An overview of Question Answering

MLIA Language group
17/01/2019

Different types of datasets

Open-domain VS contextual QA

Open-domain QA: model provided with the question only

→ mainly a *retrieval* task inside a knowledge base

How much reasoning involved?

→ concentrate on QA about a visual or textual extract, i.e. **contextual QA**

Reading comprehension task

Generative QA

The NarrativeQA Reading Comprehension Challenge, Kočiský et al. 2017

+ MS MARCO

Title: Ghostbusters II

Question: How is Oscar related to Dana?

Answer: her son

Summary snippet: …Peter's former girlfriend Dana Barrett has had a son, Oscar…

Story snippet:

DANA (setting the wheel brakes on the buggy)
Thank you, Frank. I'll get the hang of this eventually.

She continues digging in her purse while Frank leans over the buggy and makes funny faces at the baby, OSCAR, a very cute nine-month old boy.

FRANK (to the baby)
Hiya, Oscar. What do you say, slugger?

FRANK (to Dana)
That's a good-looking kid you got there, Ms. Barrett.

Snippets are not provided at test time, only the *full* script

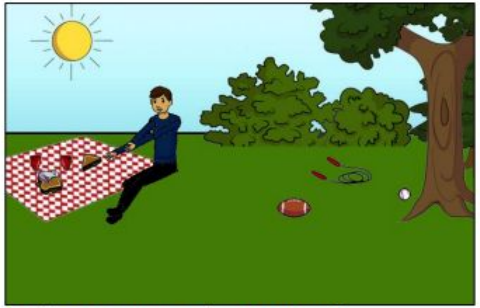
VQA

VQA: Visual Question Answering, Agrawal et al. 2015 (ICCV)

+ others

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

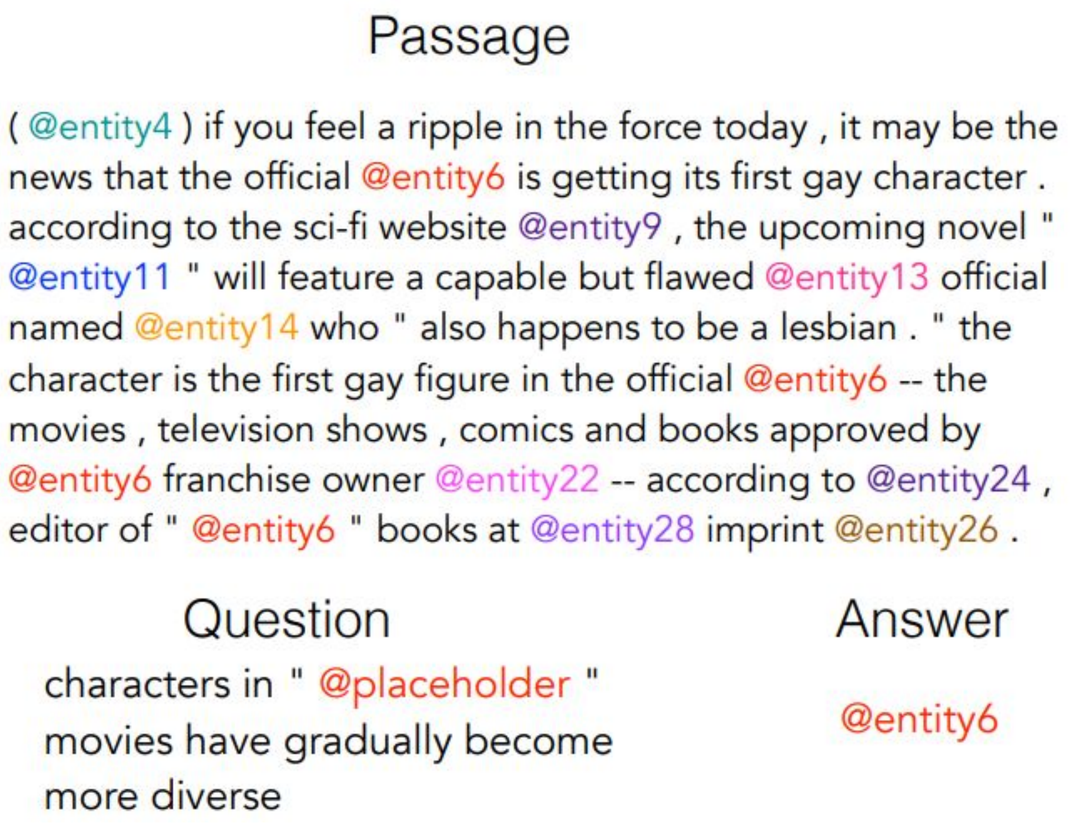
Does it appear to be rainy?
Does this person have 20/20 vision?

