcanic activity

(B) chemical reactions caused by high surface temperatures

(C) bursts of radio energy from the plane's surface

(D) strong winds that blow dust into the atmosphere

Audio – SQuAD Style

- Link: https://github.com/chiahsuan156/ODSQA

Dataset	QA-pairs	Hours	M-spkrs	F-spkrs	WER-D(%)	WER-Q(%)	Avg D Len	AvgQ Len
ODSQA	3654	25.28	7	13	19.11	18.57	428	22
DRCD-TTS	16746	--	--	--	33.63	--	332	20

SPOKEN OPEN-DOMAIN QUESTION ANSWERING DATASET

Audio – SQuAD Style

- Link: https://github.com/chiahsuan156/ODSQA

Dataset	QA-pairs	Hours	M-spkrs	F-spkrs	WER-D(%)	WER-Q(%)	Avg D Len	AvgQ Len
ODSQA	3654	25.28	7	13	19.11	18.57	428	22
DRCD-TTS	16746	--	--	--	33.63	--	332	20

SPOKEN OPEN-DOMAIN QUESTION ANSWERING DATASET

SOD QA
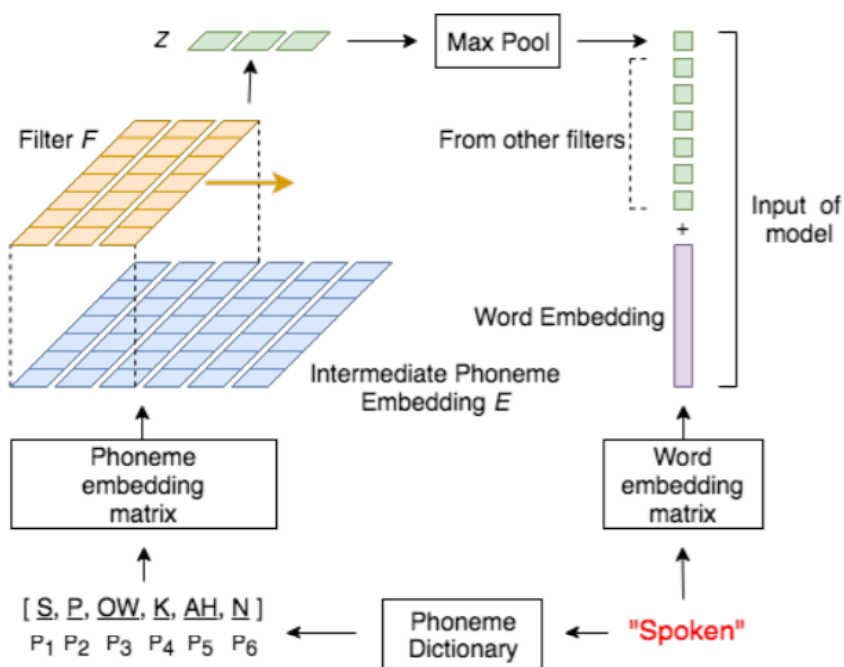
OPEN-DOMAIN SPOKEN QUESTION ANSWERING DATASET

ODS QA

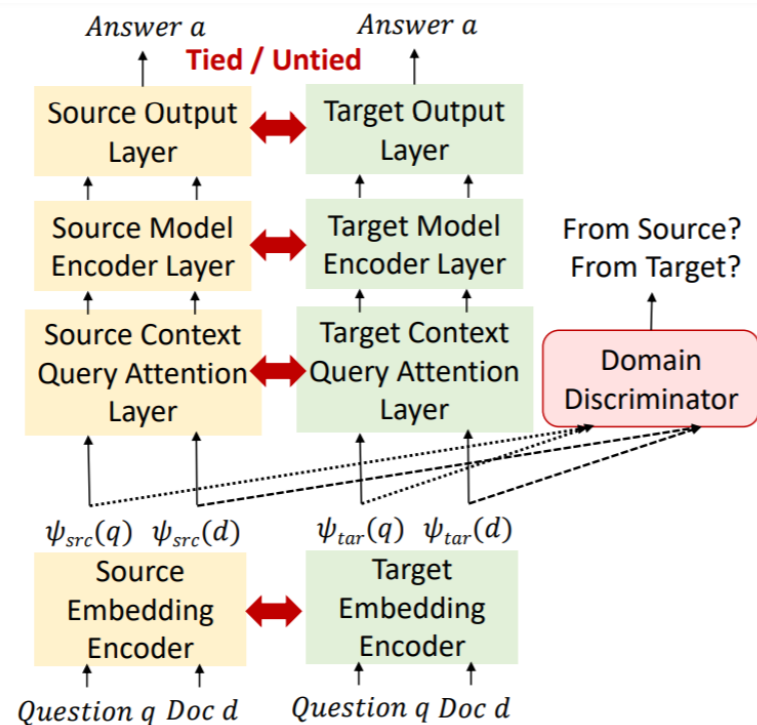[Lee, et al., SLT'18]

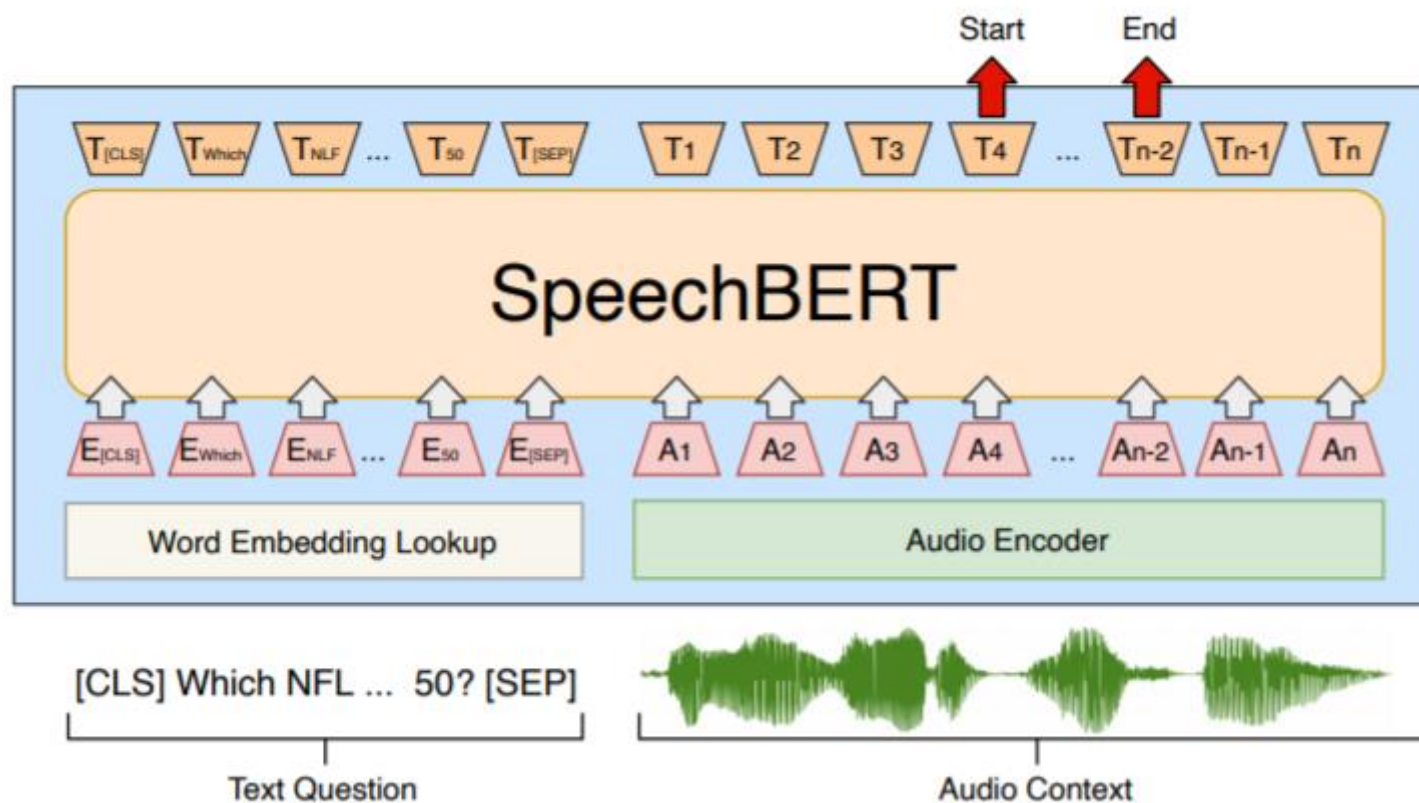# Audio - Techniques Developed

Subword Units

[Lee, et al., INTERSPEECH'18]

Adversarial Learning

[Lee, et al., ICASSP'19]
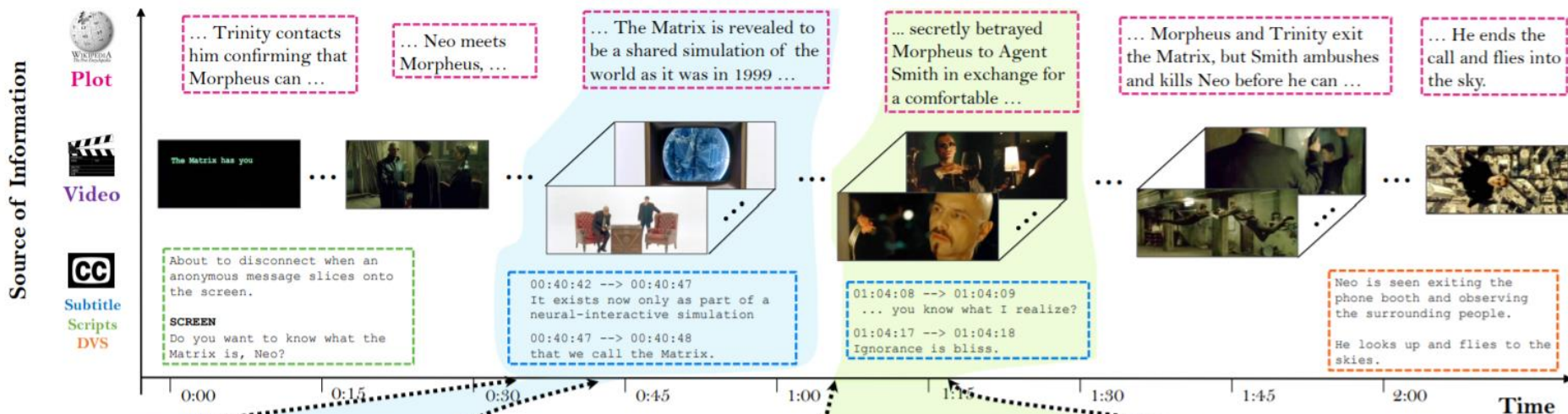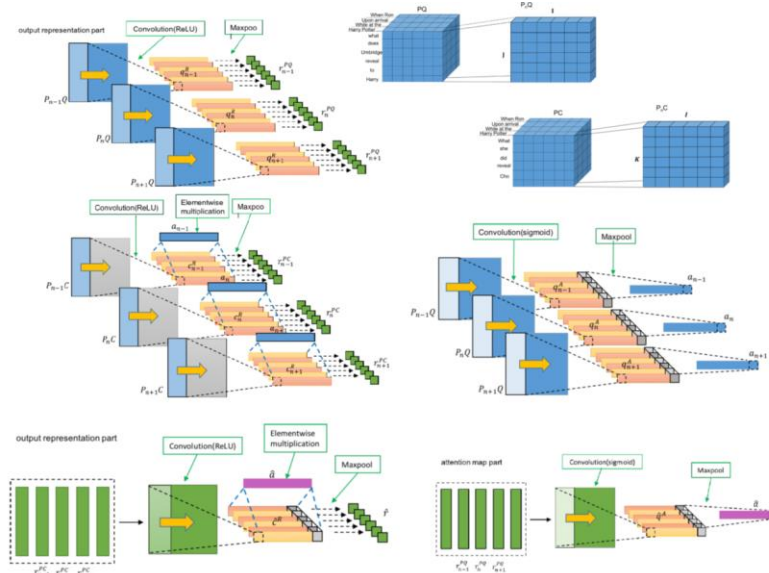
Audio - Towards End-to-end

[Chuang, et al., arXiv'19]

Movie QA

QACNN
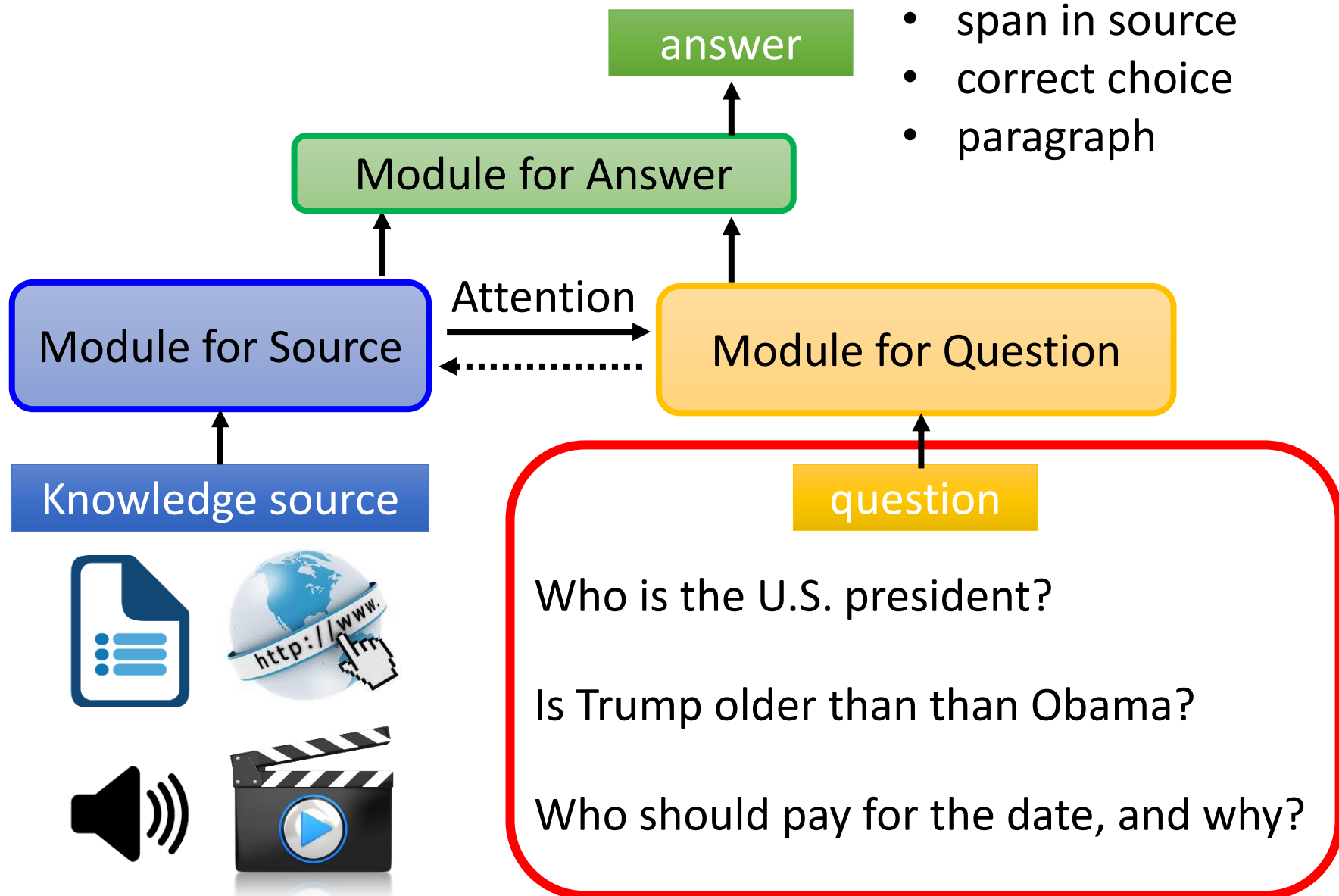
[Liu, et al., ICCV workshop'18]

Source of Information

Plot
- … Trinity contacts him confirming that Morpheus can …
- … Neo meets Morpheus, …
- … The Matrix is revealed to be a shared simulation of the world as it was in 1999 …
- … secretly betrayed Morpheus to Agent Smith in exchange for a comfortable …
- … Morpheus and Trinity exit the Matrix, but Smith ambushes and kills Neo before he can …
- … He ends the call and flies into the sky.

