Question Answering
Hung-yi Lee
李宏毅
Question Answering

Who is the U.S. president?

Is Trump older than Obama?

Who should pay for the date, and why?
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]
Answer is simply a word

Classification problem!
Multiple Choices

Knowledge source question Choice A

Choice B Choice C Choice D

Yes / No

Module for Answer

Module for Source

Knowledge source

Module for Question

question

Module for Choice

Choice A
Span in Source / Extraction-based

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

<table>
<thead>
<tr>
<th>Start Score</th>
<th>0.1</th>
<th>0.1</th>
<th>0.7</th>
<th>0.1</th>
<th>0.0</th>
<th>0.0</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Score</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.6</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Source</td>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$w_4$</td>
<td>$w_5$</td>
<td>$w_6$</td>
<td>$w_7$</td>
</tr>
</tbody>
</table>

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]

<table>
<thead>
<tr>
<th>All words in the answer are in the passage but from multiple text spans.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q</strong>: who did odysseus see in the underworld</td>
</tr>
<tr>
<td><strong>P₁</strong>: The souls that Odysseus saw in the Underworld On seeing <strong>Achilles’</strong> 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.</td>
</tr>
<tr>
<td><strong>P₂</strong>: 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 <strong>Elphenor</strong>, the youngest member of his crew, in the Underworld.</td>
</tr>
<tr>
<td><strong>A</strong>: Elphenor and Achilles.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>All words in the answer are in the passage and question.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q</strong>: what do <strong>producers need to make food</strong></td>
</tr>
<tr>
<td><strong>P</strong>: Plants are producers. Producers are living things that can make their own food using <strong>air</strong>, <strong>light</strong>, <strong>soil</strong>, and <strong>water</strong>. Plants use a process called photosynthesis to make food.</td>
</tr>
<tr>
<td><strong>A</strong>: Producers need air, light, soil, and water to make food.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part of words in the answer are not found in the passage or question.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q</strong>: why conversion observed in body</td>
</tr>
<tr>
<td><strong>P</strong>: Conversion disorder <strong>symptoms</strong> 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.</td>
</tr>
<tr>
<td><strong>A</strong>: Due to symptoms in the body</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>All Words in the answer are not found in the passages or question.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q</strong>: is there an age limit for learning speech</td>
</tr>
<tr>
<td><strong>P</strong>: 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.</td>
</tr>
<tr>
<td><strong>A</strong>: No</td>
</tr>
</tbody>
</table>
S-net
[Tan, et al., AAAI’18]
並不是所有問題都有答案......
這就是正確的答案。

沉默！

評審答對，他答對。

可是，剛剛那個人：

過去，只是要他。

就這樣，不能回答。

就是，他答對。

評審答對，他答對。

可是，剛剛那個人：

過去，只是要他。

就這樣，不能回答。

就是，他答對。

評審答對，他答對。

可是，剛剛那個人：

過去，只是要他。

就這樣，不能回答。

就是，他答對。

評審答對，他答對。

可是，剛剛那個人：

過去，只是要他。

就這樣，不能回答。

就是，他答對。
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]
Know what you don’t know

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

[Devlin, et al., NAACL’19]
Knowledge source

**Question Answering**

- Module for Source
- Module for Question
- Module for Answer

**Attention**

- Who is the U.S. president?
- Is Trump older 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?


[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

• 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 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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>QA-pairs</th>
<th>Hours</th>
<th>M-spkrs</th>
<th>F-spkrs</th>
<th>WER-D(%)</th>
<th>WER-Q(%)</th>
<th>Avg D Len</th>
<th>AvgQ Len</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODSQA</td>
<td>3654</td>
<td>25.28</td>
<td>7</td>
<td>13</td>
<td>19.11</td>
<td>18.57</td>
<td>428</td>
<td>22</td>
</tr>
<tr>
<td>DRCD-TTS</td>
<td>16746</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>33.63</td>
<td>--</td>
<td>332</td>
<td>20</td>
</tr>
</tbody>
</table>

