

Deep Learning for Natural Language Processing

Machine Reading Comprehension with Attention



Jinyoung Yeo

AAI5003

Teaching Machines to Read



Sara wanted to play on a baseball team . However, she had never tried to swing a bat and hit a baseball before .

Teaching Machines to Read

1 Sara wanted to play on a baseball team .

2 However , she had never tried to swing a bat and hit a baseball before .

1 Sara wanted to play on a baseball team .

2 However , she had never tried to swing a bat and hit a baseball before .

1 **Mention** Sara wanted to play on a baseball team .

2 However , **Mention** she had never tried to swing a bat and hit a baseball before .

Reading Comprehension as Question Answering

Wendy Lehnert. 1977. "The Process of Question Answering"

THE PROCESS OF QUESTION ANSWERING

May 1977

Research Report #88

Wendy Lehnert

When a person understands a story, he can demonstrate his understanding by answering questions about the story. Since questions can be devised to query any aspect of text comprehension, the ability to answer questions is the strongest possible demonstration of understanding. Question answering is therefore a task criterion for evaluating reading skills.

If a computer is said to understand a story, we must demand of the computer the same demonstrations of understanding that we require of people. Until such demands are met, we have no way of evaluating text understanding programs. Any computer programmer can write a program which inputs text. If the programmer assures us that his program understands text, it is a bit like being reassured by a used car salesman about a suspiciously low speedometer reading. Only when we can ask a program to answer questions about what it reads will we be able to begin to assess that program's comprehension.

"Since questions can be devised to query **any aspect** of text comprehension, the ability to answer questions is the **strongest possible demonstration of understanding.**"

Reading Comprehension as Question Answering

Sara wanted to play on a baseball team. She had never tried to swing a bat and hit a baseball before. Her Dad gave her a bat and together they went to the park to practice.

Why was Sara practicing?

```
graph TD; A[Sara wanted to play on a baseball team. She had never tried to swing a bat and hit a baseball before. Her Dad gave her a bat and together they went to the park to practice.] --> C[reading comprehension system]; B[Why was Sara practicing?] --> C; C --> D[She wanted to play on a team.]
```

**reading
comprehension
system**

She wanted to play on a team.

Reading comprehension is a “new” field

Before **2015**, we hadn't had any statistical NLP systems which are capable of reading a simple passage and answering questions.

Datasets (passage, question, answer)

- MCTest: 2600 questions
- ProcessBank: 500 questions

Systems

- Hand-built systems
- Classifier with linguistic features

Reading comprehension is a “new” field

Since 2015...

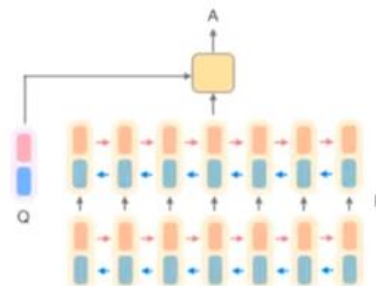
Datasets (passage, question, answer)



