Building a Smart Question Answering System from Scratch

Minjoon Seo
PhD Student
University of Washington

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

UW NLP
What is “Question Answering System”? 
Q: Which airports are in New York City?

There are four airports in NYC: JFK, LeGuardia, Newark, and Stewart.
Why do you care about it?
“If you got a billion dollars to spend on a huge research project, what would you like to do?”

“I'd use the billion dollars to build a NASA-size program focusing on natural language processing (NLP), in all of its glory (semantics, pragmatics, etc).”

Michael Jordan
Professor of Computer Science
UC Berkeley
Towards Artificial General Intelligence...

Natural language is the best tool to describe and communicate “thoughts”

Asking and answering questions is the best way to develop deeper “thoughts”
QA in Our Lives

• Amazon Alexa
• Apple Siri
• Facebook M
• Google Now
• IBM Watson
• Microsoft Cortana
• Etc.
Limitations of industrial QA systems

• Carefully engineered modules, rules and features by humans
  • Can machines **learn end-to-end**, and learn new things easily?

• Little capability for reasoning
  • Can machines perform **reasoning**: induction, deduction, conditional expression, etc.?
Limitations of industrial QA systems

- Carefully engineered modules, rules and features by humans
  - Can machines learn end-to-end, and learn new things easily?

- Little capability for reasoning
  - Can machines perform reasoning: induction, deduction, conditional expression, etc.?
End-to-end learning (minimal supervision)

• Machines are learning from **question-answer pairs** only (supervised by answers only)

• No supervision of how-to, prior knowledge, etc. **Latently learn** these things instead.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which airports are in New York City?</td>
<td>JFK, LeGuardia, Newark, and Stewart</td>
</tr>
<tr>
<td>Which NFL team represented the AFC</td>
<td>Denver Broncos</td>
</tr>
<tr>
<td>Who wrote Harry Potter?</td>
<td>J. K. Rowling</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Q: Which NFL team represented the AFC at Super Bowl 50?
Super Bowl 50

Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi’s Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.

Q: Which NFL team represented the AFC at Super Bowl 50?

A: Denver Broncos

• Document is given.
• User asks a document-specific question
Our Model: Bi-directional Attention Flow (BiDAF)

Donald Trump is the president of the U.S. Who leads the United States?
VGG-16

BiDAF (ours)
**SQuAD Leaderboard**

(Stanford-qa.com) as of 12pm, 2 Dec 2017

## Test Set Leaderboard

Since the release of our dataset (and paper), the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test and development sets of v1.1.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Model</th>
<th>Test EM</th>
<th>Test F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BiDAF (ensemble) Allen Institute for AI &amp; University of Washington (Seo et al. '16)</td>
<td>73.3</td>
<td>81.1</td>
</tr>
<tr>
<td>2</td>
<td>Dynamic Coattention Networks (ensemble) Salesforce Research (Xiong &amp; Zhong et al. '16)</td>
<td>71.6</td>
<td>80.4</td>
</tr>
<tr>
<td>2</td>
<td>r-net (ensemble) Microsoft Research Asia</td>
<td>72.1</td>
<td>79.7</td>
</tr>
<tr>
<td>4</td>
<td>r-net (single model) Microsoft Research Asia</td>
<td>68.4</td>
<td>77.5</td>
</tr>
<tr>
<td>5</td>
<td>BiDAF (single model) Allen Institute for AI &amp; University of Washington (Seo et al. '16)</td>
<td>68.0</td>
<td>77.3</td>
</tr>
<tr>
<td>5</td>
<td>Multi-Perspective Matching (ensemble) IBM Research</td>
<td>68.2</td>
<td>77.2</td>
</tr>
</tbody>
</table>
Leaderboard as of 24 Mar 2017

- BiDAF still third!
- 23 Submissions
- Microsoft, IBM, Salesforce, Facebook, Google, ...

Since the release of our dataset, the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test and development sets of v1.1. Will your model outperform humans on the QA task?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Model</th>
<th>EM</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>r-net (ensemble)</td>
<td>76.922</td>
<td>84.006</td>
</tr>
<tr>
<td></td>
<td>Microsoft Research Asia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ReasoNet (ensemble)</td>
<td>75.034</td>
<td>82.552</td>
</tr>
<tr>
<td></td>
<td>MSR Redmond</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BiDAF (ensemble)</td>
<td>73.744</td>
<td>81.525</td>
</tr>
<tr>
<td></td>
<td>Allen Institute for AI &amp; University of Washington</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Multi-Perspective Matching (diversity-ensemble)</td>
<td>73.765</td>
<td>81.257</td>
</tr>
<tr>
<td></td>
<td>IBM Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>r-net (single model)</td>
<td>72.401</td>
<td>80.751</td>
</tr>
<tr>
<td></td>
<td>Microsoft Research Asia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Dynamic Coattention Networks (ensemble)</td>
<td>71.625</td>
<td>80.383</td>
</tr>
<tr>
<td></td>
<td>Salesforce Research</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.