Cloze-style

CNN & Daily Mail dataset

from *Teaching Machines to Read and Comprehend,* Hermann et al. 2015 (NeurIPS)

Picture from *A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task*, Chen et al. 2016



Multiple choice

RACE dataset

RACE: Large-scale ReAding Comprehension Dataset From Examinations, Lai et al. 2017

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.
Alice looked at the envelope for a minute, and then handed it back to the mailman.
"I'm sorry I can't take it, I don't have enough money to pay it", she said.
A gentleman standing around were very sorry for her. Then he came up and paid the postage for her.
When the gentleman gave the letter to her, she said with a smile, " Thank you very much, This letter is from Tom. I'm going to marry him. He went to London to look for work. I've waited a long time for this letter, but now I don't need it, there is nothing in it."
"Really? How do you know that?" the gentleman said in surprise.
"He told me that he would put some signs on the envelope. Look, sir, this cross in the corner means that he is well and this circle means he has found work. That's good news."
The gentleman was Sir Rowland Hill. He didn't forgot Alice and her letter.
"The postage to be paid by the receiver has to be changed," he said to himself and had a good plan.
"The postage has to be much lower, what about a penny? And the person who sends the letter pays the postage. He has to buy a stamp and put it on the envelope." he said . The government accepted his plan. Then the first stamp was put out in 1840. It was called the "Penny Black". It had a picture of the Queen on it.

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

Extractive QA

SQuAD: 100,000+ Questions for Machine Comprehension of Text, Rajpurkar et al. 2016