Video

The Matrix has you

Subtitle Scripts DVS

About to disconnect when an anonymous message slices onto the screen.

SCREEN
Do you want to know what the Matrix is, Neo?

00:40:42 --> 00:40:47
It exists now only as part of a neural-interactive simulation

00:40:47 --> 00:40:48
that we call the Matrix.

01:04:08 --> 01:04:09
... you know what I realize?

01:04:17 --> 01:04:18
Ignorance is bliss.

Neo is seen exiting the phone booth and observing the surrounding people.

He looks up and flies to the skies.

0:00   0:15   0:30   0:45   1:00   1:15   1:30   1:45   2:00   Time

Quiz

What is the Matrix?	Who kills Neo in the Matrix?	Why does Cypher betray Morpheus?	How does the movie end?
A: A shared simulation of the world	A: Smith kills Neo	A: In exchange for a comfortable life	A: With Neo flying into the sky
A: A group of robots	A: Trinity kills Neo	A: In exchange for money	A: With the Machines chasing after Neo
A: A human body	A: Morpheus kills Neo after he realizes that Neo is not the one	A: Because he is threatened by Agent Smith	A: We see Mr. Smith torture Morpheus
A: A set of numbers stored as a table			

[Tapaswi, et al., CVPR'16]

Question Answering

- word
- span in source
- correct choice
- paragraph

answer

Module for Answer

Module for Source → Attention → Module for Question

Knowledge source

question

Who is the U.S. president?

Is Trump older than than Obama?

Who should pay for the date, and why?

Types of Questions

Simple Question: Match & Extract

Complex Question: Reasoning

Dialogue QA

Difficulty

Types of Questions

Simple Question: Match & Extract

Complex Question: Reasoning

Dialogue QA

Simple Questions

Match & Extract

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".
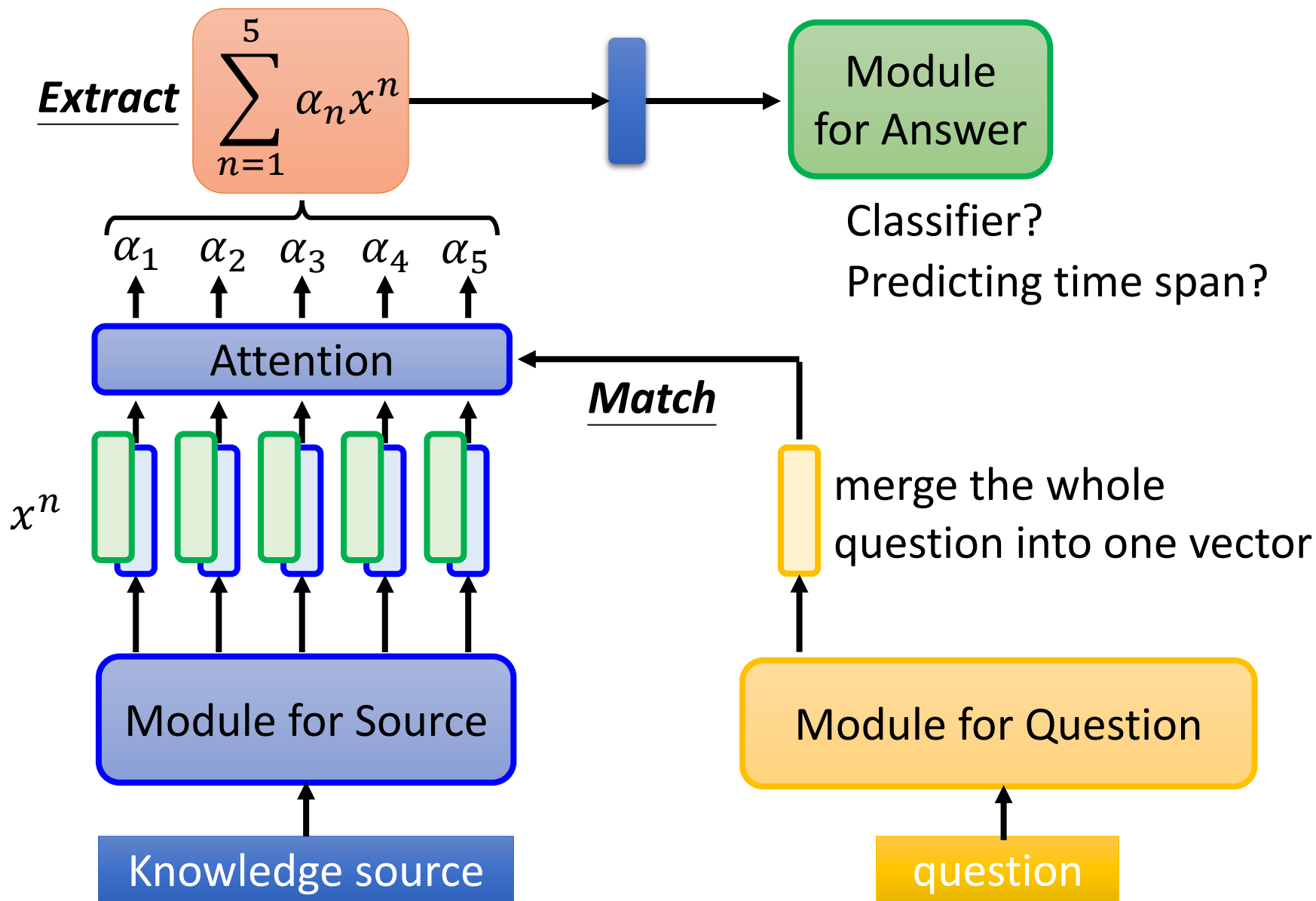
What causes precipitation to fall?
gravity

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

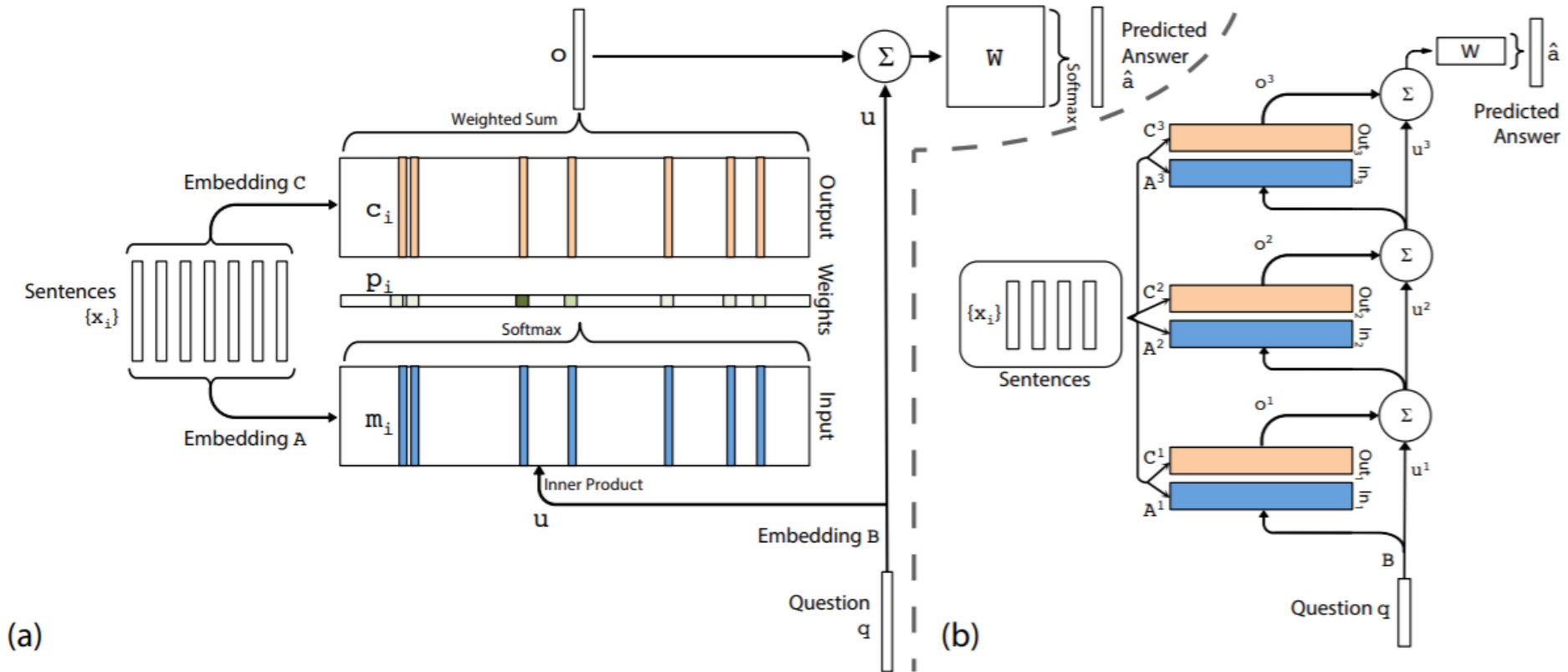*SQuAD*

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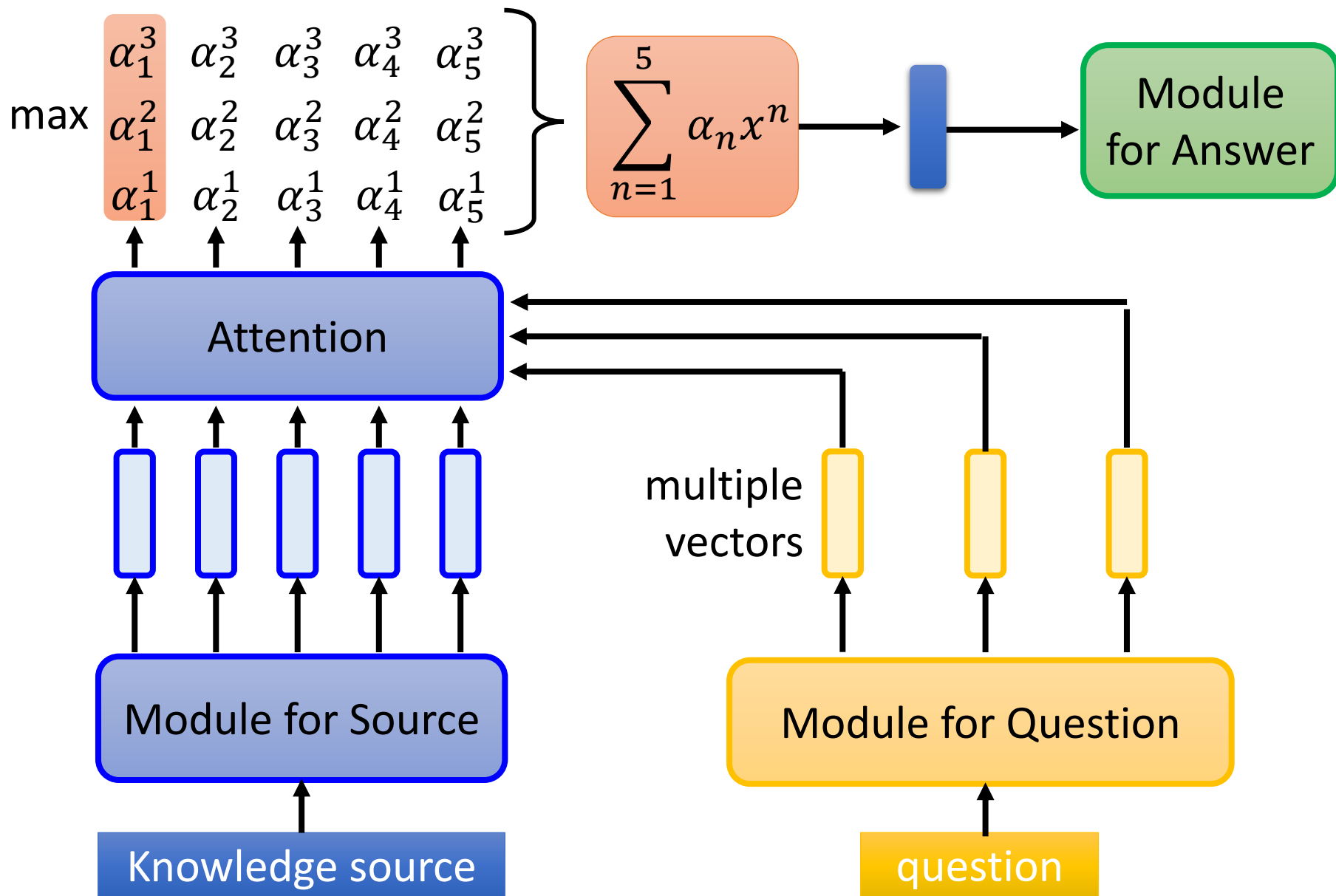within a cloud

Query-to-context Attention

Extract

$$\sum_{n=1}^{5} \alpha_n x^n$$

Module for Answer

Classifier?
Predicting time span?