SPOKEN OPEN-DOMAIN QUESTION ANSWERING DATASET
Audio – SQuAD Style

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>QA-pairs</th>
<th>Hours</th>
<th>M-spkrs</th>
<th>F-spkrs</th>
<th>WER-D(%)</th>
<th>WER-Q(%)</th>
<th>Avg D Len</th>
<th>AvgQ Len</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODSQA</td>
<td>3654</td>
<td>25.28</td>
<td>7</td>
<td>13</td>
<td>19.11</td>
<td>18.57</td>
<td>428</td>
<td>22</td>
</tr>
<tr>
<td>DRCD-TTS</td>
<td>16746</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>33.63</td>
<td>--</td>
<td>332</td>
<td>20</td>
</tr>
</tbody>
</table>

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

Adversarial Learning

[Lee, et al., INTERSPEECH’18]

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

[Chuang, et al., arXiv’19]
Movie QA

QACNN

[Liu, et al., ICCV workshop’18]

[Tapaswi, et al., CVPR’16]
**Question Answering**

- **Module for Source**
- **Module for Question**
- **Module for Answer**
- **Knowledge source**
- **Attention**

**Questions:**

- Who is the U.S. president?
- Is Trump older 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

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?
graupe

Where do water droplets collide with ice crystals to form precipitation?
within a cloud
Query-to-context Attention

\[
\sum_{n=1}^{5} \alpha_n x^n
\]

Extract

\[\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\]

Attention

\[x^n\]

Module for Source

Knowledge source

Match

Module for Question

merge the whole question into one vector

Classifier?
Predicting time span?

Module for Answer

Predicting time span?
Query-to-context Attention

End-to-end Memory Network

[Sukhbaatar, et al., NIPS’15]
Query-to-context Attention (v2)

\[
\begin{align*}
\max & \quad \alpha_1^3 \quad \alpha_2^3 \quad \alpha_3^3 \\
& \quad \quad \alpha_2^2 \quad \alpha_3^2 \quad \alpha_4^2 \\
& \quad \quad \alpha_1^1 \quad \alpha_2^1 \quad \alpha_3^1 \\
& \quad \quad \alpha_2^1 \quad \alpha_3^1 \quad \alpha_4^1 \\
\end{align*}
\]

\[
\sum_{n=1}^{5} \alpha_n x^n
\]

Module for Source

Attention

Module for Question

Knowledge source

question

multiple vectors
Query-to-context Attention (v2)

Is there a cat in the basket?

[Xu, et al., CVPR’16]
Context-to-query

Attention
R-Net

Answer Prediction

Question-Passage Matching

Passage Self-Matching

Self-attention

[Wang, et al., ACL’17]
The above can be used to capture long range info.

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

- BiDAF  [Seo, et al., ICLR ’17]
Dynamic Coattention Networks

[Xiong, et al., ICLR’17]
QANet

BERT 家族之外的最後一道榮光

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
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, ...}
<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>%</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| 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 Discrete Reasoning Over the text in the Paragraph

[Dua, et al., NAACL’19]

<table>
<thead>
<tr>
<th>Reasoning</th>
<th>Passage (some parts shortened)</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtraction (28.8%)</td>
<td>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.</td>
<td>How many more dollars was the Untitled (1981) painting sold for than the 12 million dollar estimation?</td>
<td>4300000</td>
</tr>
<tr>
<td>Comparison (18.2%)</td>
<td>In 1517, the seventeen-year-old King sailed to Castile. There, his Flemish court . . . In May 1518, Charles traveled to Barcelona in Aragon.</td>
<td>Where did Charles travel to first, Castile or Barcelona?</td>
<td>Castile</td>
</tr>
<tr>
<td>Selection (19.4%)</td>
<td>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.</td>
<td>Who was the University professor that helped produce The Ballad Of Black Jack, Ivan Boyd or Don Mueller?</td>
<td>Don Mueller</td>
</tr>
<tr>
<td>Addition (11.7%)</td>
<td>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.</td>
<td>What date did the JNA form a battlegroup to counterattack after the village of Nos Kalik was captured?</td>
<td>3 March 1992</td>
</tr>
<tr>
<td>Count (16.5%) and Sort (11.7%)</td>
<td>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.</td>
<td>Which kicker kicked the most field goals?</td>
<td>John Kasay</td>
</tr>
</tbody>
</table>
Multiple-hop

Match → Extract ("frog") → Change

Match → Extract ("Brian") → Change

Match → Extract ("yellow") → ......