Q: Which NFL team represented the AFC at Super Bowl 50?
Q: Which NFL team represented the AFC at Super Bowl 50?
Pipelined Approach

“Where was Barack Obama born?”

Search Alg.

Document 1

Document 2

... 

Document n

QA System

“Hawaii, USA”
Open-domain QA Demo
(prototype)
400 lines of code!
Isn’t Google doing this already?

• Requires Structured Knowledge Base

August 4, 1961 (age 55 years), Kapiolani Medical Center for Women and Children, Honolulu, HI

Donald Trump
June 14, 1946

Michelle Obama
January 17, 1964

Hillary Clinton
October 26, 1947
Isn’t Google doing this already?

• Carefully engineered and not specifically giving you the answer

A clear cloudless day-time **sky** is **blue** because molecules in the air scatter **blue** light from the sun more than they scatter red light. When we look towards the sun at sunset, we see red and orange colours because the **blue** light has been scattered out and away from the line of sight.

**Why is the sky Blue?**
math.ucr.edu/home/baez/physics/General/BlueSky/blue_sky.html
What the model can and can’t do...

• You can ask any question that is *directly answerable* by a document

• You can’t ask *reasoning questions*
Reasoning questions

“If frogs eats insects and flies are insects, do frogs eat flies?”

“If John has an apple and John went to bathroom, where is the apple?”
New Assumption:

We make the syntax of sentences and question *simple*
Reasoning Question Answering

**Task 1: Single Supporting Fact**
- Mary went to the bathroom.
- John moved to the hallway.
- Mary travelled to the office.
- Where is Mary? **A: office**

**Task 2: Two Supporting Facts**
- John is in the playground.
- John picked up the football.
- Bob went to the kitchen.
- Where is the football? **A: playground**

**Task 3: Three Supporting Facts**
- John picked up the apple.
- John went to the office.
- John went to the kitchen.
- John dropped the apple.
- Where was the apple before the kitchen? **A: office**

**Task 4: Two Argument Relations**
- The office is north of the bedroom.
- The bedroom is north of the bathroom.
- The kitchen is west of the garden.
- What is north of the bedroom? **A: office**
- What is the bedroom north of? **A: bathroom**
Our approach: Query-Reduction

Reduced query:

Where is the apple?

Where is Sandra?

Where is Sandra?

Where is Daniel?

Where is Daniel?

Where is Daniel? → garden

Q: Where is the apple? A: garden
Query-Reduction Networks

- Reduce the query into an easier-to-answer query over the sequence of state-changing triggers (sentences), *in vector space*
bAbI QA Results (10k)
Conclusion

• Real question answering system with *minimal* supervision
• Machines that are able to reason with *minimal* supervision

• Reasoning for real, complex questions is *still hard*, but *not far away*
Industrial Impacts on QA Systems

• Models will become simpler and more elegant
  • Easier maintenance
  • Less man-hours needed

• Model behaviors are determined by data, not humans
  • Learns new things and improve performance with more data

• Models will be able to perform reasoning
  • Soon!
Thank you!

Minjoon Seo
PhD Student of Computer Science
University of Washington

seominjoon@gmail.com
seominjoon.github.io
There are 13 natural reserves in Warsaw – among others, Bielany Forest, Kabaty Woods, Czerniaków Lake. About 15 kilometres (9 miles) from Warsaw, the Vistula river’s environment changes strikingly and features a perfectly preserved ecosystem, with a habitat of animals that includes the otter, beaver and hundreds of bird species. There are also several lakes in Warsaw – mainly the oxbow lakes, like Czerniaków Lake, the lakes in the Łazienki or Wilanów Parks, Kamionek Lake. There are lot of small lakes in the parks, but only a few are permanent – the majority are emptied before winter to clean them of plants and sediments.

Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi’s Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.
Opening in May 1852 at the state may not aid of these may be more effect and may result in the state may not aid from 28 January to 25
but by September had been debut on May 5 ,