SQuAD 2, HotpotQA, QuAC, SWAG

The task is much more restricted

Evaluation with *exact match* and token F1

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

What causes precipitation to fall?
gravity

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

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

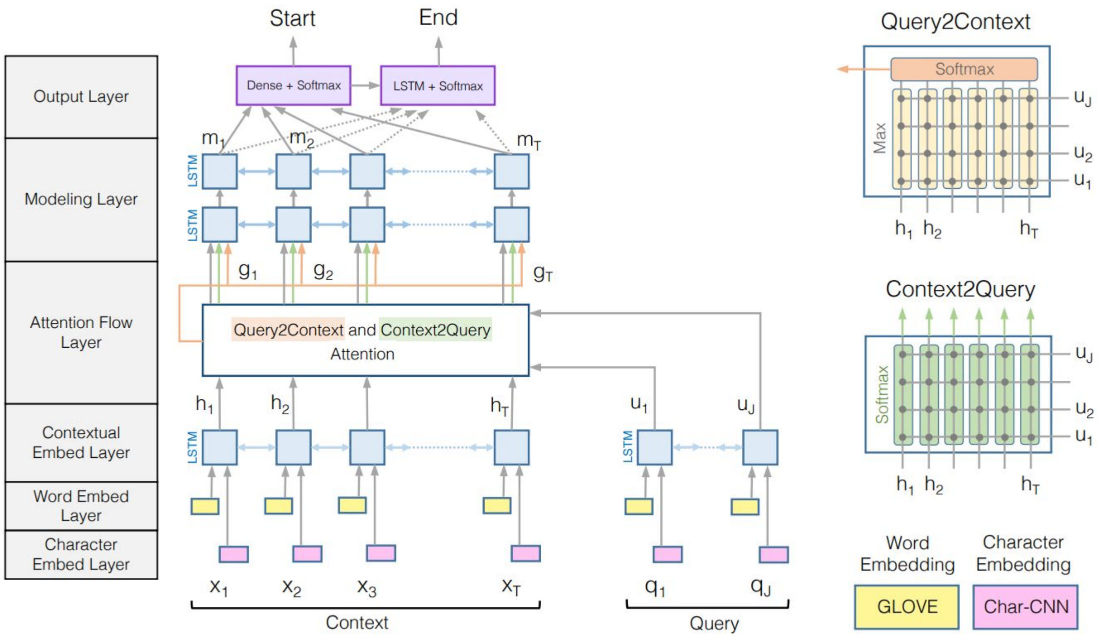
Reading comprehension models

Representation learning at its finest

Learn a **enhanced context representation** and add two linear layers

Originality of this work is that attention is used for both context and query

Picture from *Bidirectional Attention Flow for Machine Comprehension*, Seo et al. 2017 (ICLR)



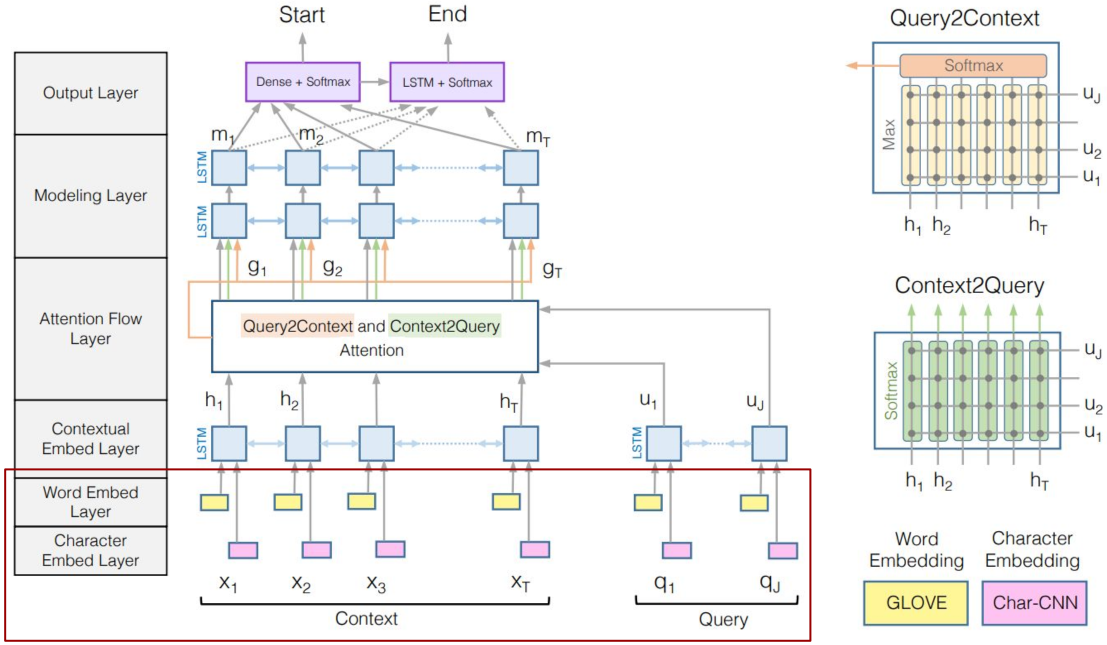
The BiDAF model learns attentive Question & Context representations that are combined and passed into a BiLSTM

Representation learning at its finest

Learn a **enhanced context representation** and add two linear layers

Originality of this work is that attention is used for both context and query

Picture from *Bidirectional Attention Flow for Machine Comprehension*, Seo et al. 2017 (ICLR)