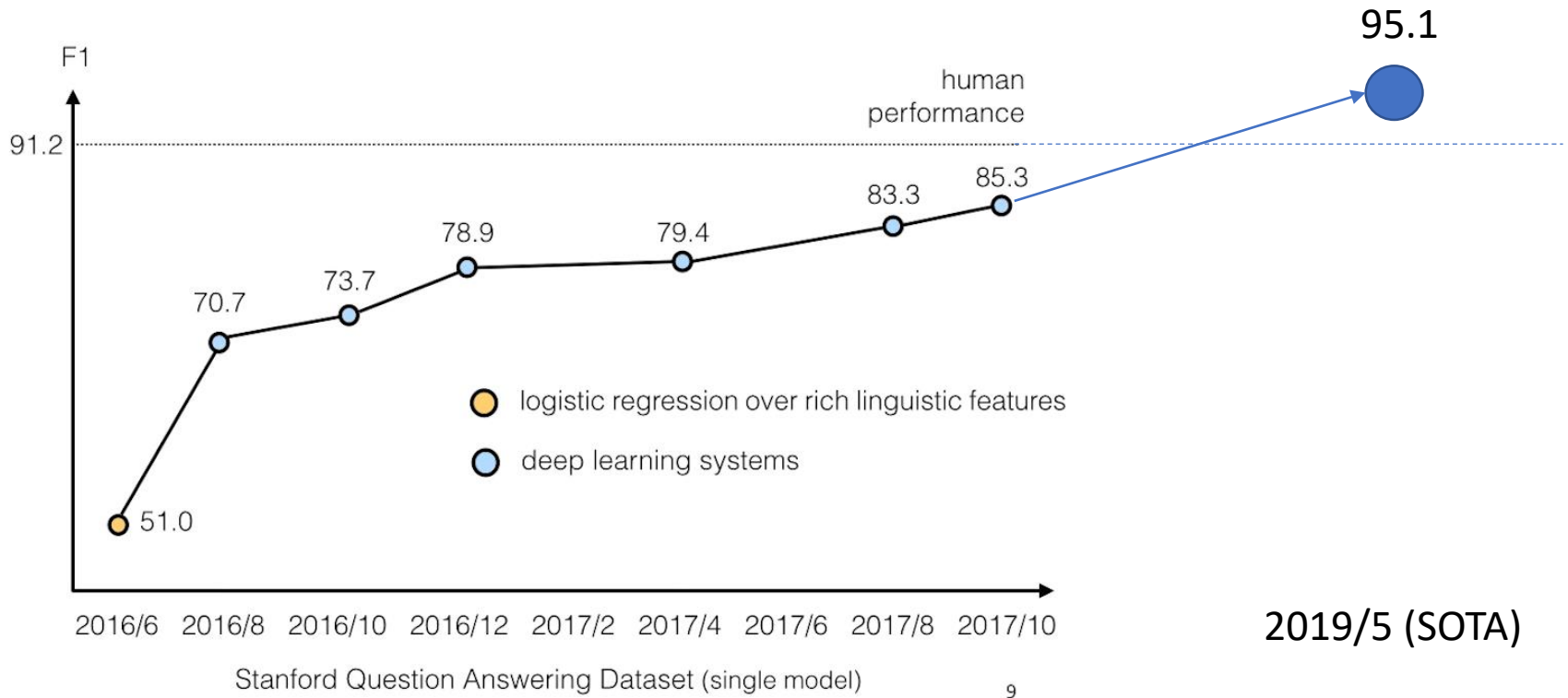
More than 100k examples!

Systems

End-to-end neural networks!



Progress is rapid!




Outline


- Dataset: CNN/Daily Mail Dataset
- Model: Stanford Attentive Reader
- Model: DeepMind Attentive Reader

- Dataset: Stanford Question Answering Dataset (SQuAD)
- Model: Stanford Attentive Reader*

- Task: Open-domain QA
- Model: DrQA

CNN/Daily Mail Datasets


 Entertainment » 'Star Wars' universe gets its first gay character



Story highlights

Official "Star Wars" universe gets its first gay character, a lesbian governor

The character appears in the upcoming novel "Lords of the Sith"

Characters in  movies have gradually become more diverse

(CNN) — If you feel a ripple in the Force today, it may be the news that the official Star Wars universe is getting its first gay character.

According to the sci-fi website Big Shiny Robot, the upcoming novel "Lords of the Sith" will feature a capable but flawed Imperial official named Moff Mors who "also happens to be a lesbian."

The character is the first gay figure in the official Star Wars universe -- the movies, television shows, comics and books approved by Star Wars franchise owner Disney -- according to Shelly Shapiro, editor of "Star Wars" books at Random House imprint Del Rey Books.

CNN/Daily Mail Datasets

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 .

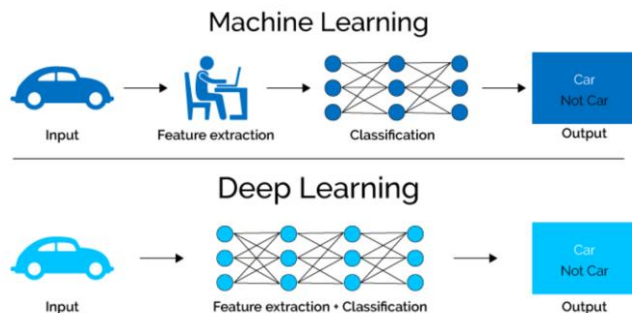
question

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

answer @entity6

CNN: 380k, Daily Mail: 879k training - free!

A Cateogrical-feature Classifier



Can we build a simple, end-to-end neural network to tackle this problem?

How is it different from feature-based linear classifiers?

