Building a Smart Question Answering System from Scratch

Minjoon Seo

PhD Student

University of Washington

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?

ASK ME ANYTHING

“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.

Question	Answer
Which airports are in New York City?	JFK, LeGuardia, Newark, and Stewart
Which NFL team represented the AFC ..	Denver Broncos
Who wrote Harry Potter?	J. K. Rowling
…	…

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

Context-aware Question Answering

Super Bowl 50

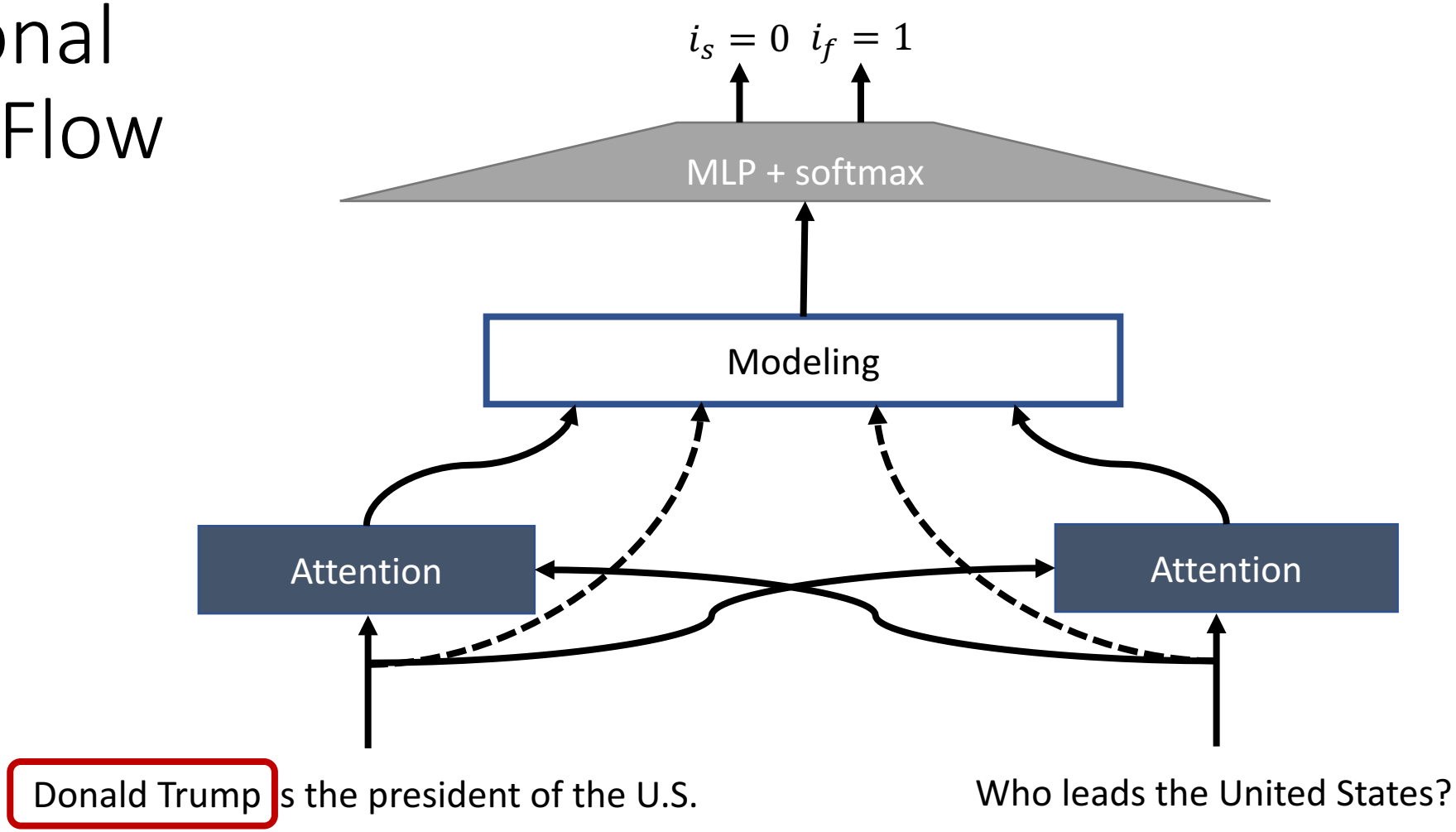
Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl 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.