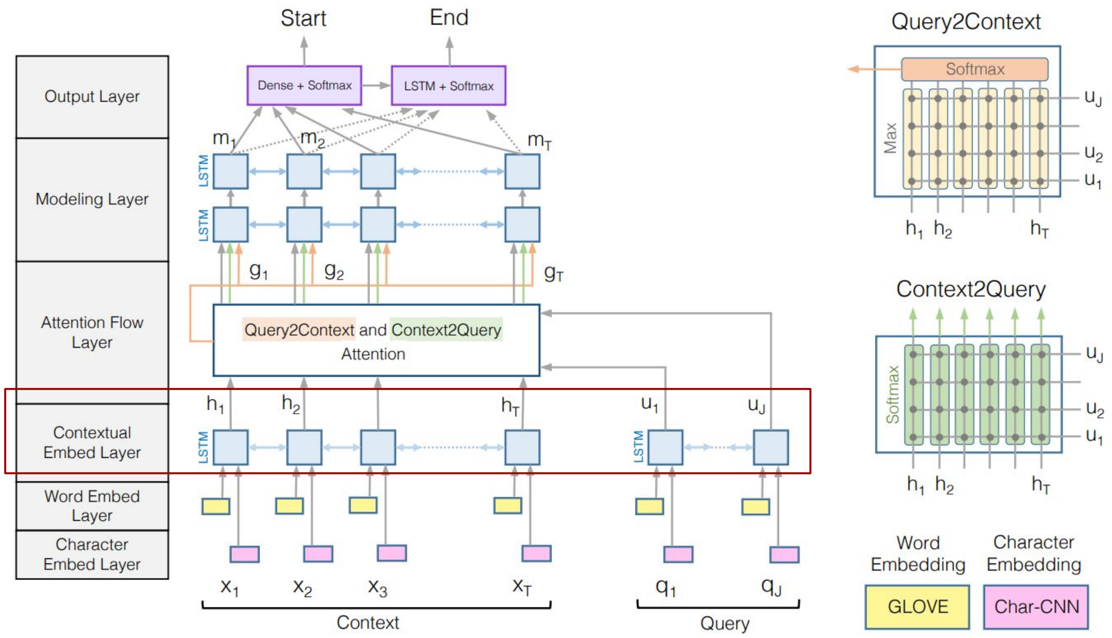
The BiDAF model learns attentive Question & Context representations that are combined and passed into a BiLSTM

Representation learning at its finest

Learn a **enhanced context representation** and add two linear layers

Originality of this work is that attention is used for both context and query

Picture from *Bidirectional Attention Flow for Machine Comprehension*, Seo et al. 2017 (ICLR)



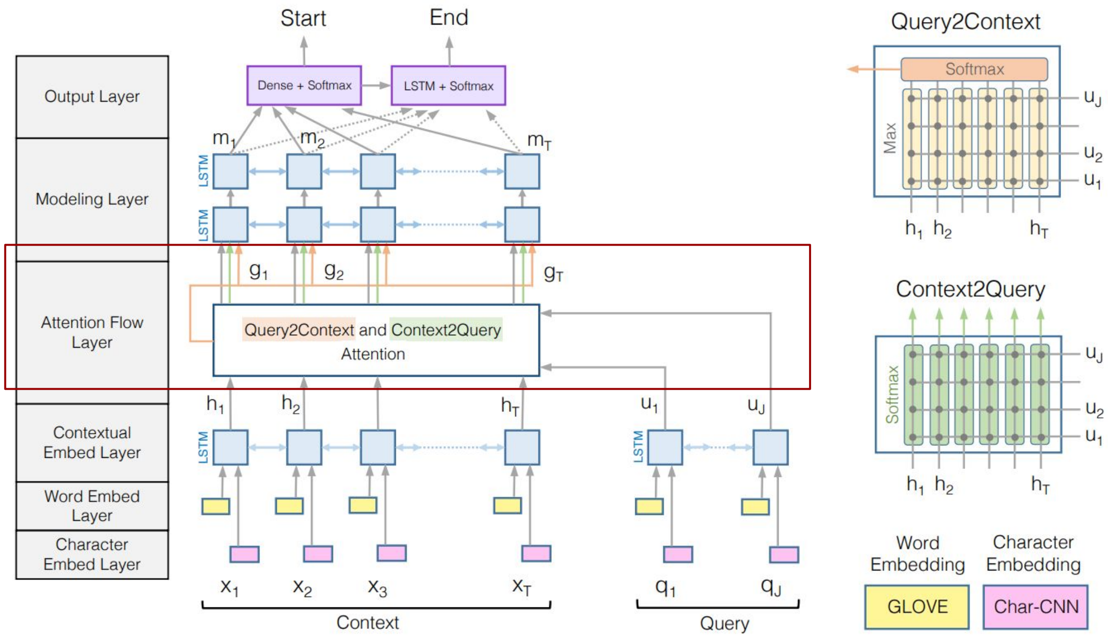
The BiDAF model learns attentive Question & Context representations that are combined and passed into a BiLSTM