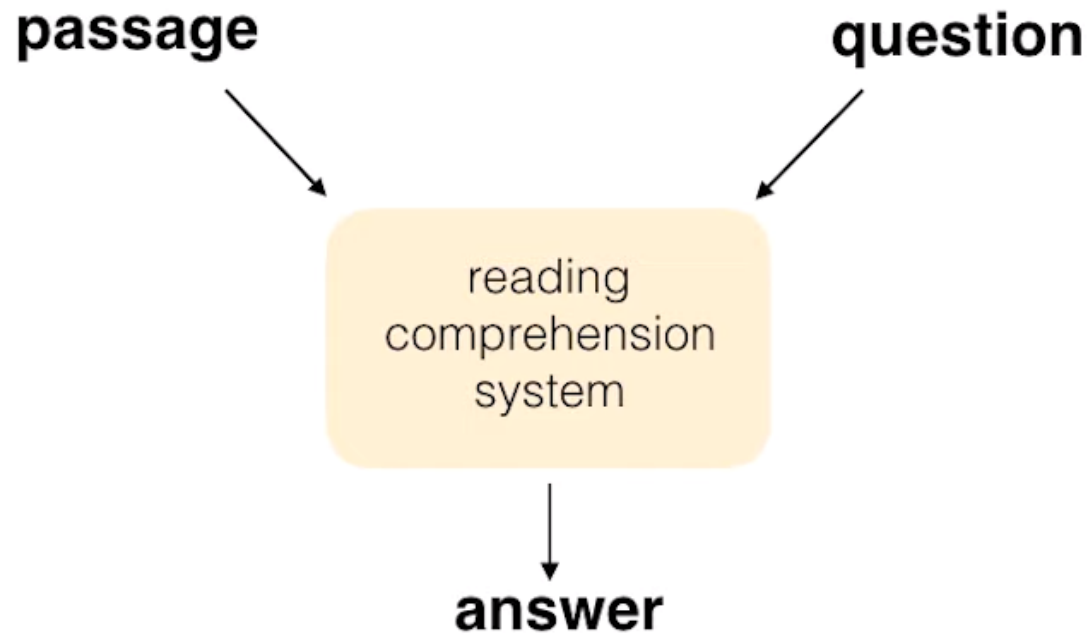
For each candidate entity e , we build a symbolic feature vector:

$$f_{P,Q}(e) > f_{P,Q}(e')$$

1. Whether e occurs in P
2. Whether e occurs in Q
3. Frequency of e in P
4. First position of e in P

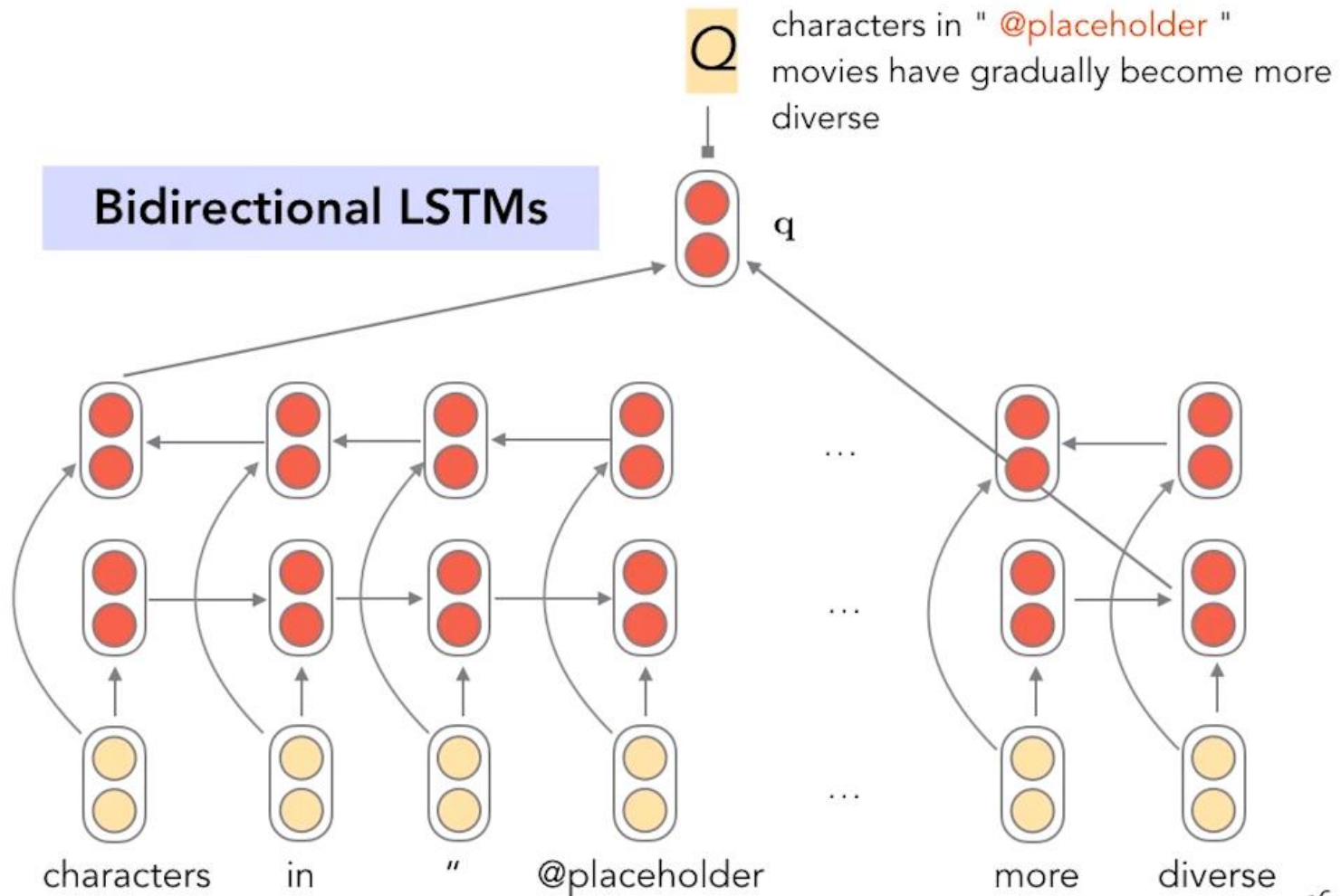
5. Whether e co-occurs with another Q word in P .
6. word distance
7. **n-gram** exact match
8. **dependency parse** match

