Natural QA System for Wikipedia

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Qian Ye | Deep Learning CSE599 Autumn 2019 | University of Washington

Abstract

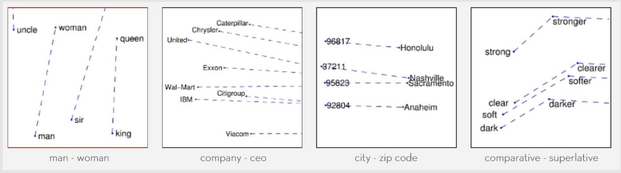
- "How a rainbow come into being?"
- This is a question an question answering (QA) system should be able to respond to. Existing QA models have been focused on extracting answers from a short paragraph rather than reading an entire page. My goal is to predict both short and long answer responses to real questions about Wikipedia articles.
- Long answers might be paragraphs, lists, list items, tables, or table rows.
- A short answers can be a sentence or phrase nested in a long answer.

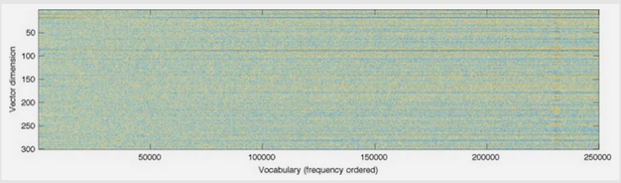
Dataset

- Google Natural Questions corpus.
- 307,373 training examples
- 7,842 test examples.

Vocabulary Dictionary and Embedding

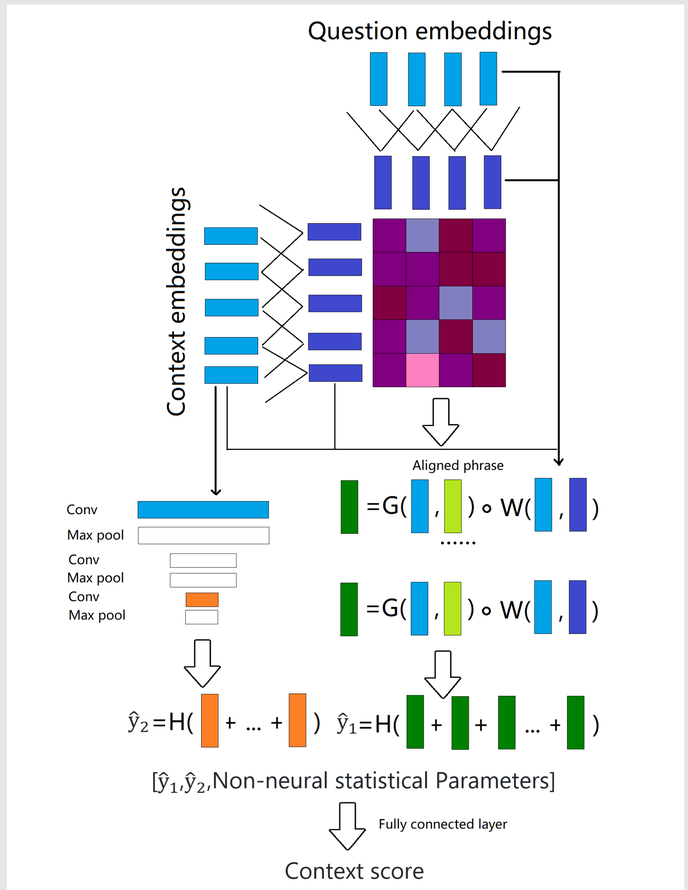
- GloVe: Vectors for Word Representation





Model

- **Long answer model.**
- Key features:
- -Use full GloVe.
- -Pass through convolution layers to get local -context before calculating attention score.
- -Weighted Comparation result.
- -Use network with 3 max pooling layer to get structure information of a context.



- **Short answer model**
- Implement the Document Reader model (Chen et al., 2017)

Experience

	Long answer Test			Short answer Test		
	P	R	F1	P	R	F1
DocuentQA	48.9	43.3	45.7	40.6	31.0	35.1
DecAtt	54.3	55.7	55.0	-	-	-
DocReader	-	-	-	31.9	31.1	31.5
My method	59.2	60.1	59.7	-	-	-

- **Examples of predicted results**
- Example 1 Question: when did they finish building the sydney opera house?
- Short Answer: 20 October 1973
- Long Answer: Designed by Danish architect Jorn Utzon, but completed by an Australian architectural team headed up by Peter Hall, the building was formally opened on 20 October 1973[4] after a … circumstances that followed, including cost and scheduling overruns as well as the architect's ultimate resignation.[5]
- Example 2 Qustion: where was the first sample of ascorbic acid isolated from?
- Short Answer: paprika
- Long Answer: In 1933, sugar chemist Walter Norman Haworth, working with samples of "hexuronic acid" that Szent-Györgyi had isolated from paprika and sent him in the previous year, …" for the compound, and later specifically l-ascorbic acid.[4] Because of their work, in 1937 the Nobel Prizes for chemistry and medicine were awarded to Haworth and Szent-Györgyi, respectively.

References

- Ankur Parikh, Oscar T¨ackstr¨om et al. 2016. A decomposable attention model for natural language inference.
- Tri Nguyen, Mir Rosenberg et al. 2016. MS MARCO: A human generated machine reading comprehension dataset.
- Tom Kwiatkowski, Jennimaria Palomaki et al. Natural Questions: a Benchmark for Question Answering Research.