Representation learning at its finest

Learn a **enhanced context representation** and add two linear layers

Originality of this work is that attention is used for both context and query

Picture from *Bidirectional Attention Flow for Machine Comprehension*, Seo et al. 2017 (ICLR)



$$\alpha(h, u) = w^T [h; u; h \circ u]$$

Then compute a attentive context matrix U’ and question matrix H’ with the weights α

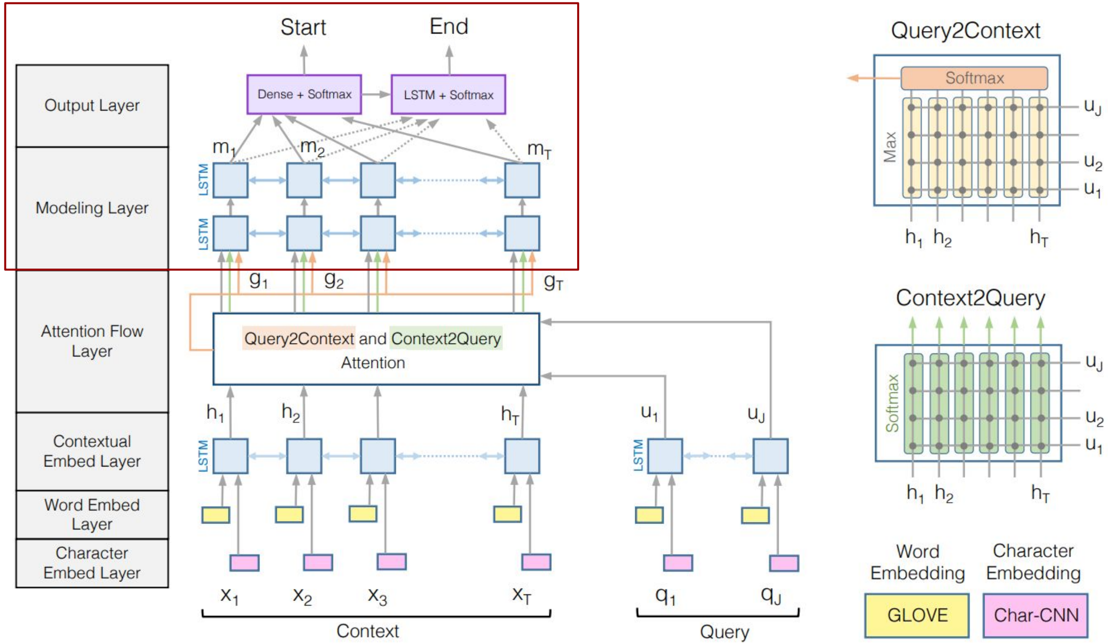
$$\mathbf{G}_{:t} = \beta(\mathbf{H}_{:t}, \mathbf{U}'_{:t}, \mathbf{H}'_{:t})$$

Representation learning at its finest

Learn a **enhanced context representation** and add two linear layers

Originality of this work is that attention is used for both context and query

Picture from *Bidirectional Attention Flow for Machine Comprehension*, Seo et al. 2017 (ICLR)



$$\alpha(h, u) = w^T[h; u; h \circ u]$$

Then compute a attentive context matrix U’ and question matrix H’ with the weights α

$$\mathbf{G}_{:t} = \beta(\mathbf{H}_{:t}, \mathbf{U}'_{:t}, \mathbf{H}'_{:t})$$

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Change the type of attention

- Dot product / bilinear and add self-attention

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Change the type of attention

- Dot product / bilinear and add self-attention
- Co-attention (*Dynamic coattention networks for question answering*, Xiong et al. 2016 (ICLR))

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Change the type of attention

- Dot product / bilinear and add self-attention
- Co-attention (*Dynamic coattention networks for question answering*, Xiong et al. 2016 (ICLR))
- Iteratively compute attention (*Reinforced Mnemonic Reader for Machine Reading Comprehension*, Hu et al. 2017 (IJCAI))