Attentive Reader



1. Encode the question
 2. Encode the passage
 3. Model the interaction between passage and question
 4. Infer the answer
- ... all in vector space!**

Attentive Reader



Attentive Reader

Bidirectional LSTMs

Q

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



q

P



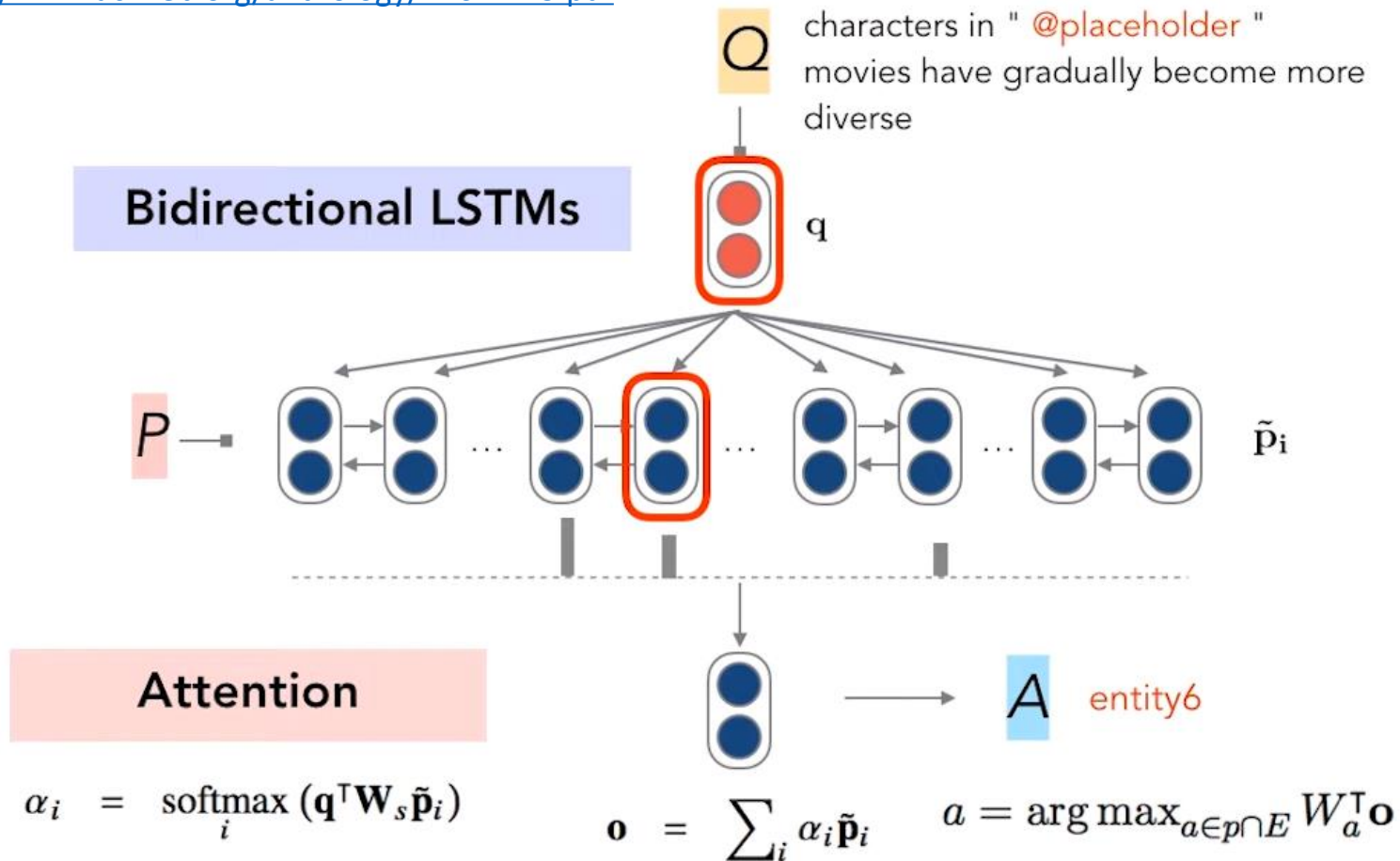
\tilde{P}_i

(@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 "

“Stanford” Attentive Reader (2016)

A thorough examination of the CNN/Daily Mail reading comprehension task

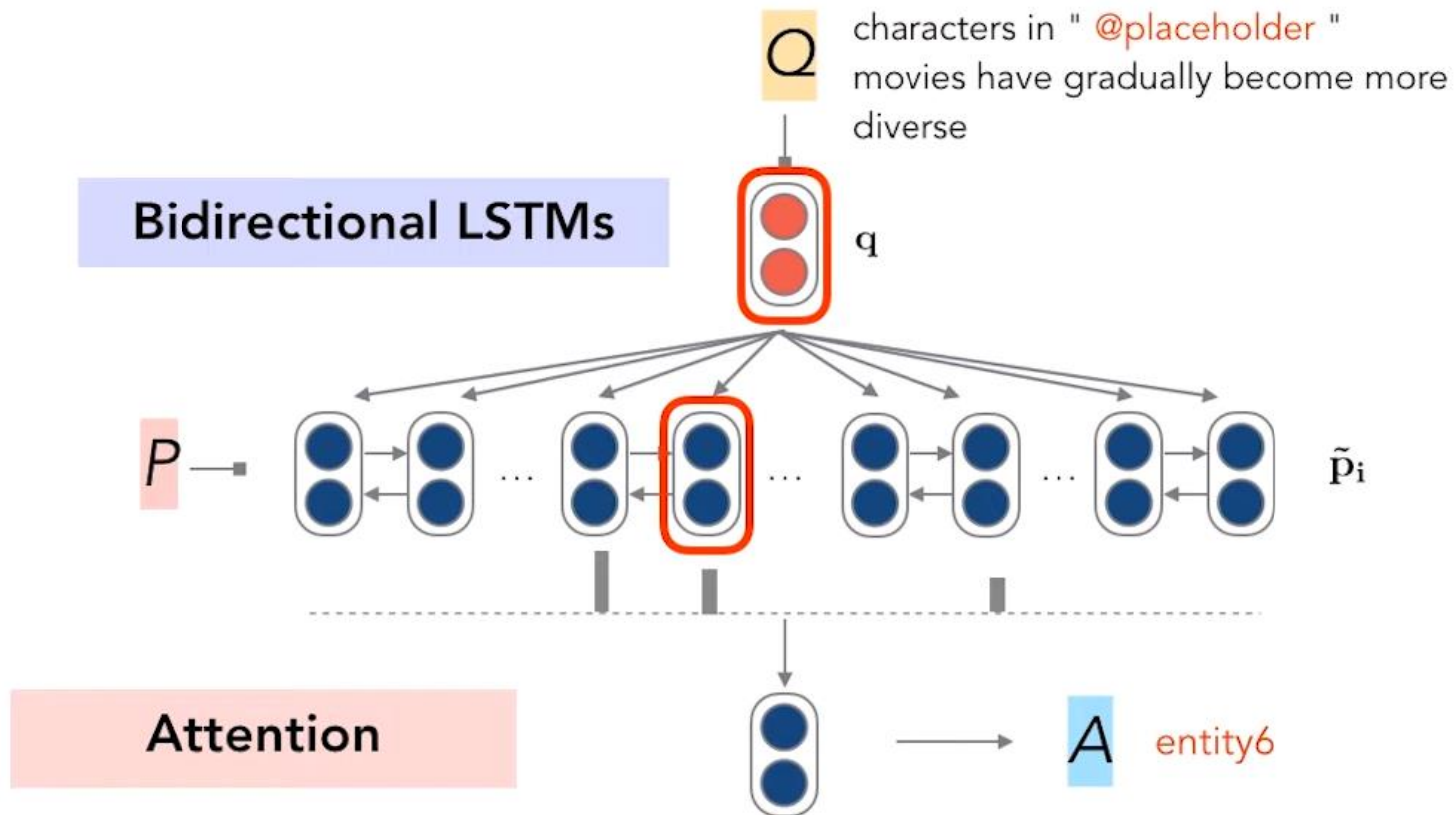
<https://www.aclweb.org/anthology/P16-1223.pdf>