<table>
<thead>
<tr>
<th>Story (16: basic induction)</th>
<th>Support</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brian is a frog.</td>
<td>yes</td>
<td>0.00</td>
<td>0.98</td>
<td>0.00</td>
</tr>
<tr>
<td>Lily is gray.</td>
<td></td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Brian is yellow.</td>
<td>yes</td>
<td>0.07</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Julius is green.</td>
<td></td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Greg is a frog.</td>
<td>yes</td>
<td>0.76</td>
<td>0.02</td>
<td>0.00</td>
</tr>
</tbody>
</table>

What color is Greg? Answer: yellow Prediction: yellow
Memory Network

Extract

Match

$\sum$

Change what to match

Extract

Match

$\sum$

Change what to 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

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 ...

Q1: What are the candidates running for?
A1: Governor
R1: The Virginia governor’s race

Q2: Where?
A2: Virginia
R2: The Virginia governor’s race

Q3: Who is the democratic candidate?
A3: Terry McAuliffe
R3: Democrat Terry McAuliffe

Q4: Who is his opponent?
A4: Ken Cuccinelli
R4 Republican Ken Cuccinelli

Q5: What party does he belong to?
A5: Republican
R5: Republican Ken Cuccinelli

Q6: Which of them is winning?
A6: Terry McAuliffe
R6: 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]
Dialogue QA

D  Q1  →  QA Model  →  A1

D  Q2  →  QA Model  →  A2

D  Q3  →  QA Model  →  A3

...  ...  ...
Dialogue QA

[Huang, et al., ICLR’19]
Dialogue QA

D Q1 ----> I-th layer

0.1 Attention Weight

D Q2 ----> I-th layer

0.3

D Q3 ----> I-th layer

0.6

D Q1 ----> (I+1)-th layer

D Q2 ----> (I+1)-th layer

D Q3 ----> (I+1)-th layer

[Qu, et al., CIKM’19]
Problem Solved?
<table>
<thead>
<tr>
<th>Rank</th>
<th>Model</th>
<th>EM</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Human Performance</td>
<td>86.831</td>
<td>89.452</td>
</tr>
<tr>
<td></td>
<td><em>Stanford University</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(Rajpurkar &amp; Jia et al. '18)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SA-Net on Albert (ensemble)</td>
<td>90.724</td>
<td>93.011</td>
</tr>
<tr>
<td></td>
<td>QIANXIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SA-Net-V2 (ensemble)</td>
<td>90.679</td>
<td>92.948</td>
</tr>
<tr>
<td></td>
<td>QIANXIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Retro-Reader (ensemble)</td>
<td>90.578</td>
<td>92.978</td>
</tr>
<tr>
<td></td>
<td><em>Shanghai Jiao Tong University</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ELECTRA+ALBERT+EntitySpanFocus (ensemble)</td>
<td>90.442</td>
<td>92.839</td>
</tr>
<tr>
<td></td>
<td>SRCB_DML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ALBERT + DAAF + Verifier (ensemble)</td>
<td>90.386</td>
<td>92.777</td>
</tr>
<tr>
<td></td>
<td>PINGAN Omni-Sinitic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Retro-Reader on ALBERT (ensemble)</td>
<td>90.115</td>
<td>92.580</td>
</tr>
<tr>
<td></td>
<td><em>Shanghai Jiao Tong University</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Thank you, Wei Hsiun, for providing the experimental results.
### bAbI Tasks 1-10

<table>
<thead>
<tr>
<th>Dataset</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>True dataset</td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td>39%</td>
<td><strong>100%</strong></td>
<td><strong>99%</strong></td>
<td><strong>100%</strong></td>
<td><strong>94%</strong></td>
<td><strong>97%</strong></td>
<td><strong>99%</strong></td>
<td><strong>98%</strong></td>
</tr>
</tbody>
</table>