α_1 α_2 α_3 α_4 α_5

Attention

Match

x^n

merge the whole
question into one vector

Module for Source

Module for Question

Knowledge source

question

Query-to-context Attention

End-to-end Memory Network

[Sukhbaatar, et al., NIPS'15]

Query-to-context Attention (v2)

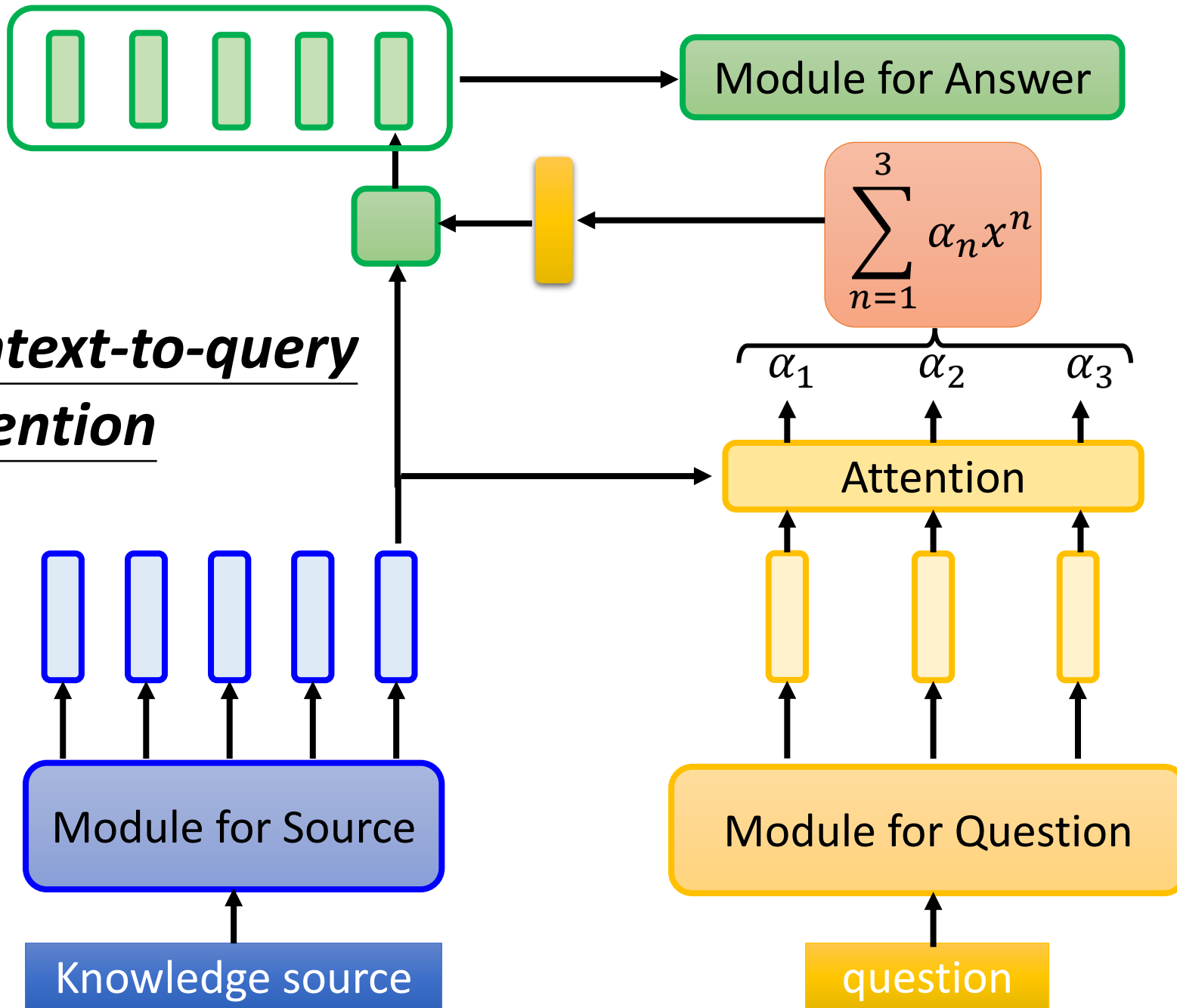$$\max \begin{matrix} \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \alpha_4^3 & \alpha_5^3 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 \\ \alpha_1^1 & \alpha_2^1 & \alpha_3^1 & \alpha_4^1 & \alpha_5^1 \end{matrix} \Bigg\} \quad \sum_{n=1}^{5} \alpha_n x^n \rightarrow$$

Module for Answer

Attention

multiple vectors

Module for Source

Module for Question

Knowledge source

question

Query-to-context Attention (v2)

[Xu, et al., CVPR'16]

Module for Answer

$$\sum_{n=1}^{3} \alpha_n x^n$$

α_1 α_2 α_3
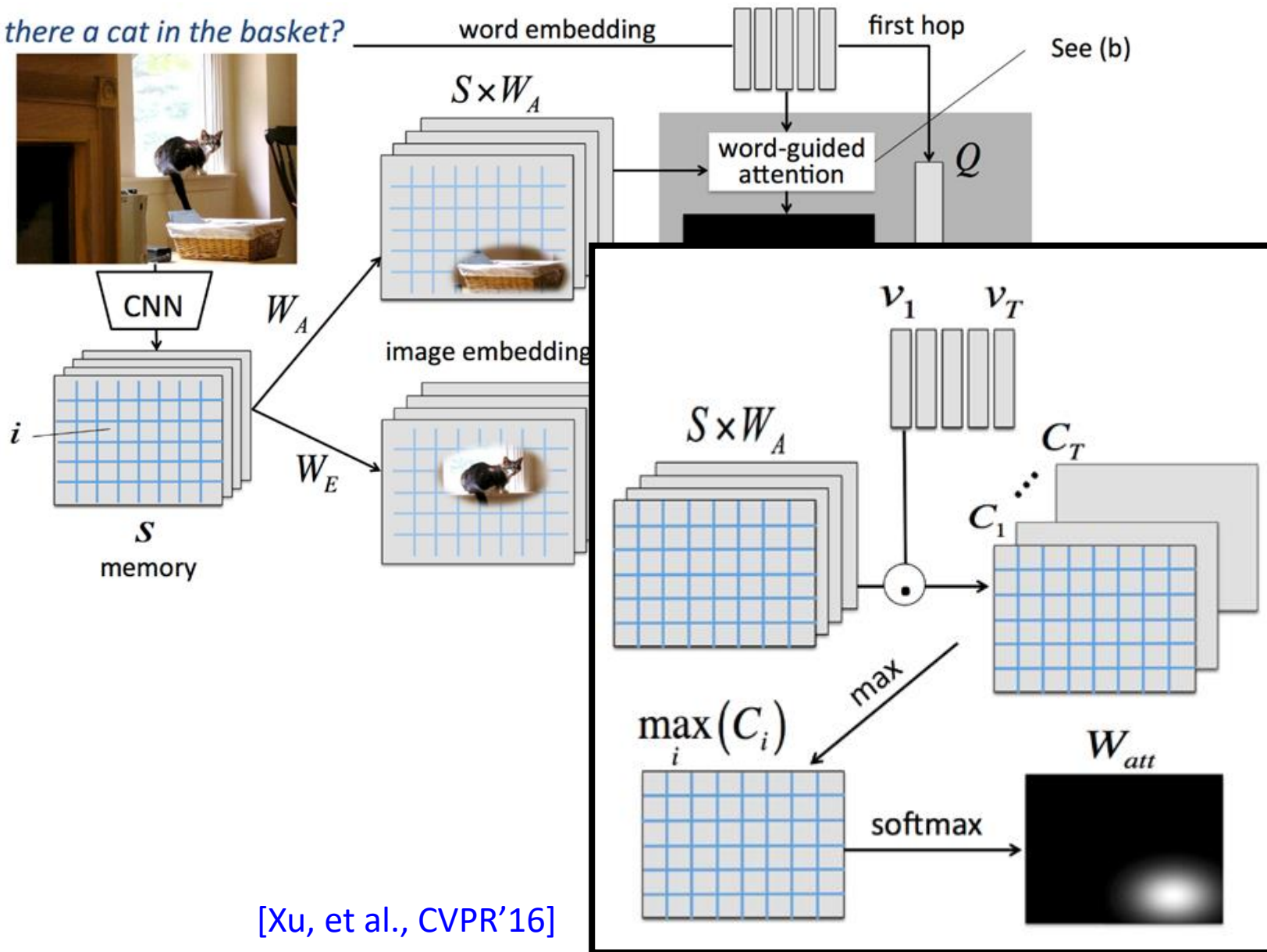
Attention

Context-to-query Attention

Module for Source

Knowledge source

Module for Question

question

R-Net

[Wang, et al., ACL'17]

r^Q

P(Begin)   P(End)

pooling

Answer Prediction

PP attention

h^P

Passage Self-Matching

Self-attention

v^P

QP attention

Question-Passage Matching

u^Q

u^P

Question & Passage Encoding

Word

Character

Question   Passage
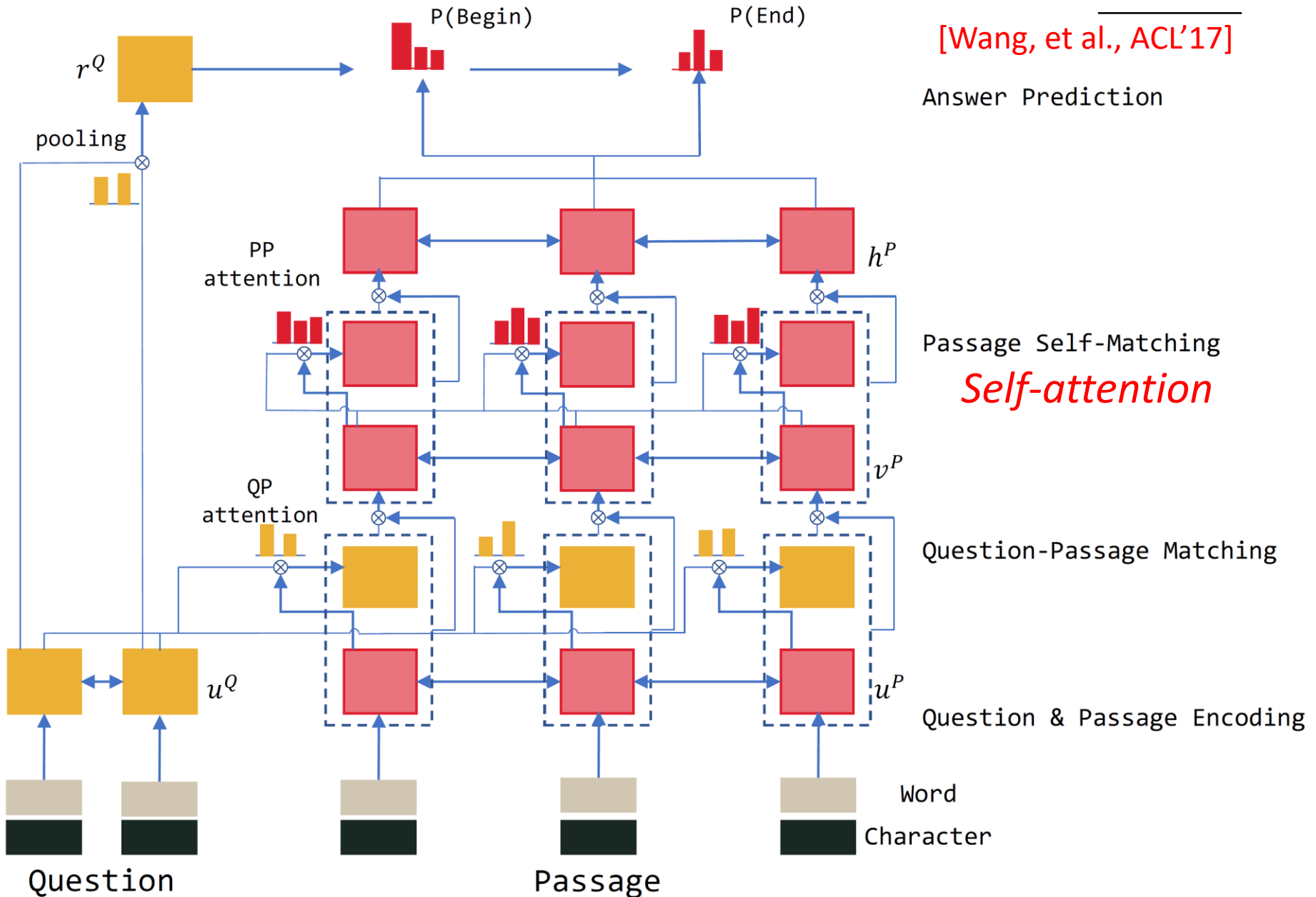
Fully-Aware Fusion Network

Fully-Aware Self-Boosted Fusion

Self-attention

Context Understanding

Question Understanding
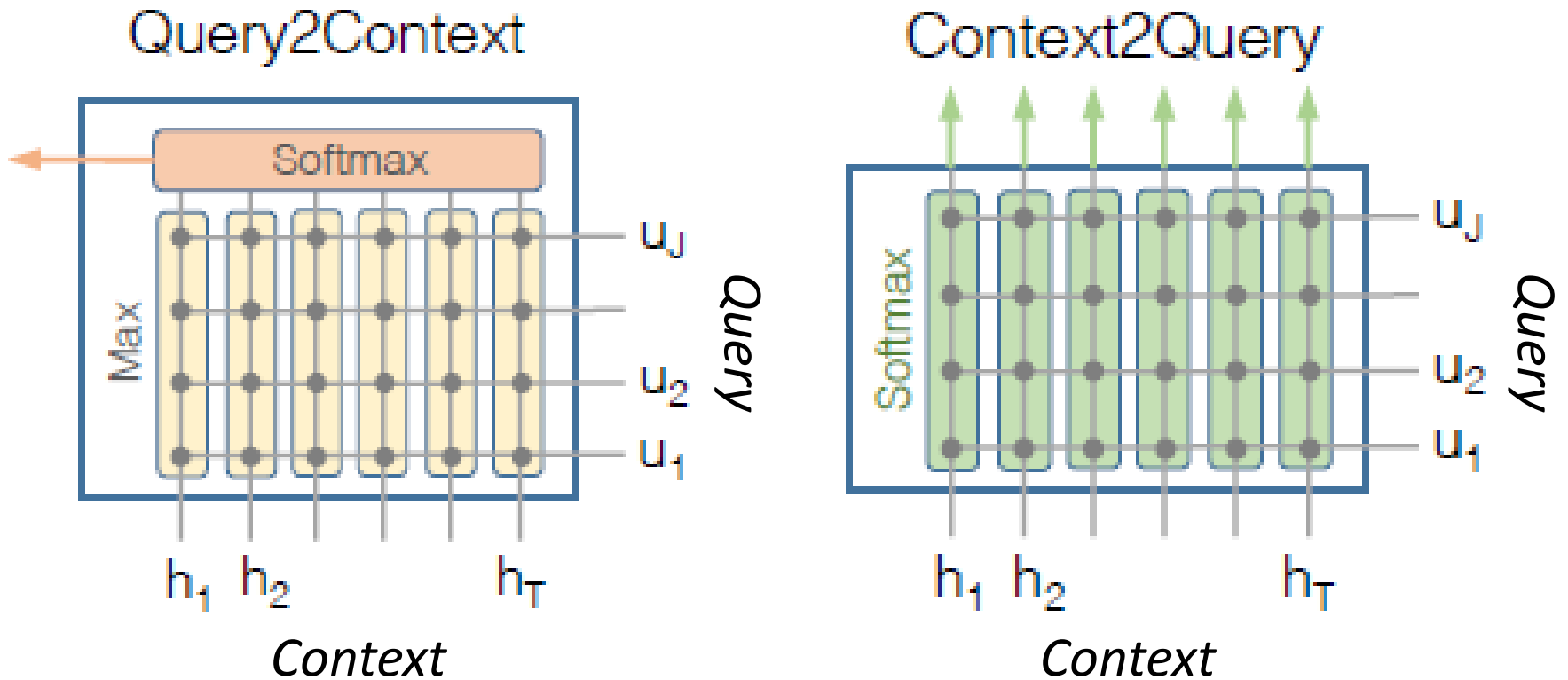
The above can be used to capture long range info.

Fully-Aware Multi-level Fusion

Context Understanding   Question Understanding

▨ : Understanding

■ : High-level Concept

▨ : Low-level Concept

☐ : Input Vector

Context   Word-level   Question

BiLSTM

Concat w/ Distant C Infor.

4

3

BiLSTM

7-Level Fully-Aware Attention

2

Concat w/ Multi-level Q Infor.