“DeepMind” Attentive Reader (2015)

Teaching Machines to Read and Comprehend

<https://arxiv.org/pdf/1506.03340.pdf>



$$\begin{aligned}
 m(t) &= \tanh(W_{ym}y_d(t) + W_{um}u), & g^{\text{AR}}(d, q) &= \tanh(W_{rg}r + W_{ug}u). \\
 s(t) &\propto \exp(w_{ms}^T m(t)), & p(a|d, q) &\propto \exp(W(a)g(d, q)), \quad \text{s.t. } a \in V, \\
 r &= y_d s,
 \end{aligned}$$

Attendance Quiz

A thorough examination of the CNN/Daily Mail reading comprehension task
<https://www.aclweb.org/anthology/P16-1223.pdf>

	CNN	Daily Mail
DeepMind: Attentive Reader	63.8	68.0
Stanford: Attentive Reader	73.6	76.6

Analysis

Categorical feature classifier	67.9	68.3
Stanford: Attentive Reader	73.6	76.6

What level of language understanding is needed?

What have current models learned?

1. exact match
2. paraphrasing
3. partial clue
4. multiple sentence
5. coreference errors
6. ambiguous/hard cases

Analysis

exact match

@placeholder and @entity8 were tried together and convicted of murder , but now cleared

@entity12 and @entity8 were tried together and convicted of murder by two separate courts .

paraphrasing

@placeholder says he understands why @entity0 won't play at this tournament

“entity0 called me personally to let me know that he wouldn't be playing here at @entity23” **entity3** said.

Analysis

partial clue

a tv movie based on @entity12 's book " @placeholder " casts a @entity76 actor as @entity5

@entity12 professed that his "entity11" is not a religious book...

multiple sentences

" he is doing a his - and - her duet all by himself, " @entity6 said of @placeholder

" we got some groundbreaking performances , here too , tonight , " @entity6 said . " we got @entity17 , who will be doing some musical performances . he 's doing a his - and - her duet all by himself . "