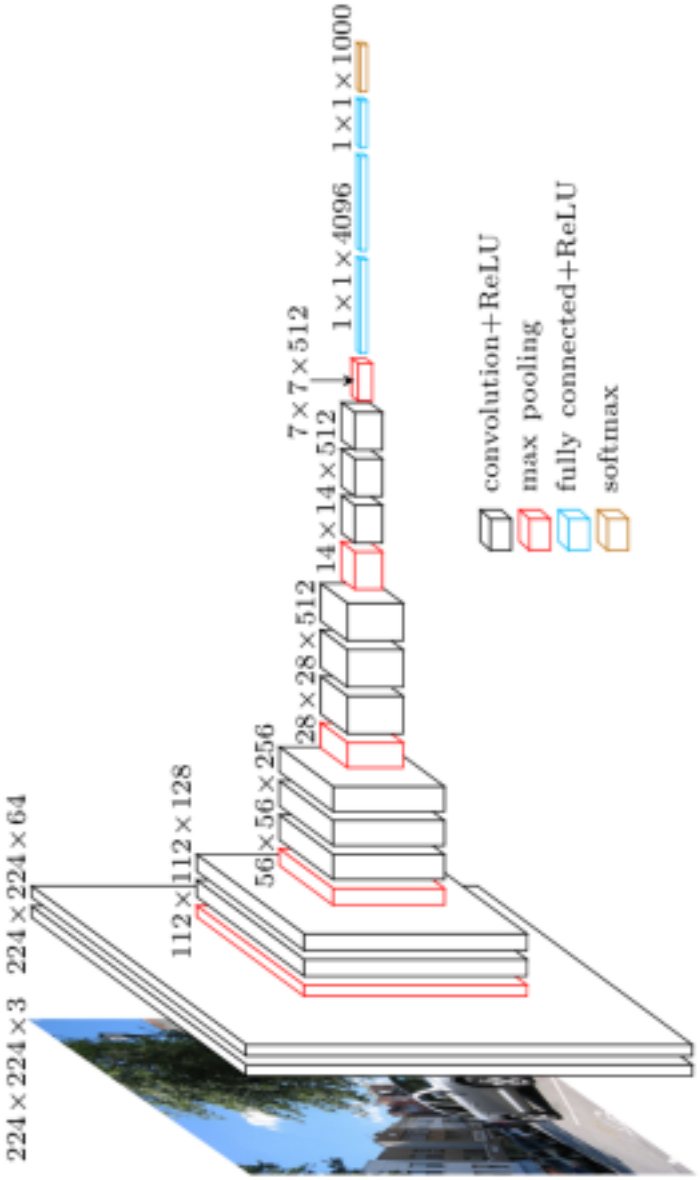
Q: Which NFL team represented the AFC at Super Bowl 50?