### bAbI Tasks 11-20

<table>
<thead>
<tr>
<th>Dataset</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>True dataset</td>
<td><strong>94%</strong></td>
<td><strong>100%</strong></td>
<td><strong>94%</strong></td>
<td><strong>96%</strong></td>
<td><strong>100%</strong></td>
<td><strong>48%</strong></td>
<td><strong>57%</strong></td>
<td><strong>93%</strong></td>
<td><strong>30%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

### Key-Value Model Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>EM</th>
<th>F1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CBT-NE</td>
<td>35.0%</td>
<td>29.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT-CN</td>
<td>37.6%</td>
<td>32.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT-V</td>
<td>52.5%</td>
<td>55.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT-P</td>
<td>55.2%</td>
<td>56.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>EM</th>
<th>F1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EM</td>
<td><strong>70.7%</strong></td>
<td>0.6%</td>
<td><strong>10.9%</strong></td>
<td><strong>−59.8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td><strong>79.1%</strong></td>
<td>4.0%</td>
<td><strong>14.8%</strong></td>
<td><strong>−64.3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Performance of QANet on SQuAD

[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**.

<table>
<thead>
<tr>
<th>Original</th>
<th>What did Tesla spend Astor’s money on?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced</td>
<td>did</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.78 → 0.91</td>
</tr>
</tbody>
</table>

### VQA

![Image of a sunflower in a vase]

<table>
<thead>
<tr>
<th>Original</th>
<th>What color is the flower?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced</td>
<td>flower?</td>
</tr>
<tr>
<td>Answer</td>
<td>yellow</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.827 → 0.819</td>
</tr>
</tbody>
</table>

[Feng, et al., EMNLP’18]
Human Performance

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Original</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQuAD</td>
<td>80.58</td>
<td>31.72</td>
</tr>
<tr>
<td>SNLI-E</td>
<td>76.40</td>
<td>27.66</td>
</tr>
<tr>
<td>SNLI-N</td>
<td>55.40</td>
<td>52.66</td>
</tr>
<tr>
<td>SNLI-C</td>
<td>76.20</td>
<td>60.60</td>
</tr>
<tr>
<td>VQA</td>
<td>76.11</td>
<td>40.60</td>
</tr>
</tbody>
</table>

Human Performance
<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset</th>
<th>Overall</th>
<th>Yes/No</th>
<th>Number</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>per Q-type prior [5]</td>
<td>VQA v1</td>
<td>35.13</td>
<td>71.31</td>
<td>31.93</td>
<td>08.86</td>
</tr>
<tr>
<td>d-LSTM Q [5]</td>
<td>VQA v1</td>
<td>48.23</td>
<td>79.05</td>
<td>33.70</td>
<td>28.81</td>
</tr>
<tr>
<td>d-LSTM Q + norm I [24]</td>
<td>VQA v1</td>
<td>54.40</td>
<td>79.82</td>
<td>33.87</td>
<td>40.54</td>
</tr>
<tr>
<td>MNM [3]</td>
<td>VQA v1</td>
<td>54.83</td>
<td>80.39</td>
<td>33.45</td>
<td>41.07</td>
</tr>
<tr>
<td>SAN [39]</td>
<td>VQA v1</td>
<td>55.86</td>
<td>78.54</td>
<td>33.46</td>
<td>44.51</td>
</tr>
<tr>
<td>MCB [11]</td>
<td>VQA v1</td>
<td>60.97</td>
<td>81.62</td>
<td>34.56</td>
<td>52.16</td>
</tr>
</tbody>
</table>
There is still a long way to go.
Reference


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


- Chih Chieh Shao and Trois Liu and Yuting Lai and Yiyiing 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 Ly, Ming Zhou. S-Net: From Answer Extraction to Answer Generation for Machine Reading Comprehension, AAAI, 2018
Reference


• 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

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


• 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

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

- 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
- 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