1

3-Level Fully-Aware Attention

[Huang, et al., ICLR'18]
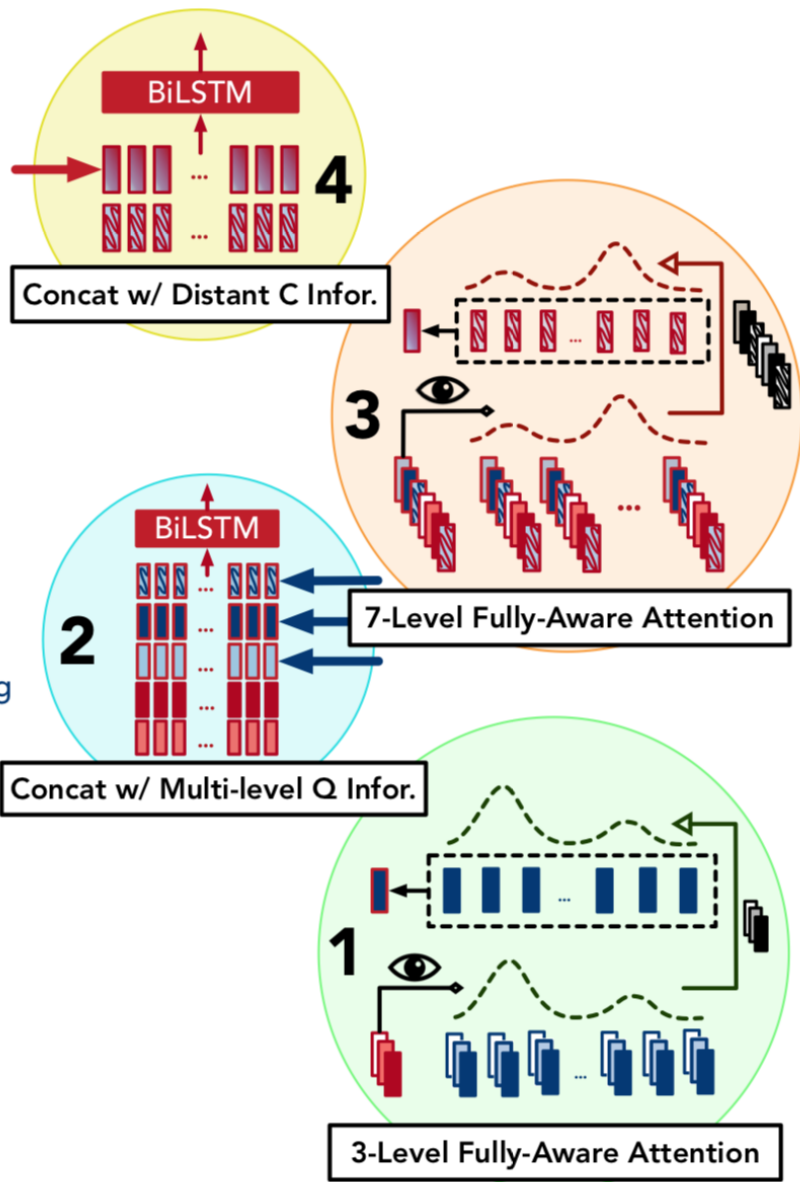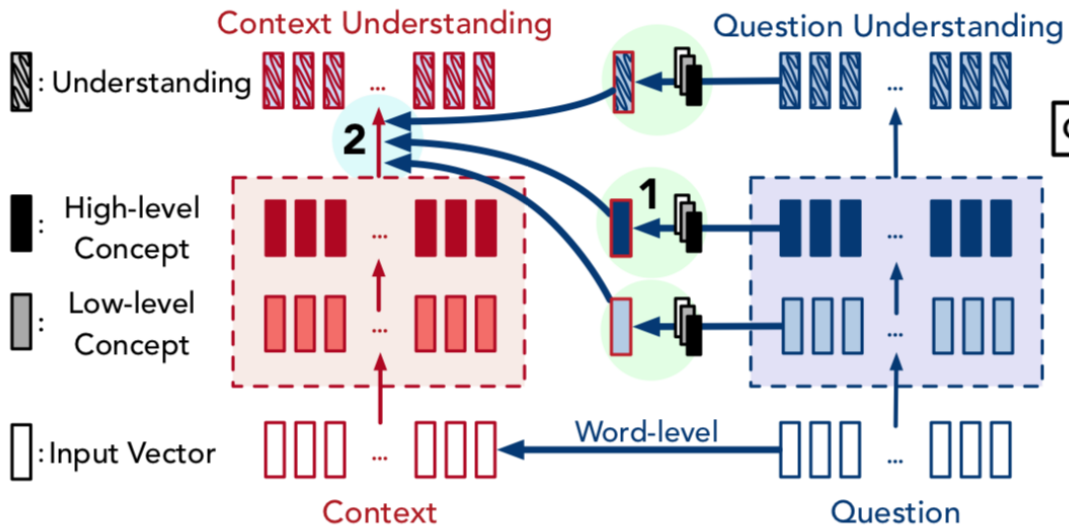
Bi-directional Attention Flow

- BiDAF   [Seo, et al., ICLR'17]

Start   End

Output Layer

Dense + Softmax

Embedding vector

Modeling Layer

m_1 m_2

LSTM

LSTM

concat

Seattle

CNN + Max Pooling

Attention Flow Layer

g_1 g_2

Query2Context and
Attention

S e a t t l e

Contextual Embed Layer

h_1 h_2 h_T u_1 u_J

LSTM   LSTM

Word Embed Layer

Character Embed Layer

x_1 x_2 x_3 x_T q_1 q_J

Context   Query

Dynamic Coattention Networks

[Xiong, et al., ICLR'17]

New
query
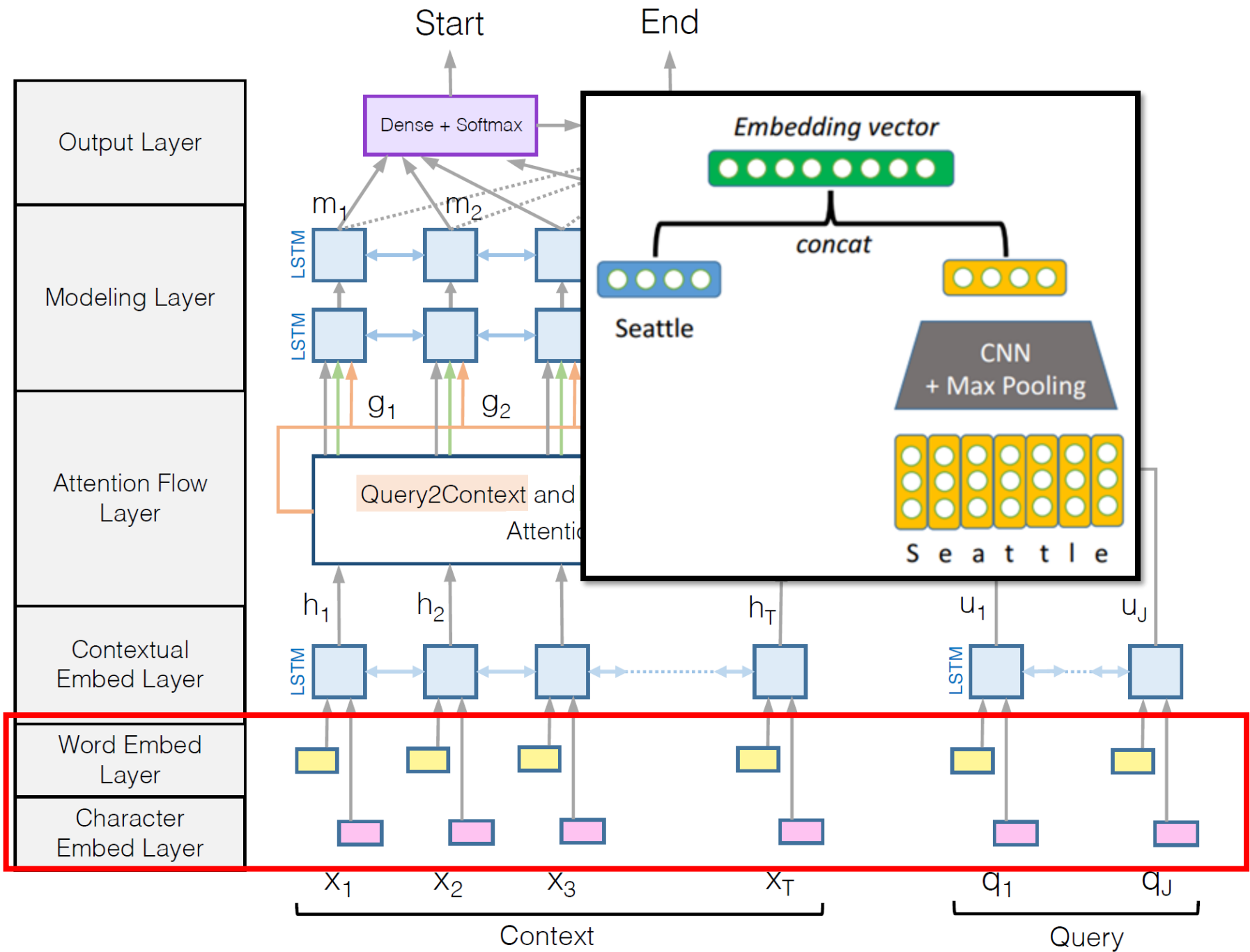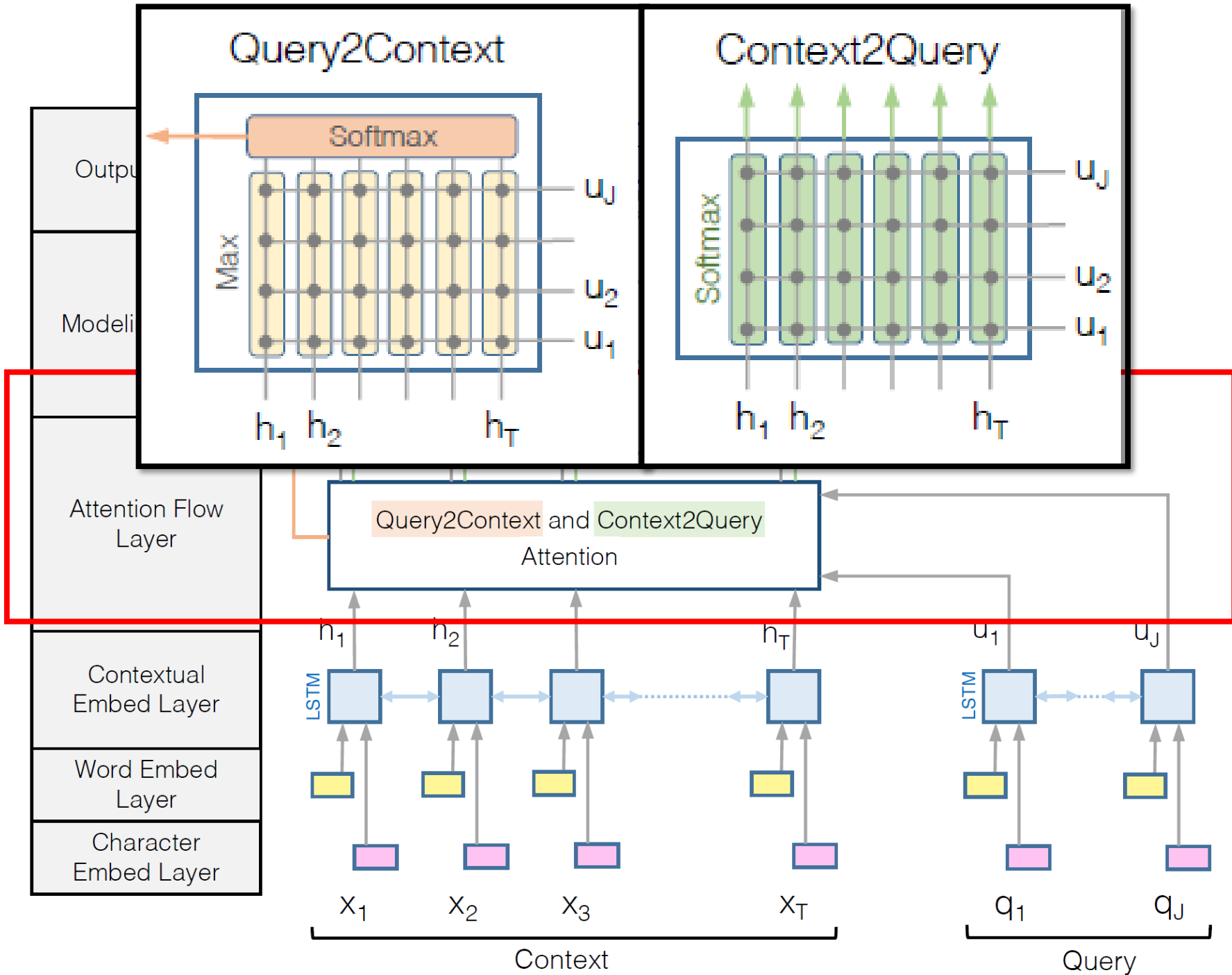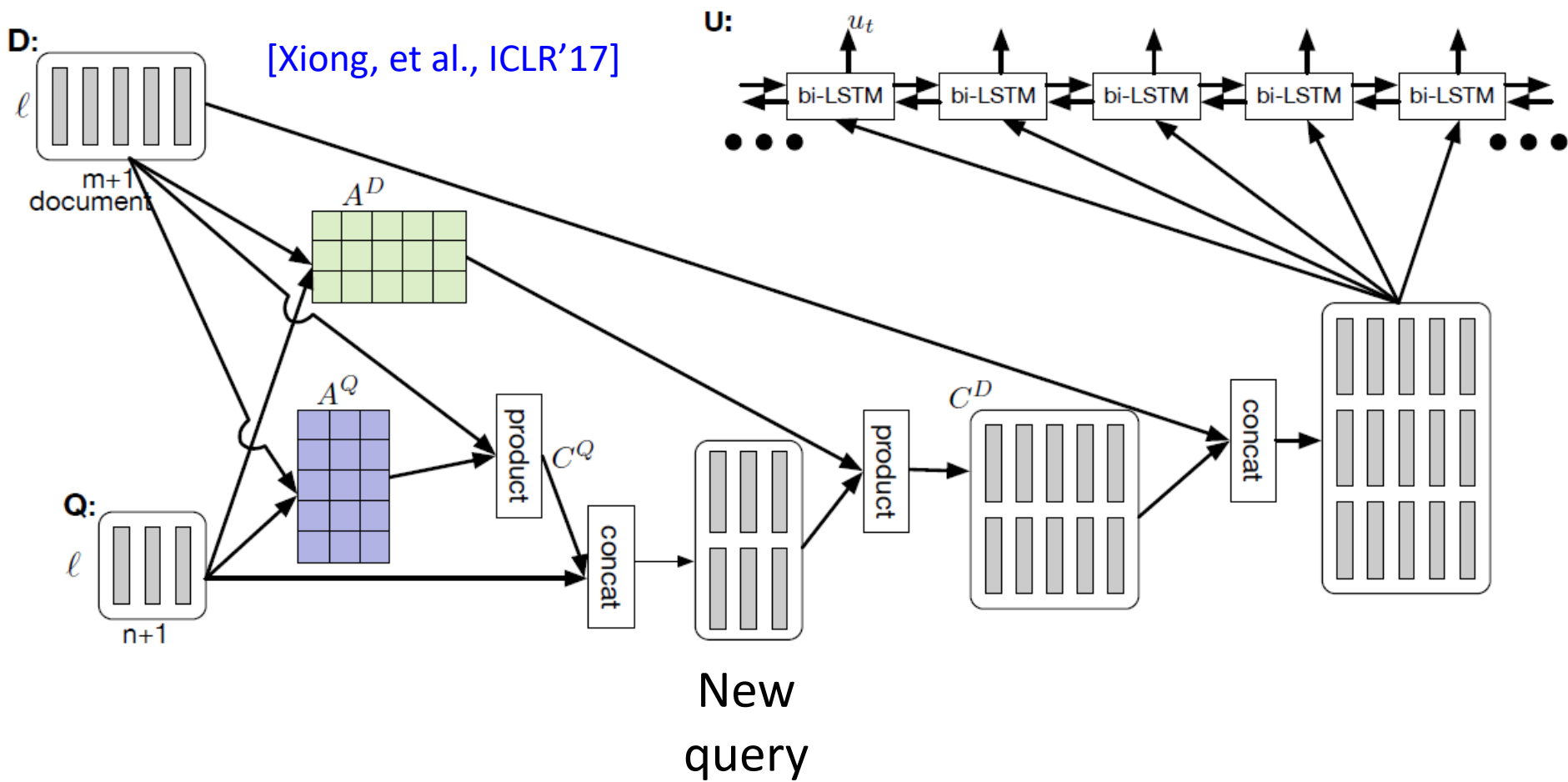
QANet