A: Denver Broncos

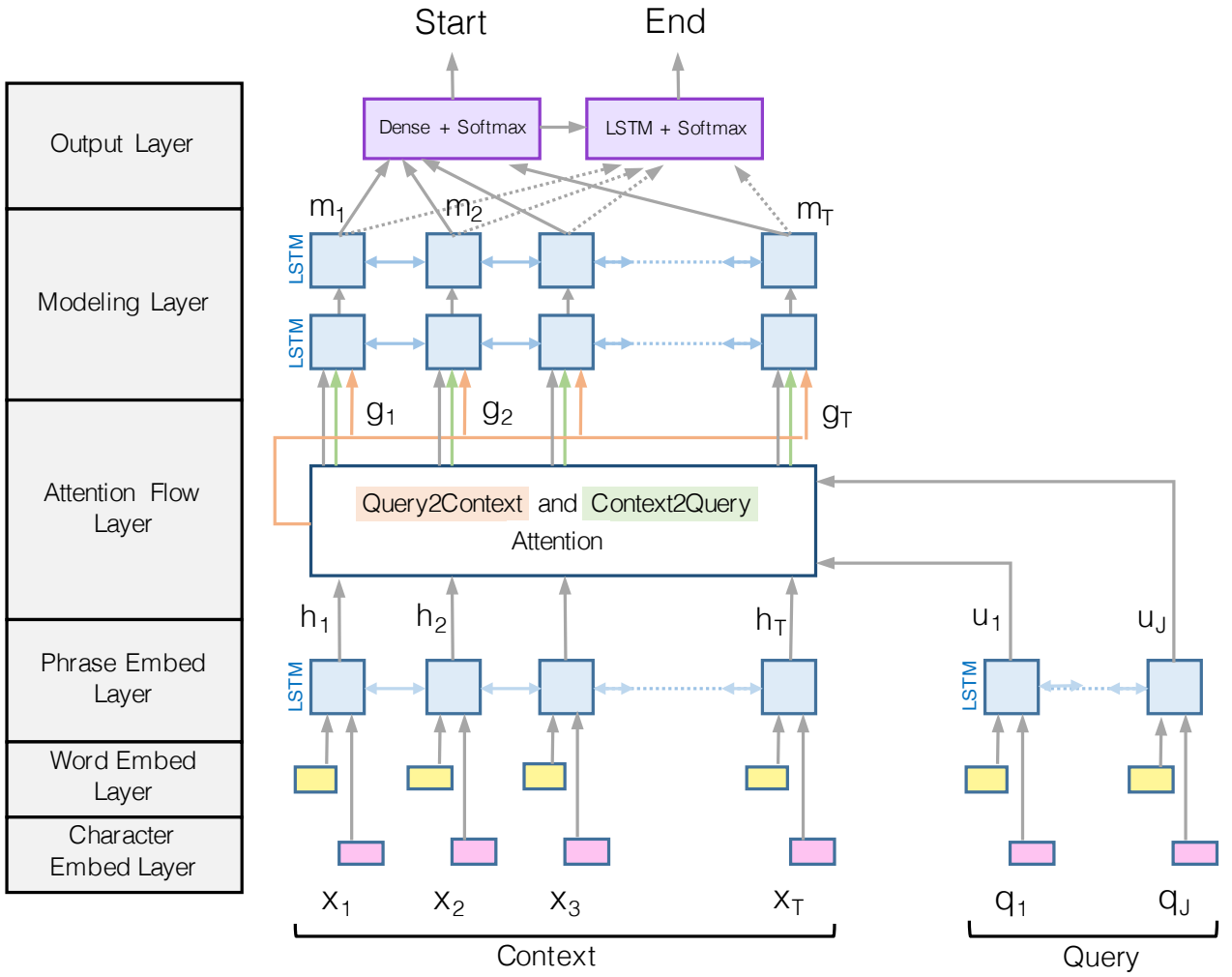
- Document is given.
- User asks a document-specific question

Our Model: Bi-directional Attention Flow (BiDAF)





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.

Rank	Model	Test EM	Test F1
1	BiDAF (ensemble) Allen Institute for AI & University of Washington (Seo et al. '16)	73.3	81.1
2	Dynamic Coattention Networks (ensemble) Salesforce Research (Xiong & Zhong et al. '16)	71.6	80.4
2	r-net (ensemble) Microsoft Research Asia	72.1	79.7
4	r-net (single model) Microsoft Research Asia	68.4	77.5
5	BiDAF (single model) Allen Institute for AI & University of Washington (Seo et al. '16)	68.0	77.3
5	Multi-Perspective Matching (ensemble) IBM Research	68.2	77.2

Leaderboard as of 24 Mar 2017

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

Leaderboard

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?

Rank	Model	EM	F1
1 9 days ago	r-net (ensemble) *Microsoft Research Asia*	76.922	84.006
2 12 days ago	ReasoNet (ensemble) *MSR Redmond*	75.034	82.552
3 a month ago	BiDAF (ensemble) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603	73.744	81.525
3 2 months ago	Multi-Perspective Matching (diversity-ensemble) *IBM Research* https://arxiv.org/abs/1612.04211	73.765	81.257
4 11 days ago	r-net (single model) *Microsoft Research Asia*	72.401	80.751
5 5 months ago	Dynamic Coattention Networks (ensemble) *Salesforce Research* https://arxiv.org/abs/1611.01604	71.625	80.383