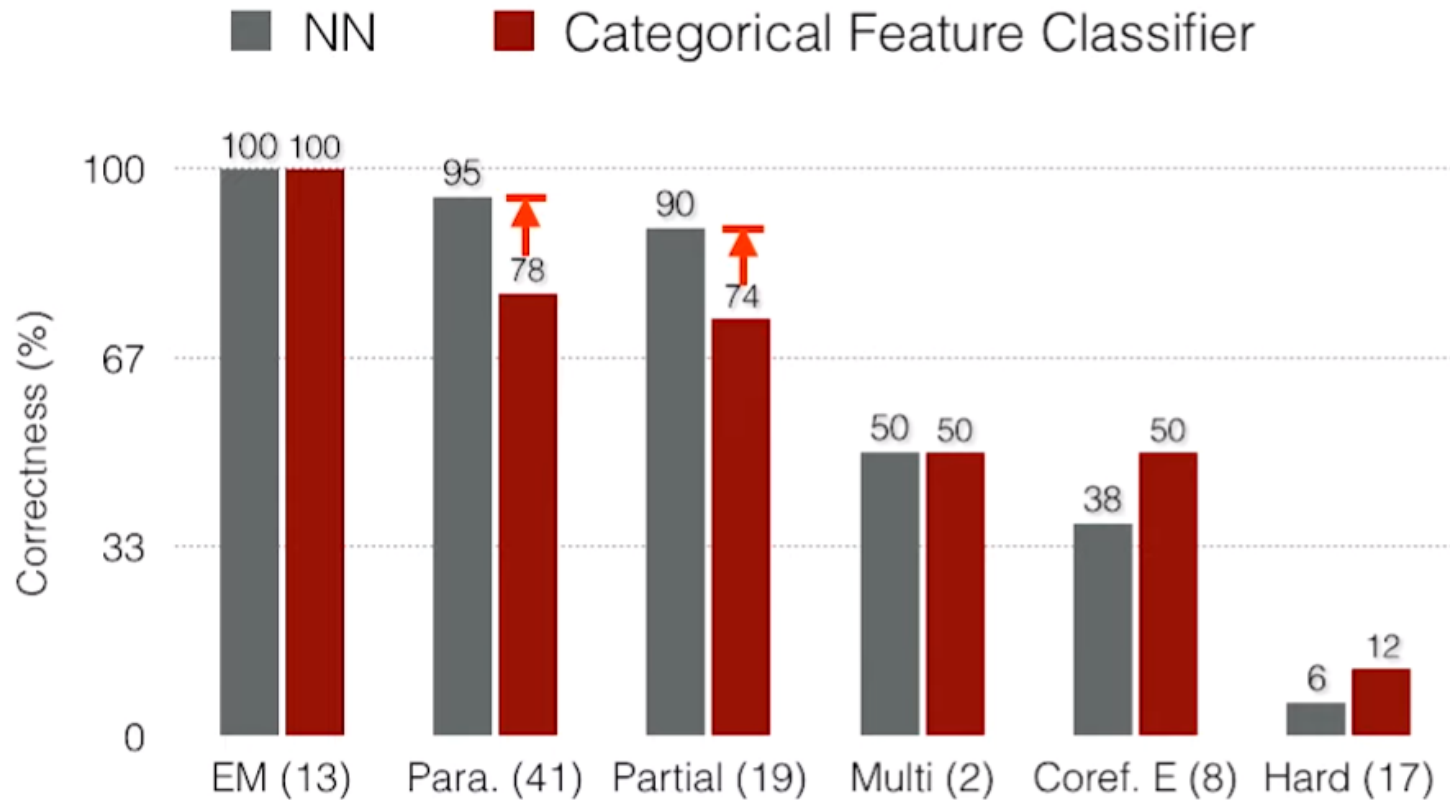
Analysis

What level of language understanding is needed?

What have current models learned?

- | | |
|-------------------------|------------|
| 1. exact match | 13% |
| 2. paraphrasing | 41% |
| 3. partial clue | 19% |
| 4. multiple sentence | 2% |
| 5. coreference errors | 8% |
| 6. ambiguous/hard cases | 17% |

Analysis



BiLSTMs + attention models are really good at learning semantic matching.

Does it work in a more real QA setup?

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

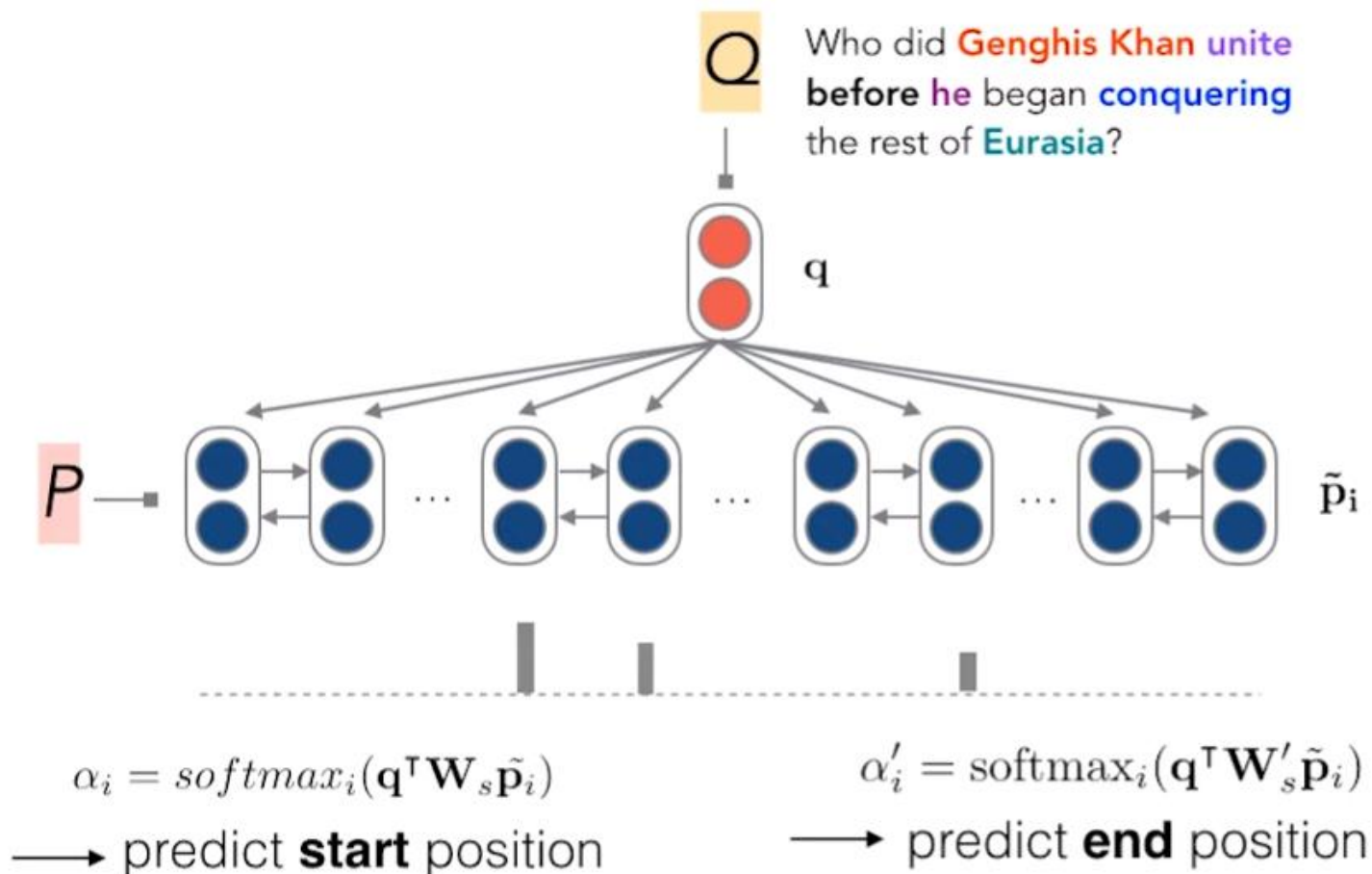
Stanford Question Answering Dataset (SQuAD)