BERT 家族之外的最後一道榮光

Model

One Encoder Block
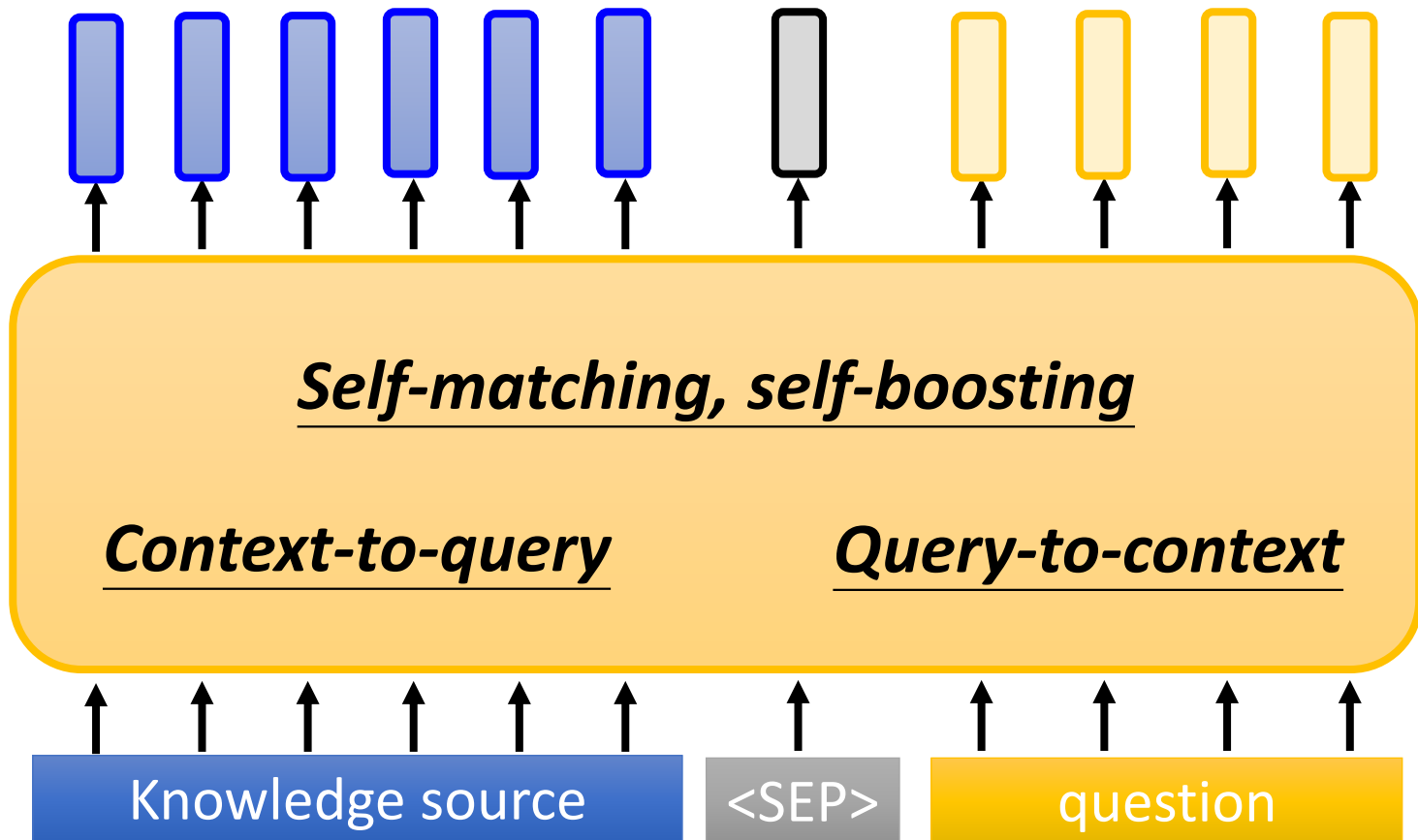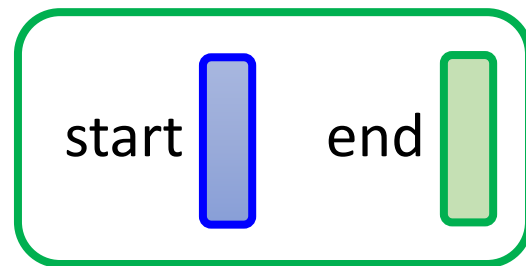
Do not use RNN

[Yu, et al., ICLR'18]
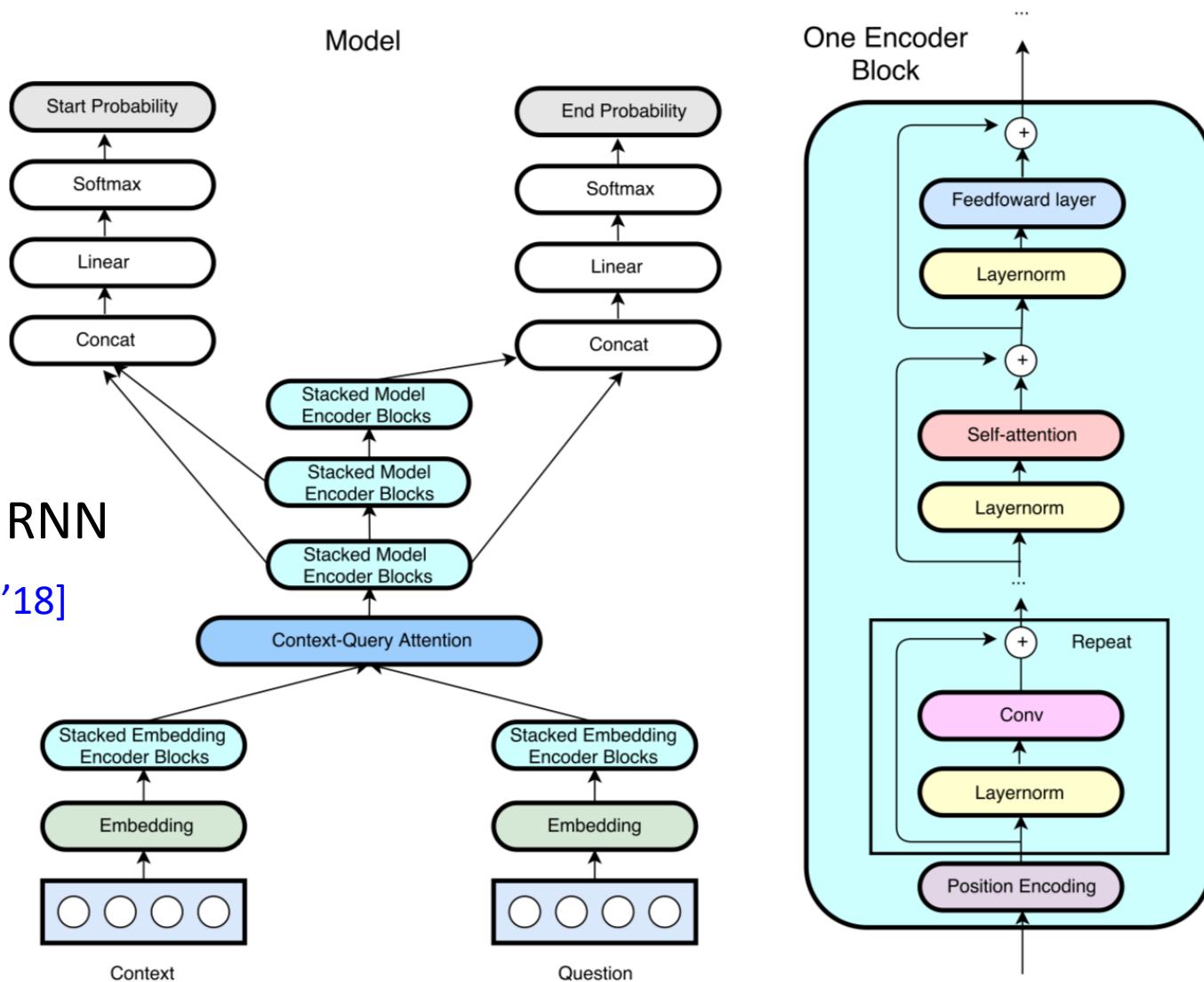
你想要的 BERT 裡面通通都有了!

Types of Questions

Simple Question: Match & Extract

Complex Question: Reasoning

Dialogue QA

Qangaroo

The Hanging Gardens, in **[Mumbai]**, also known as Pherozeshah Mehta Gardens, are terraced gardens … They provide sunset views over the **[Arabian Sea]** …

Mumbai (also known as Bombay, the official name until 1995) is the capital city of the Indian state of Maharashtra. It is the most populous city in **India** …

The **Arabian Sea** is a region of the northern Indian Ocean bounded on the north by **Pakistan** and **Iran**, on the west by northeastern **Somalia** and the Arabian Peninsula, and on the east by **India** …

Q: (Hanging gardens of Mumbai, country, ?)
Options: {Iran, **India**, Pakistan, Somalia, …}

Hoppot QA [Yang, et al., EMNLP'18]

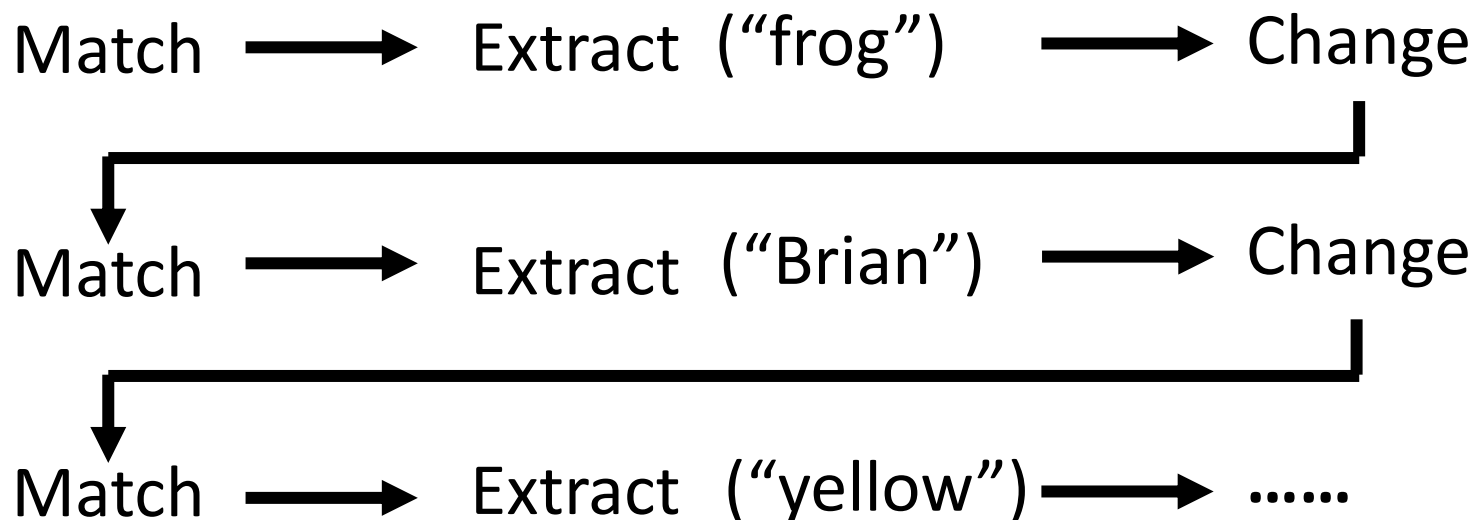| Reasoning Type	%	Example(s)
Inferring the *bridge entity* to complete the 2nd-hop question (Type I)	42	**Paragraph A:** The 2015 Diamond Head Classic was a college basketball tournament ... *Buddy Hield was named the tournament's MVP.* **Paragraph B:** *Chavano Rainier "Buddy" Hield* is a Bahamian professional basketball player for the **Sacramento Kings** of the NBA... **Q:** Which team does the player named 2015 Diamond Head Classic's MVP play for?
Comparing two entities (Comparison)	27	**Paragraph A:** LostAlone were a British rock band ... consisted of *Steven Battelle, Alan Williamson, and Mark Gibson...* **Paragraph B:** Guster is an American alternative rock band ... Founding members *Adam Gardner, Ryan Miller, and Brian Rosenworcel* began... **Q:** Did LostAlone and Guster have the same number of members? (**yes**)
Locating the **answer entity** by checking multiple properties (Type II)	15	**Paragraph A:** Several *current and former members of the Pittsburgh Pirates* ... John Milner, **Dave Parker**, and Rod Scurry... **Paragraph B:** **David Gene Parker**, *nicknamed "The Cobra"*, is an American former player in Major League Baseball... **Q:** Which former member of the Pittsburgh Pirates was nicknamed "The Cobra"?

DROP   **D**iscrete **R**easoning **O**ver the text in the **P**aragraph

[Dua, et al., NAACL'19]

懷抱惡意出題

| Reasoning	Passage (some parts shortened)	Question	Answer
Subtraction (28.8%)	That year, his Untitled (1981), a painting of a haloed, black-headed man with a bright red skeletal body, depicted amid the artists signature scrawls, was sold by Robert Lehrman for $16.3 million, well above its $12 million high estimate.	How many more dollars was the Untitled (1981) painting sold for than the 12 million dollar estimation?	4300000
Comparison (18.2%)	In 1517, the seventeen-year-old King sailed to Castile. There, his Flemish court .... In May 1518, Charles traveled to Barcelona in Aragon.	Where did Charles travel to first, Castile or Barcelona?	Castile
Selection (19.4%)	In 1970, to commemorate the 100th anniversary of the founding of Baldwin City, Baker University professor and playwright Don Mueller and Phyllis E. Braun, Business Manager, produced a musical play entitled The Ballad Of Black Jack to tell the story of the events that led up to the battle.	Who was the University professor that helped produce The Ballad Of Black Jack, Ivan Boyd or Don Mueller?	Don Mueller
Addition (11.7%)	Before the UNPROFOR fully deployed, the HV clashed with an armed force of the RSK in the village of Nos Kalik, located in a pink zone near Šibenik, and captured the village at 4:45 p.m. on 2 March 1992. The JNA formed a battlegroup to counterattack the next day.	What date did the JNA form a battlegroup to counterattack after the village of Nos Kalik was captured?	3 March 1992
Count (16.5%) and Sort (11.7%)	Denver would retake the lead with kicker Matt Prater nailing a 43-yard field goal, yet Carolina answered as kicker John Kasay ties the game with a 39-yard field goal. ... Carolina closed out the half with Kasay nailing a 44-yard field goal. ... In the fourth quarter, Carolina sealed the win with Kasay's 42-yard field goal.	Which kicker kicked the most field goals?	John Kasay

Multiple-hop

Match $\longrightarrow$ Extract ("frog") $\longrightarrow$ Change

Match $\longrightarrow$ Extract ("Brian") $\longrightarrow$ Change

Match $\longrightarrow$ Extract ("yellow") $\longrightarrow$ ......