Context-aware Question Answering

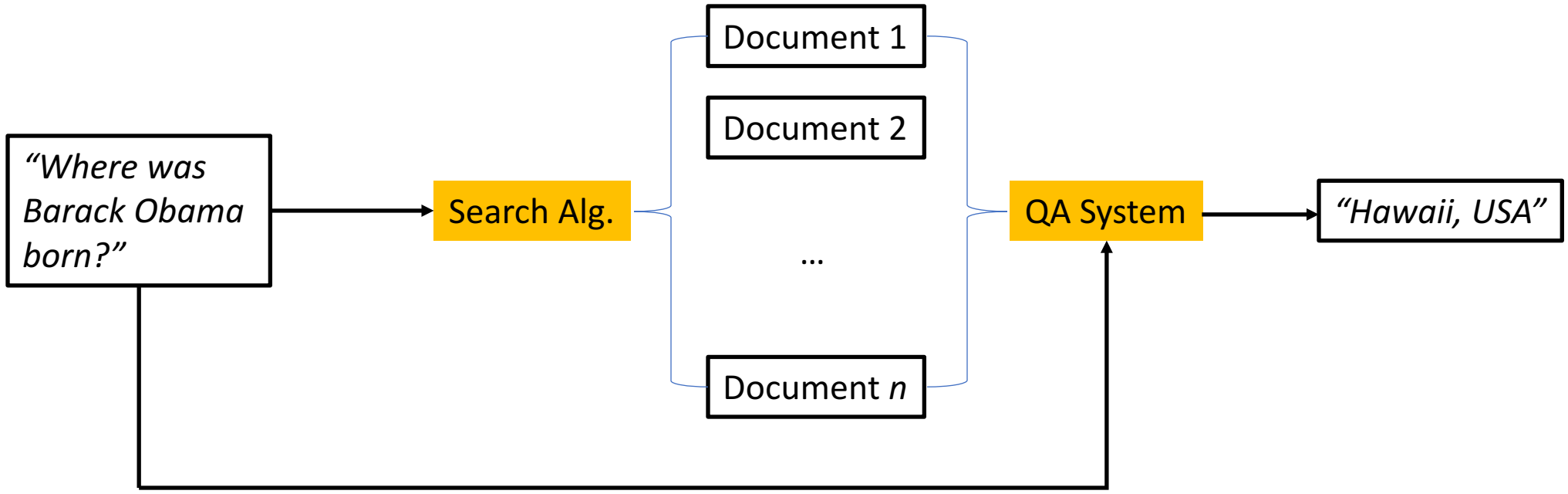
Super Bowl 50

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?

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

Pipelined Approach



Open-domain QA Demo
(prototype)
400 lines of code!

Isn’t Google doing this already?

- Requires Structured Knowledge Base

Barack Obama / Born

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

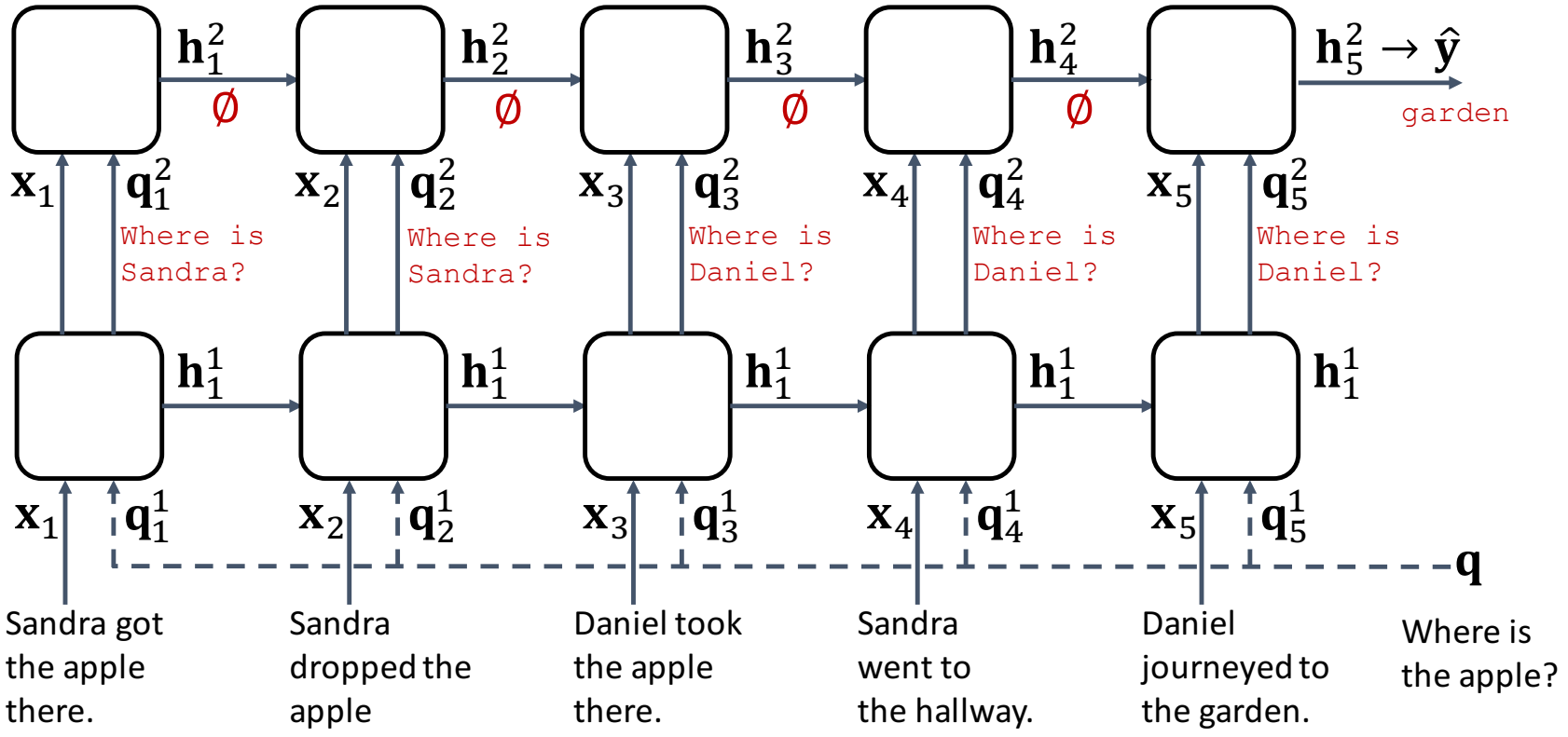
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Our approach: Query-Reduction