Representation learning at its finest

Since then, most models are variants of this idea

<u>Enhance embeddings</u>

- POS / NER information + add ELMo or CoVe

<u>Change the type of attention</u>

- Dot product / bilinear and add self-attention
- Co-attention (*Dynamic coattention networks for question answering*, Xiong et al. 2016 (ICLR))
- Iteratively compute attention (*Reinforced Mnemonic Reader for Machine Reading Comprehension*, Hu et al. 2017 (IJCAI))

- Use CNN (*QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension*, Yu et al. 2018)

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Change the type of attention

- Dot product / bilinear and add self-attention
- Co-attention (*Dynamic coattention networks for question answering*, Xiong et al. 2016 (ICLR))
- Iteratively compute attention (*Reinforced Mnemonic Reader for Machine Reading Comprehension*, Hu et al. 2017 (IJCAI))

- Use CNN (*QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension*, Yu et al. 2018)

- “Ensemble” outputs from a Memory Network (*Stochastic Answer Networks for Machine Reading Comprehension*, Liu et al. 2017 (ACL))

Representation learning at its finest

Since then, most models are variants of this idea

Enhance embeddings

- POS / NER information + add ELMo or CoVe

Change the type of attention

- Dot product / bilinear and add self-attention
- Co-attention (*Dynamic coattention networks for question answering*, Xiong et al. 2016 (ICLR))
- Iteratively compute attention (*Reinforced Mnemonic Reader for Machine Reading Comprehension*, Hu et al. 2017 (IJCAI))

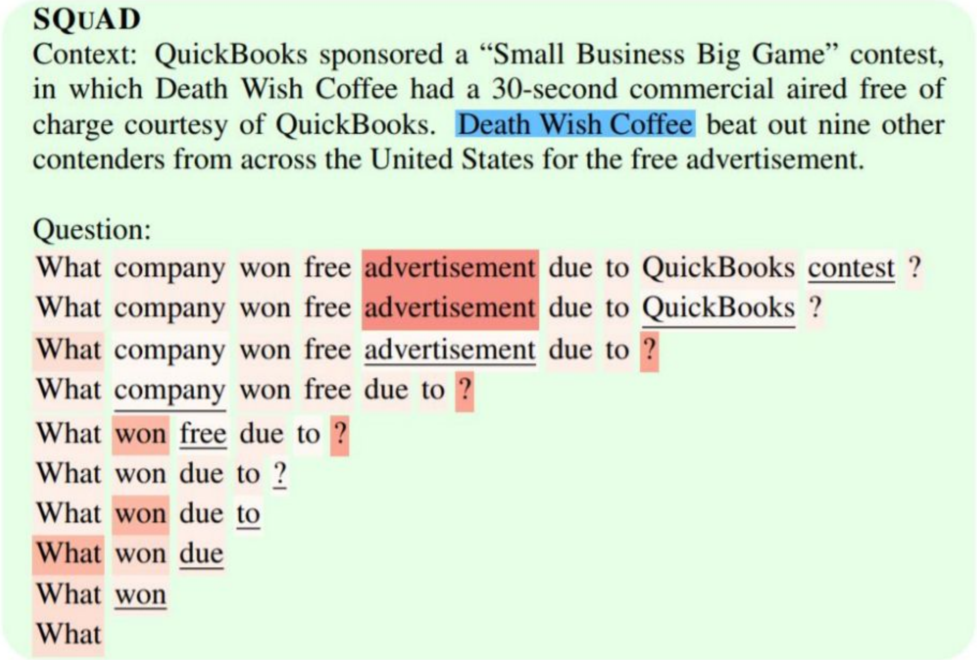
- Use CNN (*QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension*, Yu et al. 2018)

- "Ensemble" outputs from a Memory Network (*Stochastic Answer Networks for Machine Reading Comprehension*, Liu et al. 2017 (ACL))

- And, of course, use BERT

Do models really understand?

Considering only a small part of the query or context does not change the result.