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow	Prediction: yellow			

Memory Network

Change what to match

Extract

Match

Change what to match

Extract

Match

q
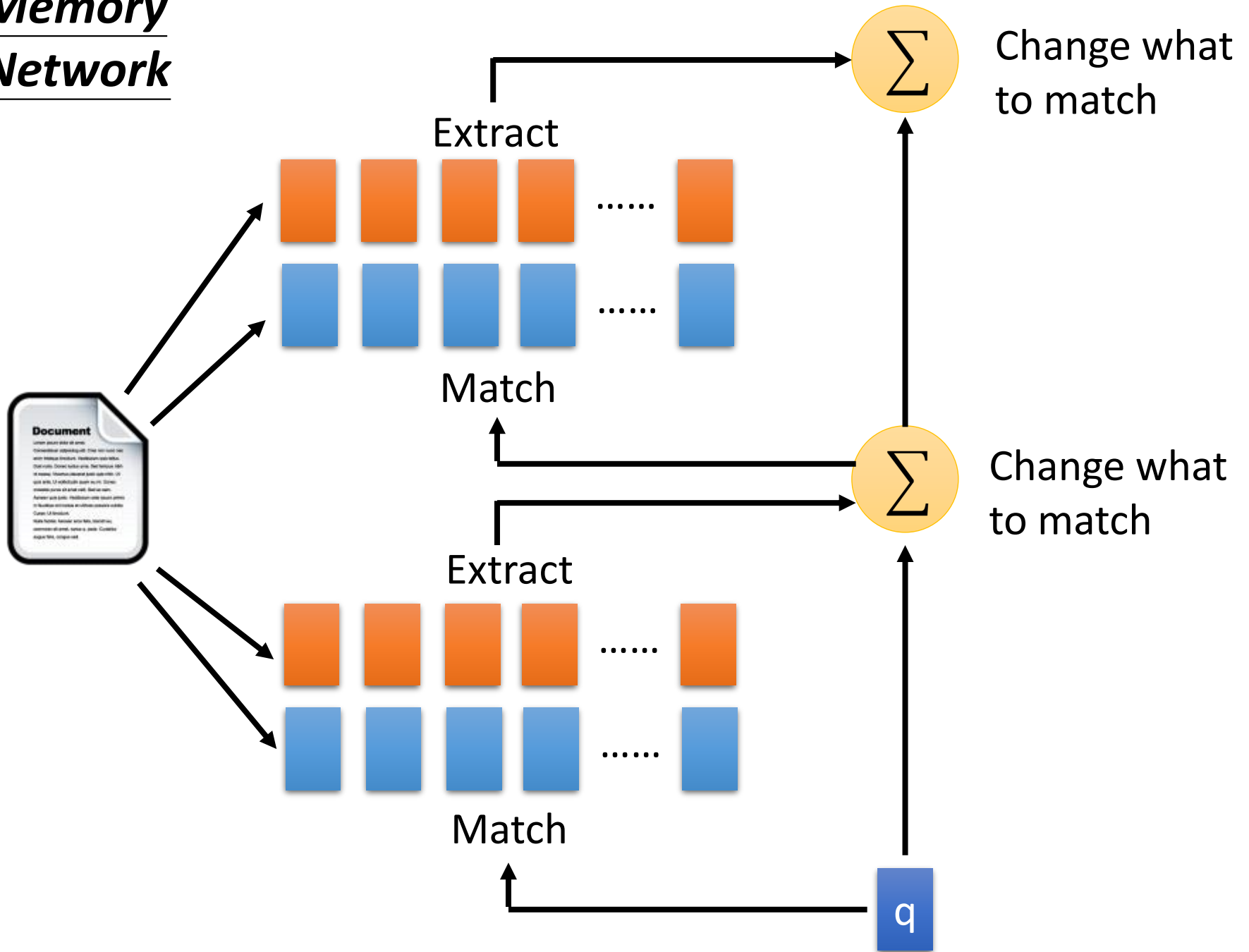
How many hops do we need?
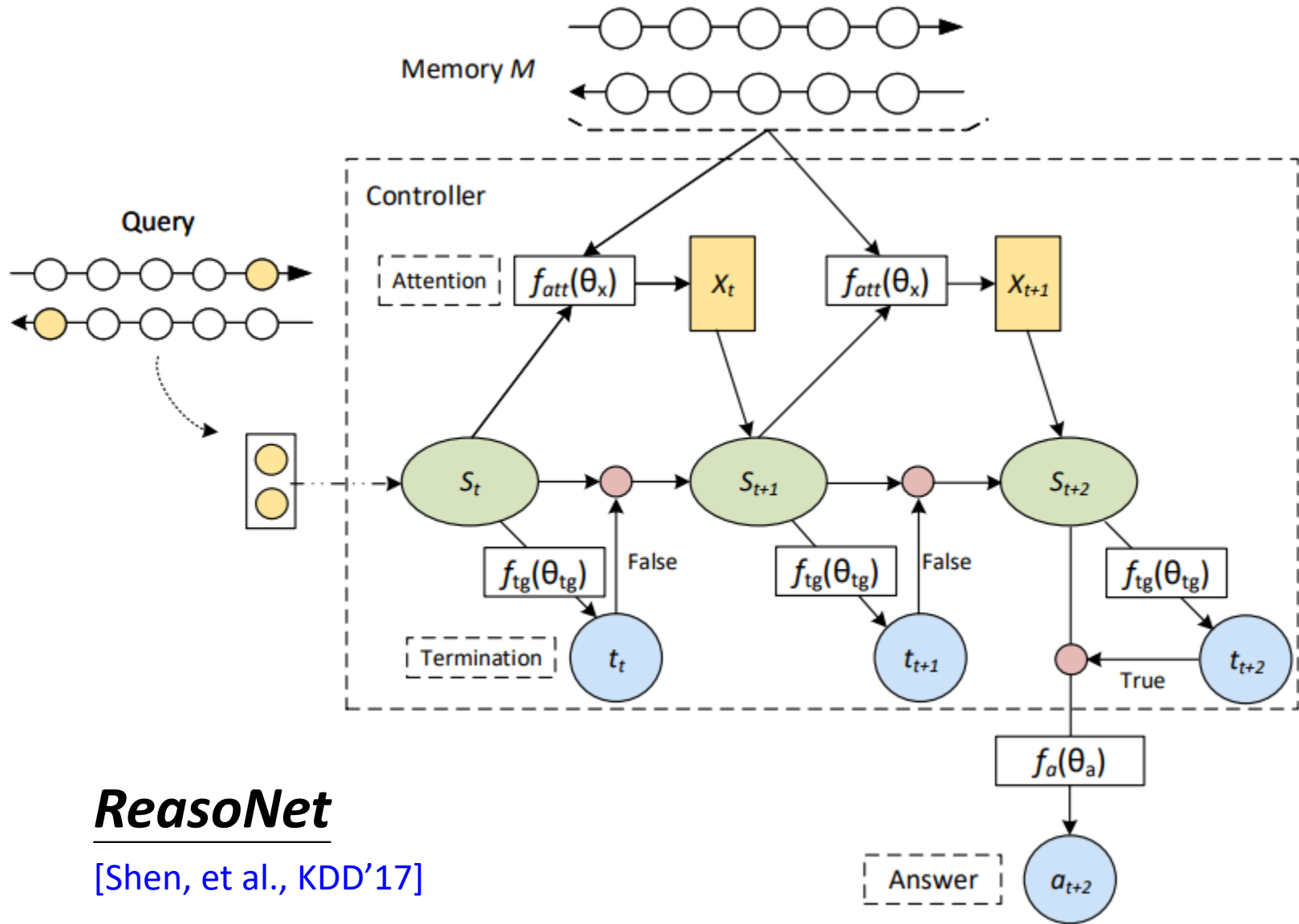
ReasoNet

[Shen, et al., KDD'17]

Graph Neural Network

Input Paragraphs:

The Sum of All Fears is a best-selling thriller novel by Tom Clancy ... It was the fourth of Clancy's Jack Ryan books to be turned into a film ...

Dr. John Patrick Jack Ryan Sr., KCVO (Hon.), Ph.D. is a fictional character created by Tom Clancy who appears in many of his novels and their respective film adaptations ...
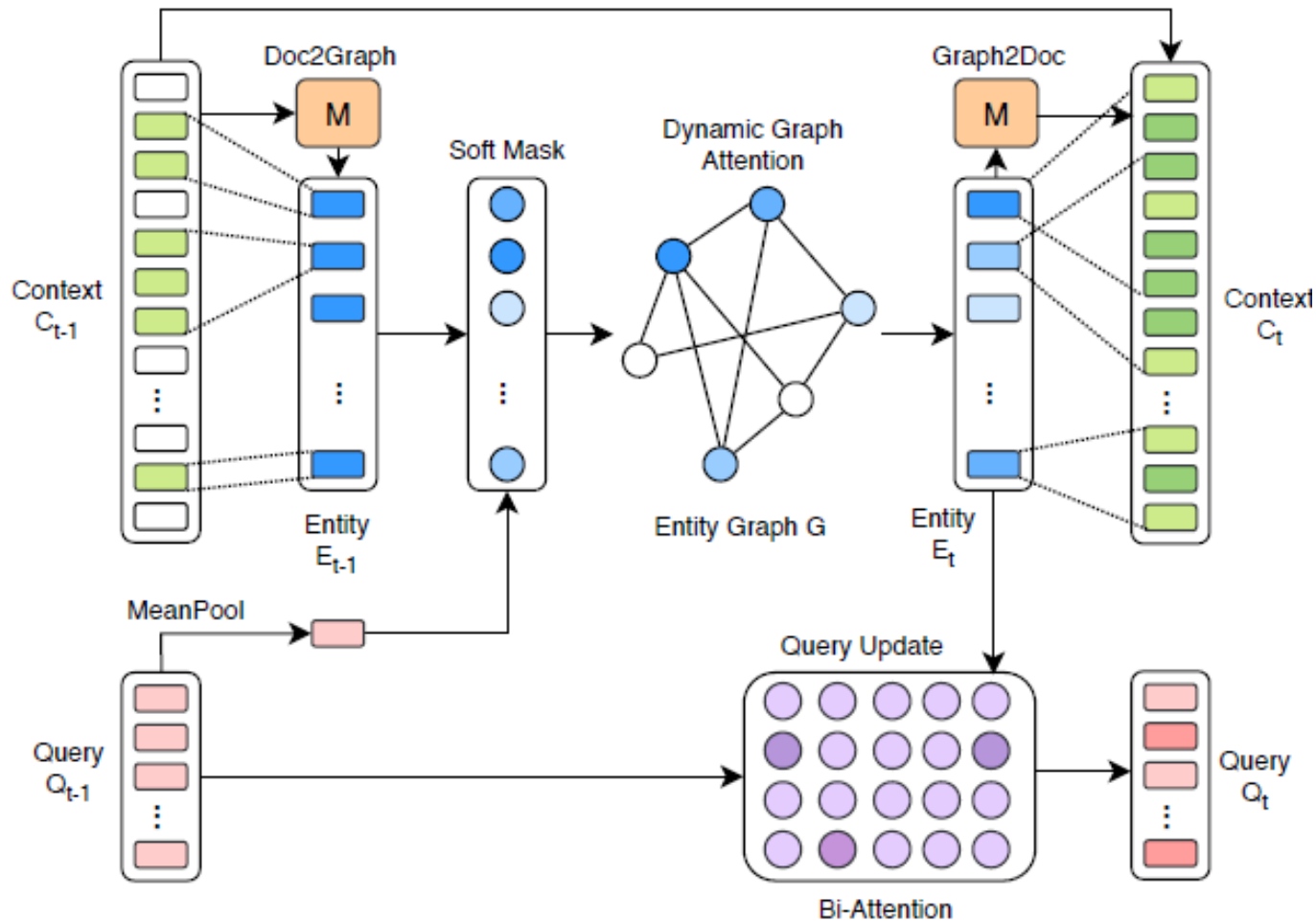
Net Force Explorers is a series of young adult novels created by Tom Clancy and Steve Pieczenik as a spin-off of the military fiction series ...

Question: What fiction character created by Tom Clancy was turned into a film in 2002?

Answer: Jack Ryan

[Qiu, et al., ACL'19]

Graph Neural Network

Graph is not needed ...

[Shao, et al., arXiv'20]

[Qiu, et al., ACL'19]

Types of Questions

Simple Question: Match & Extract

Complex Question: Reasoning

Dialogue QA

Dialogue QA

CoQA

[Reddy, et al., TACL'19]

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q_1: What are the candidates **running** for?
A_1: Governor
R_1: The Virginia governor's race

Q_2: **Where**?
A_2: Virginia
R_2: The Virginia governor's race

Q_3: Who is the democratic candidate?
A_3: **Terry McAuliffe**
R_3: Democrat Terry McAuliffe

Q_4: Who is **his** opponent?
A_4: **Ken Cuccinelli**
R_4 Republican Ken Cuccinelli

Q_5: What party does **he** belong to?
A_5: Republican
R_5: Republican Ken Cuccinelli

Q_6: Which of **them** is winning?
A_6: Terry McAuliffe
R_6: Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May

Dialogue QA

QuAC

[Choi, et al., EMNLP'18]

Section: Daffy Duck, Origin & History

STUDENT: **What is the origin of Daffy Duck?**
TEACHER: ↪ first appeared in Porky's Duck Hunt
STUDENT: **What was he like in that episode?**
TEACHER: ↪ assertive, unrestrained, combative
STUDENT: **Was he the star?**
TEACHER: ↪ No, barely more than an unnamed bit player in this short
STUDENT: **Who was the star?**
TEACHER: ↛ No answer
STUDENT: **Did he change a lot from that first episode in future episodes?**
TEACHER: ↪ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc
STUDENT: **How has he changed?**
TEACHER: ↪ Daffy was less anthropomorphic
STUDENT: **In what other ways did he change?**
TEACHER: ↪ Daffy's slobbery, exaggerated lisp (...) is barely noticeable in the early cartoons.
STUDENT: **Why did they add the lisp?**
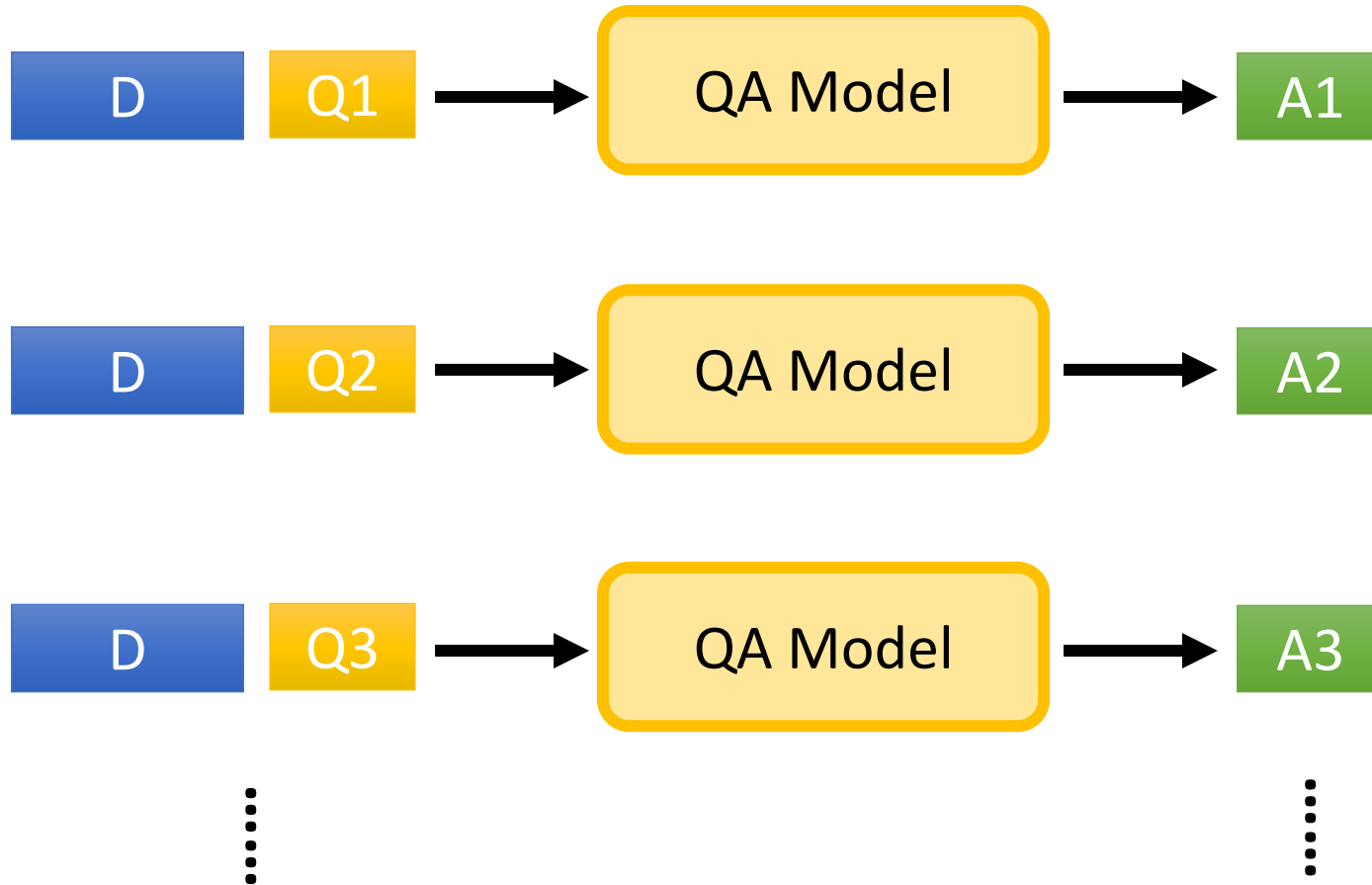TEACHER: ↪ One often-repeated "official" story is that it was modeled after producer Leon Schlesinger's tendency to lisp.
STUDENT: **Is there an "unofficial" story?**
TEACHER: ↪ Yes, Mel Blanc (...) contradicts that conventional belief
. . .