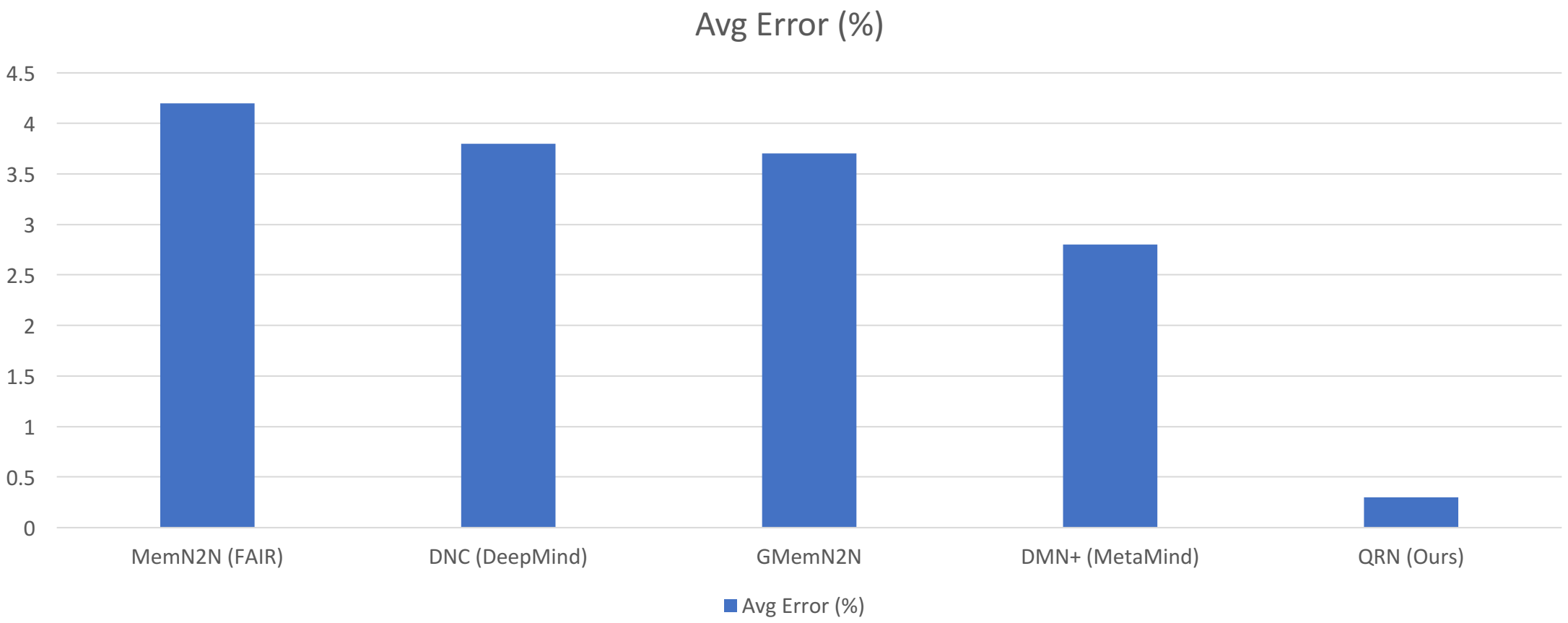
	Reduced query:
<START>	*Where is the apple?*
Sandra got the apple there.	*Where is Sandra?*
Sandra dropped the apple.	*Where is Sandra?*
Daniel took the apple there.	*Where is Daniel?*
Sandra went to the hallway.	*Where is Daniel?*
Daniel journeyed to the garden.	*Where is Daniel?* → *garden*
Q: Where 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