From the Black Box workshop of EMNLP

Pathologies of Neural Models Make Interpretations Difficult, Feng et al. 2018

What Makes Reading Comprehension Questions Easier?, Sugawara et al. 2018

Also in VQA *Did the Model Understand the Question?* Mudrakarta et al. 2016

Some datasets are intrinsically weak

Adversarial Examples for Evaluating Reading Comprehension Systems, Jia & Liang 2017

Adding a distracting sentence in the context can totally disturb the model.

	SQuAD	AddSent
BiDAF	75.5	34.3
Mnemonic	78.5	46.6

F1 score on original and adversarial datasets

Some datasets are intrinsically weak

Adversarial Examples for Evaluating Reading Comprehension Systems, Jia & Liang 2017

Adding a distracting sentence in the context can totally disturb the model.

	SQuAD	AddSent
BiDAF	75.5	34.3
Mnemonic	78.5	46.6

F1 score on original and adversarial datasets

In the original dataset, there is ***necessarily*** *an answer in the context!*

Some datasets are intrinsically weak

Adversarial Examples for Evaluating Reading Comprehension Systems, Jia & Liang 2017

Adding a distracting sentence in the context can totally disturb the model.

	SQuAD	AddSent
BiDAF	75.5	34.3
Mnemonic	78.5	46.6

F1 score on original and adversarial datasets

*In the original dataset, there is **necessarily** an answer in the context!*

Efficient and Robust Question Answering from Minimal Context over Documents, Min et al. 2018 (ACL)

92% of SQuAD questions are answerable with **one** sentence

→ *pipeline* model
sentences selector > classical QA model

Achieves 74.5 EM / 83.1 F1

Some datasets are intrinsically weak

Adversarial Examples for Evaluating Reading Comprehension Systems, Jia & Liang 2017

Adding a distracting sentence in the context can totally disturb the model.

	SQuAD	AddSent
BiDAF	75.5	34.3
Mnemonic	78.5	46.6

F1 score on original and adversarial datasets

In the original dataset, there is ***necessarily*** *an answer in the context!*

Efficient and Robust Question Answering from Minimal Context over Documents, Min et al. 2018 (ACL)

92% of SQuAD questions are answerable with **one** sentence

→ *pipeline* model
sentences selector > classical QA model

Achieves 74.5 EM / 83.1 F1

The reasoning involved is rather limited because the answer is generally obvious

→ **don't expect the model to be intelligent**

New dataset every two months

Evolution of SQuAD into **SQuAD2** which contains **unanswerable** questions

→ *Know What You Don't Know: Unanswerable Questions for SQuAD*, Rajpurkar, **Jia & Liang** 2018

New dataset every two months

Evolution of SQuAD into **SQuAD2** which contains **unanswerable** questions

→ *Know What You Don't Know: Unanswerable Questions for SQuAD*, Rajpurkar, **Jia & Liang** 2018

Ask question during a conversation about a hidden piece of text

→ ***QuAC** : Question Answering in Context*, Choi et al. 2018 (EMNLP)

New dataset every two months

Evolution of SQuAD into **SQuAD2** which contains **unanswerable** questions

→ *Know What You Don't Know: Unanswerable Questions for SQuAD*, Rajpurkar, **Jia & Liang** 2018

Ask question during a conversation about a hidden piece of text

→ ***QuAC*** *: Question Answering in Context*, Choi et al. 2018 (EMNLP)

Develop *real* multi-paragraph reasoning across distractors and **justify your answer**

→ ***HotpotQA****: A Dataset for Diverse, Explainable Multi-hop Question Answering*, Yang et al. 2018 (EMNLP)

New dataset every two months

Evolution of SQuAD into **SQuAD2** which contains **unanswerable** questions

→ *Know What You Don't Know: Unanswerable Questions for SQuAD*, Rajpurkar, **Jia & Liang** 2018

Ask question during a conversation about a hidden piece of text

→ ***QuAC*** *: Question Answering in Context*, Choi et al. 2018 (EMNLP)

Develop *real* multi-paragraph reasoning across distractors and **justify your answer**

→ ***HotpotQA****: A Dataset for Diverse, Explainable Multi-hop Question Answering*, Yang et al. 2018 (EMNLP)

The models trained on these datasets **should be** more robust thanks to the *false* examples and multi-task learning