Dialogue QA
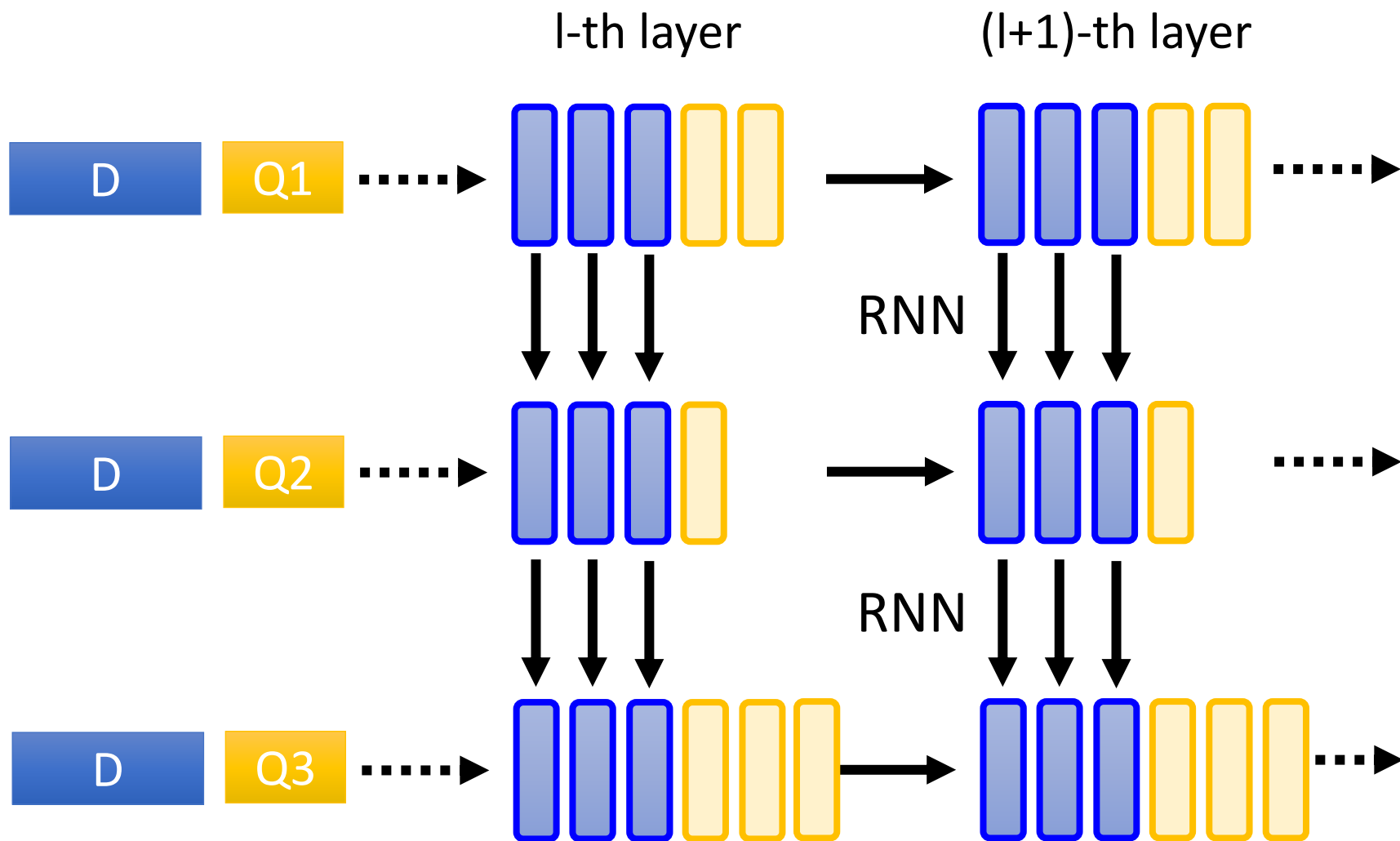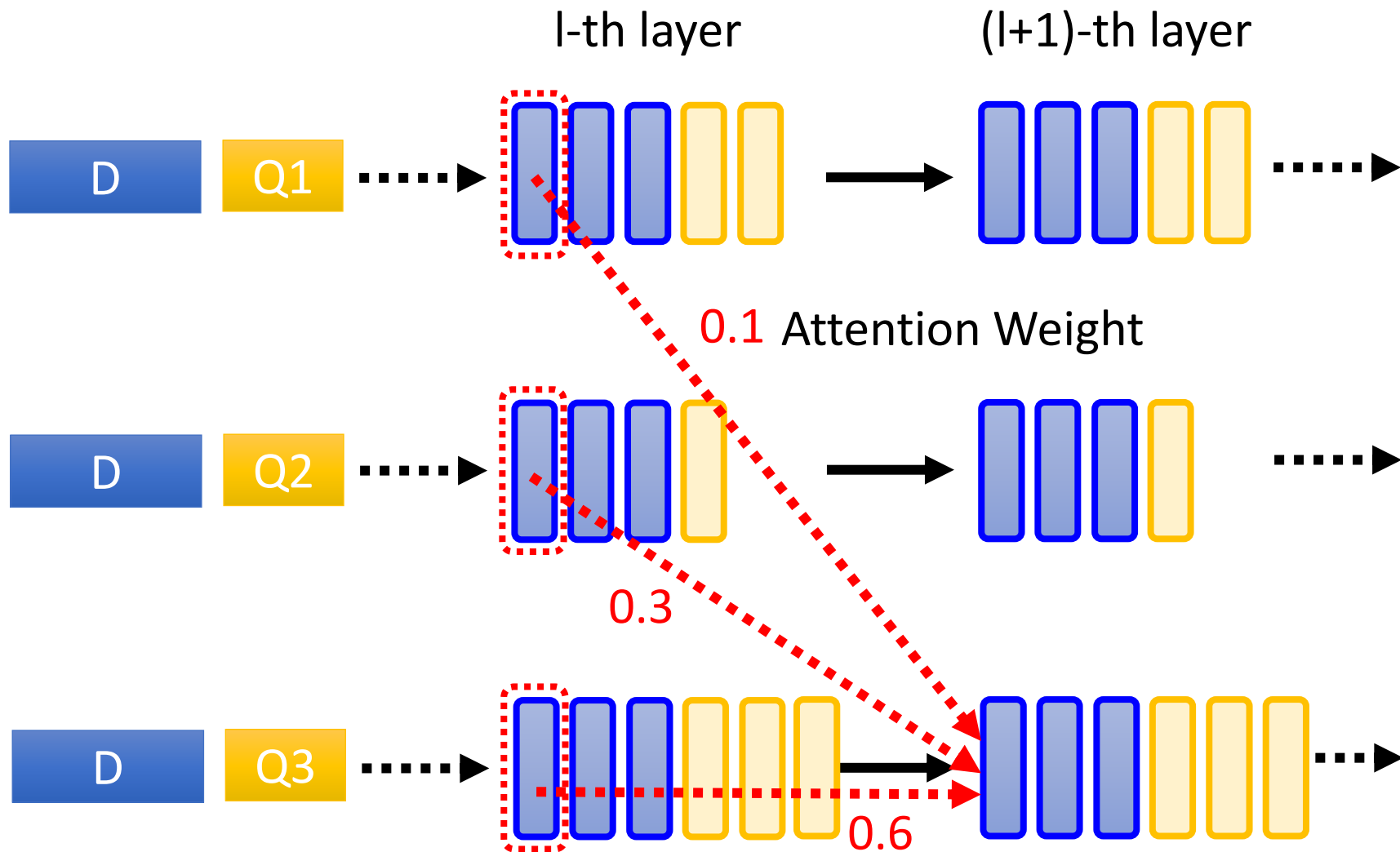
Dialogue QA

l-th layer   (l+1)-th layer

D Q1

D Q2

D Q3

RNN

RNN

Dialogue QA

l-th layer

(l+1)-th layer

D Q1

D Q2

D Q3

0.1 Attention Weight

0.3

0.6

Problem Solved?

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) *QIANXIN*	**90.724**	**93.011**
2 May 05, 2020	SA-Net-V2 (ensemble) *QIANXIN*	90.679	92.948
2 Apr 05, 2020	Retro-Reader (ensemble) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694v2	90.578	92.978
3 May 04, 2020	ELECTRA+ALBERT+EntitySpanFocus (ensemble) *SRCB_DML*	90.442	92.839
4 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) *PINGAN Omni-Sinitic*	90.386	92.777
5 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694v2	90.115	92.580

https://rajpurkar.github.io/SQuAD-explorer/

Train on SQuAD, Test on bAbI

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill
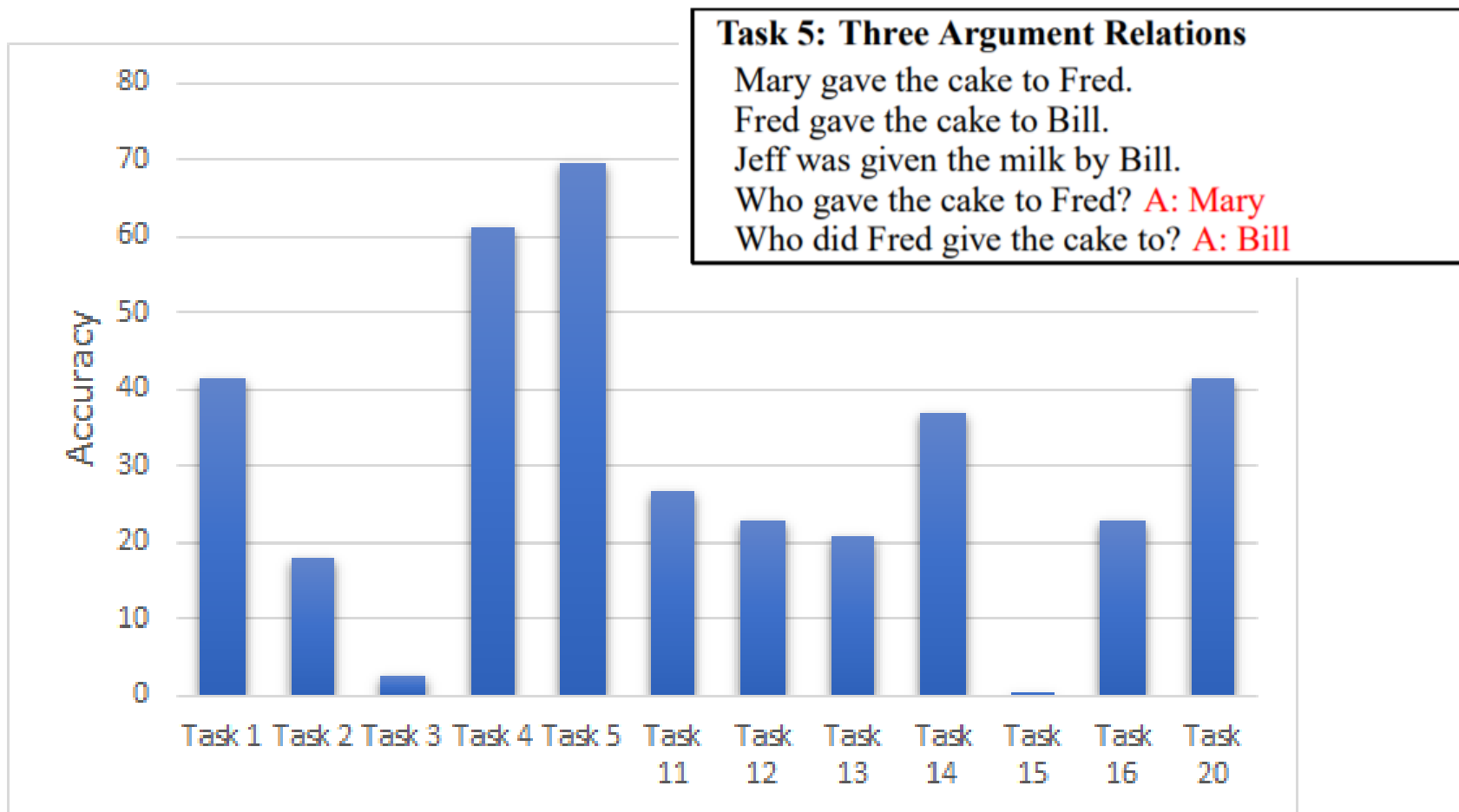
感謝吳侑學同學提供實驗結果

bAbI Tasks 1-10

Dataset	1	2	3	4	5	6	7	8	9	10
True dataset	**100%**	**100%**	39%	**100%**	99%	**100%**	**94%**	**97%**	**99%**	**98%**

bAbI Tasks 11-20

	11	12	13	14	15	16	17	18	19	20
True dataset	**94%**	**100%**	**94%**	**96%**	**100%**	48%	**57%**	**93%**	30%	**100%**

	EM	**70.7%**	0.6%	10.9%	−59.8
	F1	**79.1%**	4.0%	14.8%	−64.3

Table 4: Performance of QANet on SQuAD

CBT-NE	**35.0%**	29
CBT-CN	**37.6%**	32
CBT-V	52.5%	55
CBT-P	55.2%	**56.9%** 30.1% +1.7

[Kaushik, et al., EMNLP'18]

SQuAD

Context	In 1899, John Jacob Astor IV invested $100,000 for Tesla to further develop and produce a new lighting system. Instead, Tesla used the money to fund his Colorado Springs experiments.
Original	What did Tesla spend Astor's money on ?
Reduced	did
Confidence	$0.78 \rightarrow 0.91$

VQA

Original	What color is the flower ?
Reduced	flower ?
Answer	yellow
Confidence	$0.827 \rightarrow 0.819$

機器居然
料事如神!