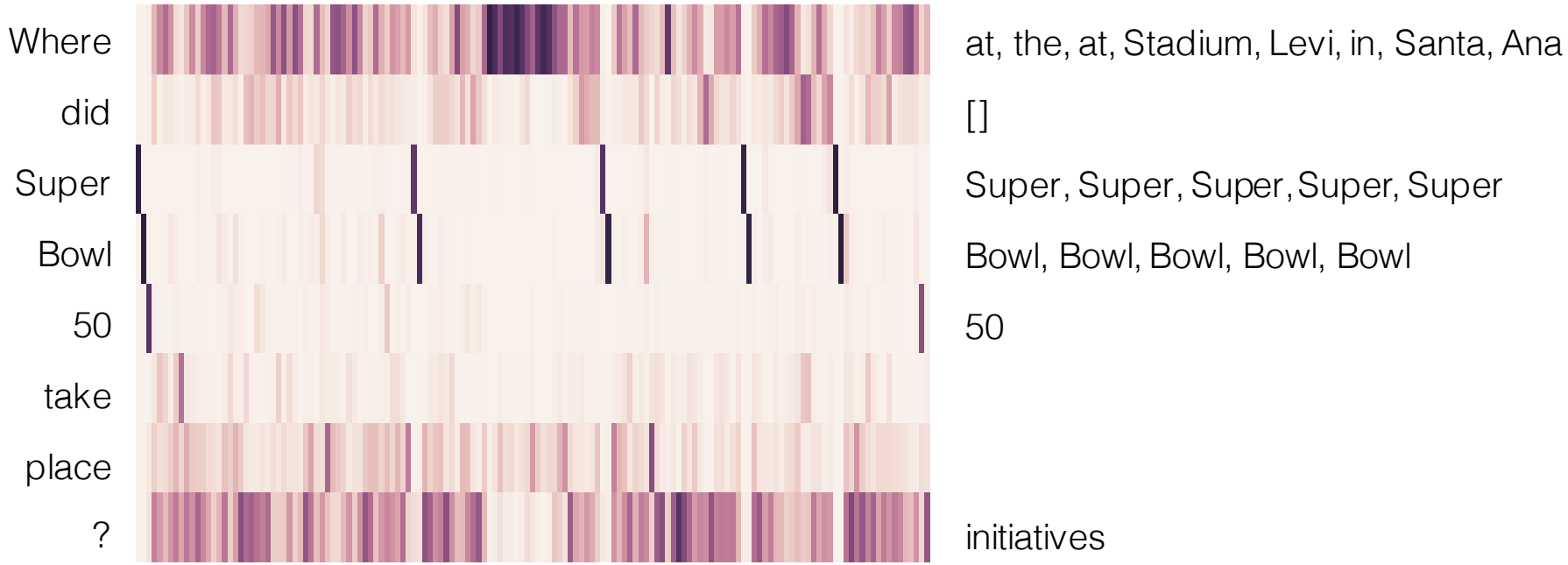
University of Washington

seominjoon@gmail.com

seominjoon.github.io

Attention Visualizations

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 .



Embedding Visualization at Word vs Phrase Layers