- **Passage:** selected from Wikipedia
- **Question:** crowdsourced
- **Answer:** must be a span in the passage

Who did **Genghis Khan** **unite** **before** **he** began **conquering** the rest of **Eurasia**?

He came to power by **uniting** many of the nomadic tribes of Northeast Asia. **After** founding the Mongol Empire and being proclaimed "**Genghis Khan**", he started the Mongol invasions that resulted in the **conquest** of most of **Eurasia**. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezm Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

Attentive Reader for SQuAD



Common Failure Cases

Question: What is the total number of professors, instructors, and lecturers at Harvard?

Harvard's **2,400** professors, lecturers, and instructors instruct **7,200** undergraduates and 14,000 graduate students. The school color is crimson, which is also the name of the Harvard sports teams and the daily newspaper, The Harvard Crimson. The color was unofficially adopted (in preference to magenta) by an 1875 vote of the student body, although the association with some form of red can be traced back to 1858, when Charles William Eliot, a young graduate student who would later become Harvard's 21st and longest-serving president (1869–1909), bought red bandanas for his crew so they could more easily be distinguished by spectators at a regatta.

Correct

Predicted

Common Failure Cases

Question: What is the population of the second largest city in California ?

Los Angeles (at 3.7 million people) and San Diego (at 1.3 million people), both in southern California, are the two largest cities in all of California (and two of the eight largest cities in the United States). In southern California there are also twelve cities with more than 200,000 residents and 34 cities over 100,000 in population. Many of southern California's most developed cities lie along or in close proximity to the coast, with the exception of San Bernardino and Riverside.

Correct

Predicted

Open-domain QA

- SQuAD is still a restricted QA setup:
 - Questions that can be answered by **span selection**
 - Annotators can see the paragraph when writing questions
- ⇒ high **lexical overlap** between question and paragraph

Can we leverage reading comprehension systems for even broader open-domain question answering?

DrQA

Reading Wikipedia to Answer Open-Domain Questions

<https://arxiv.org/abs/1704.00051>

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



DrQA

Reading Wikipedia to Answer Open-Domain Questions

<https://arxiv.org/abs/1704.00051>

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



DrQA = Information Retrieval + Reading Comprehension

DrQA

Reading Wikipedia to Answer Open-Domain Questions

<https://arxiv.org/abs/1704.00051>

Hi!



Hello! Please ask a question.

Who invented LSTM?



Hmmm. Tough one. I would say... Sepp Hochreiter

Where is Stanford University located?



I'm quite confident the answer is... California

DrQA

Reading Wikipedia to Answer Open-Domain Questions

<https://arxiv.org/abs/1704.00051>

What is the highest mountain on Earth?



I'm quite confident the answer is... Mount Everest

What year was the American Declaration of Independence?



I'm quite confident the answer is... 1776

Reference

- Danqi Chen: From Reading Comprehension to Open-Domain Question Answering
- <https://www.youtube.com/watch?v=1RN88O9C13U&t=2055s>