[Feng, et al., EMNLP'18]
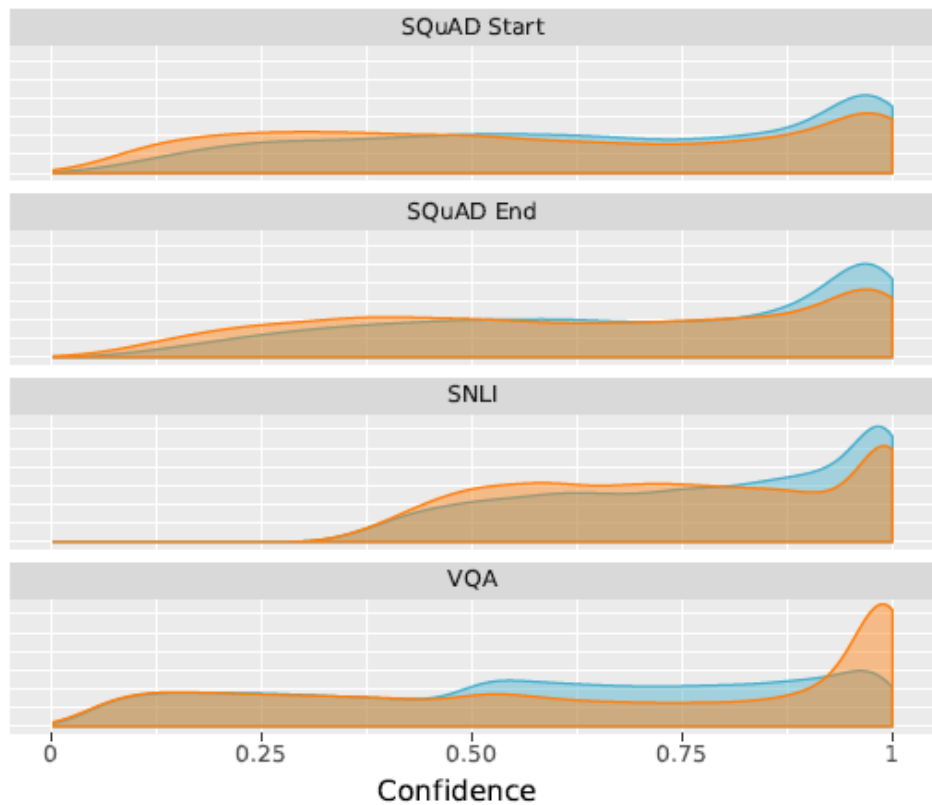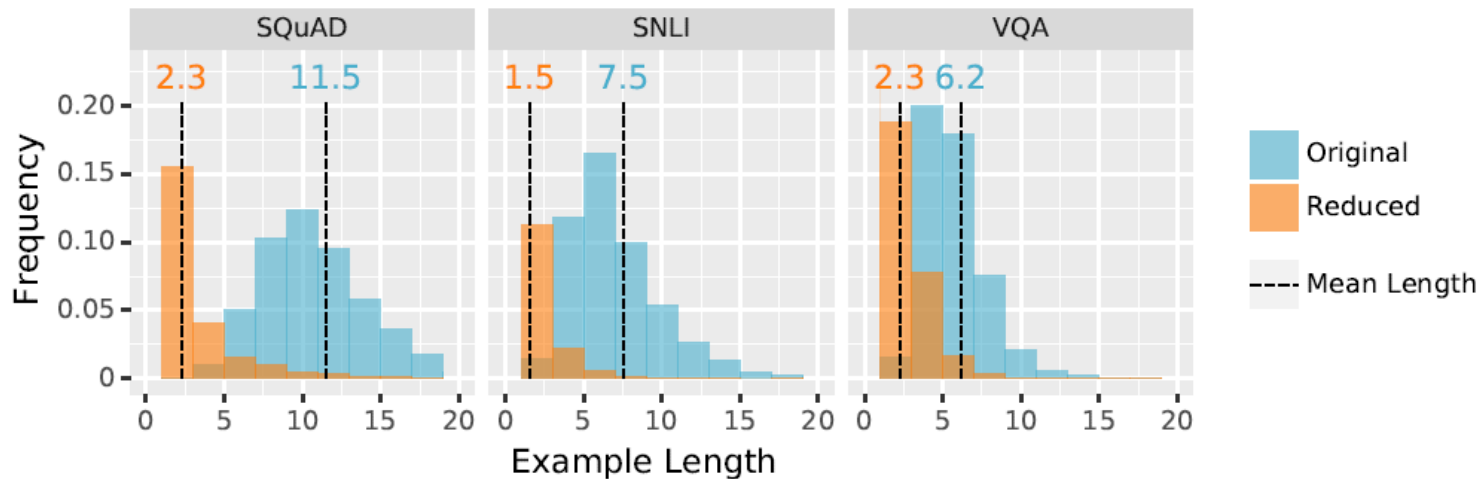
Dataset	Original	Reduced
SQuAD	80.58	31.72
SNLI-E	76.40	27.66
SNLI-N	55.40	52.66
SNLI-C	76.20	60.60
VQA	76.11	40.60

Human Performance

機器超越人類?

VQA   [Agrawal, et al., CVPR'18]
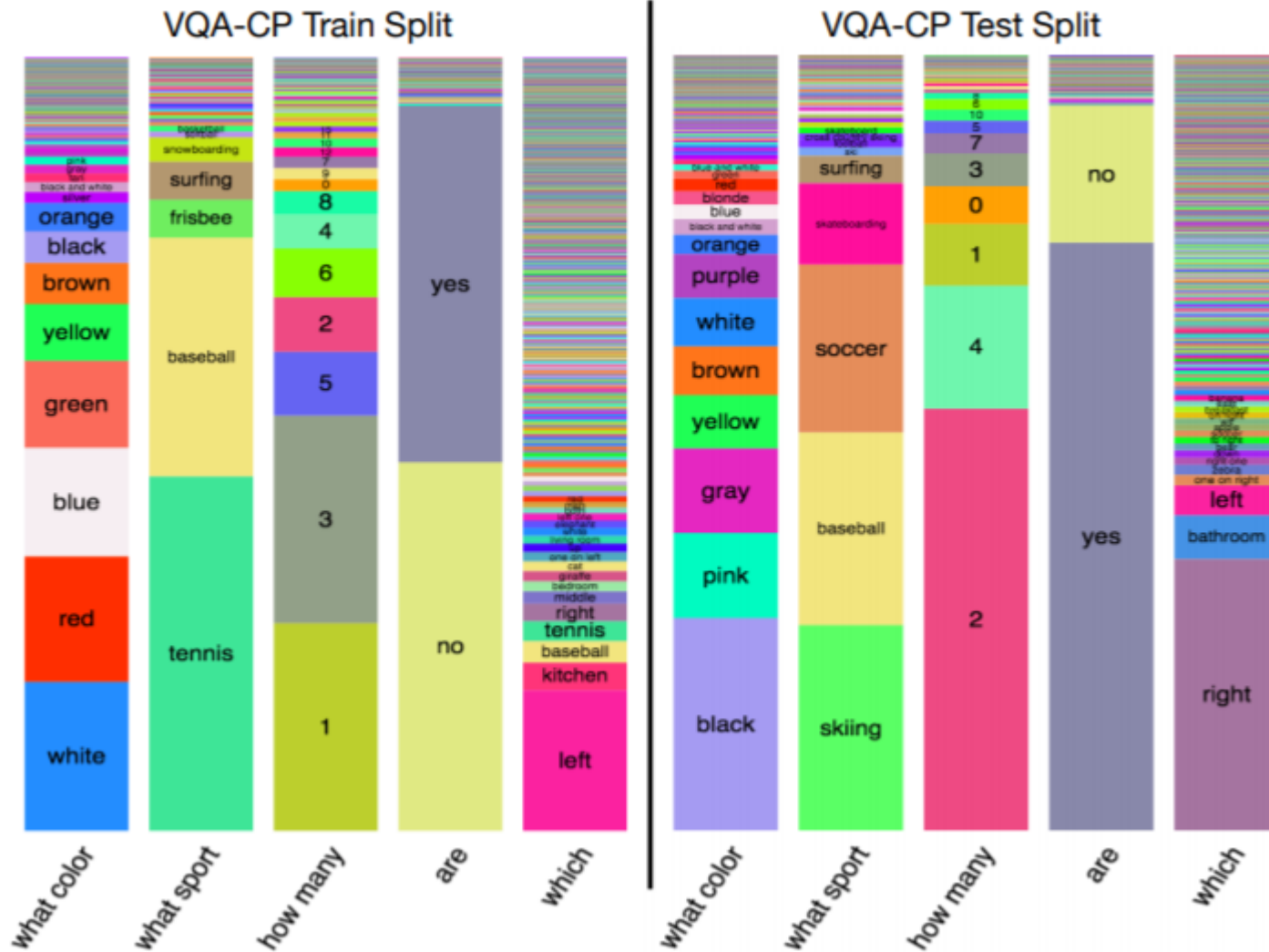
VQA   [Agrawal, et al., CVPR'18]

Model	Dataset	Overall	Yes/No	Number	Other
per Q-type prior [5] (question type only)	VQA v1	35.13	71.31	31.93	08.86
d-LSTM Q [5] (question only)	VQA v1	48.23	79.05	33.70	28.81
d-LSTM Q + norm I [24]	VQA v1	54.40	79.82	33.87	40.54
NMN [3]	VQA v1	54.83	80.39	33.45	41.07
SAN [39]	VQA v1	55.86	78.54	33.46	44.51
MCB [11]	VQA v1	60.97	81.62	34.56	52.16

There is still
a long way to go.

Reference

- Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin and Tomas Mikolov. Towards AI Complete Question Answering: A Set of Prerequisite Toy Tasks, arXiv, 2015

- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang, **SQuAD: 100,000+ Questions for Machine Comprehension of Text, EMNLP, 2016**

- Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang, MS MARCO: A Human Generated MAchine Reading COmprehension Dataset, NIPS, 2016

- Chih Chieh Shao and Trois Liu and Yuting Lai and Yiying Tseng and Sam Tsai, DRCD: a Chinese Machine Reading Comprehension Dataset, arXiv, 2018

- Wei He, Kai Liu, Jing Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Xuan Liu, Tian Wu, Haifeng Wang, **DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications, ACL workshop, 2018**

- Chuanqi Tan, Furu Wei, Nan Yang, Bowen Du, Weifeng Lv, Ming Zhou, **S-Net: From Answer Extraction to Answer Generation for Machine Reading Comprehension, AAAI, 2018**

Reference

- [Chen, et al., ACL'17] Danqi Chen, Adam Fisch, Jason Weston, Antoine Bordes, **Reading Wikipedia to Answer Open-Domain Questions, ACL, 2017**

- Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, Christopher D. Manning, HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering, EMNLP, 2018

- Pranav Rajpurkar, Robin Jia, Percy Liang, **Know What You Don't Know: Unanswerable Questions for SQuAD, ACL, 2018**

- Siva Reddy, Danqi Chen, Christopher D. Manning, CoQA: A Conversational Question Answering Challenge, TACL, 2019

- [Choi, et al., EMNLP'18] Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, Luke Zettlemoyer, **QuAC : Question Answering in Context, EMNLP, 2018**

- [Qiu, et al., ACL'19] Lin Qiu, Yunxuan Xiao, Yanru Qu, Hao Zhou, Lei Li, Weinan Zhang, Yong Yu, Dynamically Fused Graph Network for Multi-hop Reasoning, ACL, 2019

- [Shao, et al., **arXiv'20**] Nan Shao, Yiming Cui, Ting Liu, Shijin Wang, Guoping Hu, **Is Graph Structure Necessary for Multi-hop Reasoning?, arXiv, 2020**

Reference

- Yizhong Wang, Kai Liu, Jing Liu, Wei He, Yajuan Lyu, Hua Wu, Sujian Li, Haifeng Wang, Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification, ACL, 2018

- Bo-Hsiang Tseng, Sheng-syun Shen, Hung-Yi Lee, Lin-Shan Lee, Towards Machine Comprehension of Spoken Content: Initial TOEFL Listening Comprehension Test by Machine, INTERSPEECH, 2016

- Chia-Hsuan Lee, Szu-Lin Wu, Chi-Liang Liu, Hung-yi Lee, Spoken SQuAD: A Study of Mitigating the Impact of Speech Recognition Errors on Listening Comprehension, INTERSPEECH, 2018

- Chia-Hsuan Lee, Yun-Nung Chen, Hung-Yi Lee, "Mitigating the Impact of Speech Recognition Errors on Spoken Question Answering by Adversarial Domain Adaptation", ICASSP, 2019

- Chia-Hsuan Lee, Shang-Ming Wang, Huan-Cheng Chang, Hung-Yi Lee, ODSQA: Open-domain Spoken Question Answering Dataset, SLT, 2018

- Yung-Sung Chuang, Chi-Liang Liu, Hung-Yi Lee, Lin-shan Lee, SpeechBERT: An Audio-and-text Jointly Learned Language Model for End-to-end Spoken Question Answering, arXiv, 2019

Reference

- Tzu-Chien Liu, Yu-Hsueh Wu, Hung-Yi Lee, Query-based Attention CNN for Text Similarity Map, ICCV workshop, 2018

- Hsin-Yuan Huang, Eunsol Choi, Wen-tau Yih, **FlowQA: Grasping Flow in History for Conversational Machine Comprehension, ICLR, 2019**

- Xiaodong Liu, Wei Li, Yuwei Fang, Aerin Kim, Kevin Duh, Jianfeng Gao, **Stochastic Answer Networks for SQuAD 2.0, arXiv, 2018**

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL, 2019

- Minghao Hu, Furu Wei, Yuxing Peng, Zhen Huang, Nan Yang, Dongsheng Li, **Read + Verify: Machine Reading Comprehension with Unanswerable Questions, AAAI, 2019**

- Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, Matt Gardner, **DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs, NAACL, 2019**

- Johannes Welbl, Pontus Stenetorp, Sebastian Riedel, **Constructing Datasets for Multi-hop Reading Comprehension Across Documents, TACL, 2018**

Reference

- Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, Hannaneh Hajishirzi, **Bidirectional Attention Flow for Machine Comprehension, ICLR, 2017**

- Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, Quoc V. Le, **QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension, ICLR, 2018**

- *Hsin-Yuan Huang, Chenguang Zhu, Yelong Shen, Weizhu Chen,* **FusionNet: Fusing via Fully-aware Attention with Application to Machine Comprehension, ICLR, 2018**

- Huijuan Xu, Kate Saenko. Ask, Attend and Answer: Exploring Question-Guided Spatial Attention for Visual Question Answering, CVPR, 2016

- Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus, End-To-End Memory Networks, NIPS, 2015

- Caiming Xiong, Victor Zhong, Richard Socher, **Dynamic Coattention Networks For Question Answering, ICLR, 2017**

- Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, Ming Zhou, Gated Self-Matching Networks for Reading Comprehension and Question Answering, ACL, 2017

Reference

- Chen Qu, Liu Yang, Minghui Qiu, Yongfeng Zhang, Cen Chen, W. Bruce Croft, Mohit Iyyer, **Attentive History Selection for Conversational Question Answering, CIKM, 2019**

- Aishwarya Agrawal, Dhruv Batra, Devi Parikh, Aniruddha Kembhavi, **Don't Just Assume; Look and Answer: Overcoming Priors for Visual Question Answering, CVPR, 2018**

- Shi Feng, Eric Wallace, Alvin Grissom II, Mohit Iyyer, Pedro Rodriguez, Jordan Boyd-Graber, Pathologies of Neural Models Make Interpretations Difficult, EMNLP, 2018

- Divyansh Kaushik, Zachary C. Lipton, **How Much Reading Does Reading Comprehension Require? A Critical Investigation of Popular Benchmarks, EMNLP, 2018**

- Yelong Shen, Po-Sen Huang, Jianfeng Gao, Weizhu Chen, **ReasoNet: Learning to Stop Reading in Machine Comprehension, KDD, 2017**