

Iterative Attention Networks for Question Answering

Tom Henighan

Stanford Department of Physics

henighan@stanford.edu

Task: Answer question based on Reference Document

Dataset: Stanford Question Answer Dataset (SQuAD). $\sim 10^5$ Question/answer pairs. Answers are subset of reference document

Question: What remote control vehicle did Tesla make?

Document: ... He also built a wireless controlled **boat**, one of....

Concept: Allow system to look at question and document iteratively, the way a human with a bad memory (e.g. Tom Henighan) might.

Encode Step: Interpret Question with no knowledge of document, & vice-versa

First attention Iteration: Re-interpret Question with knowledge (attention) of Document, and vice-versa

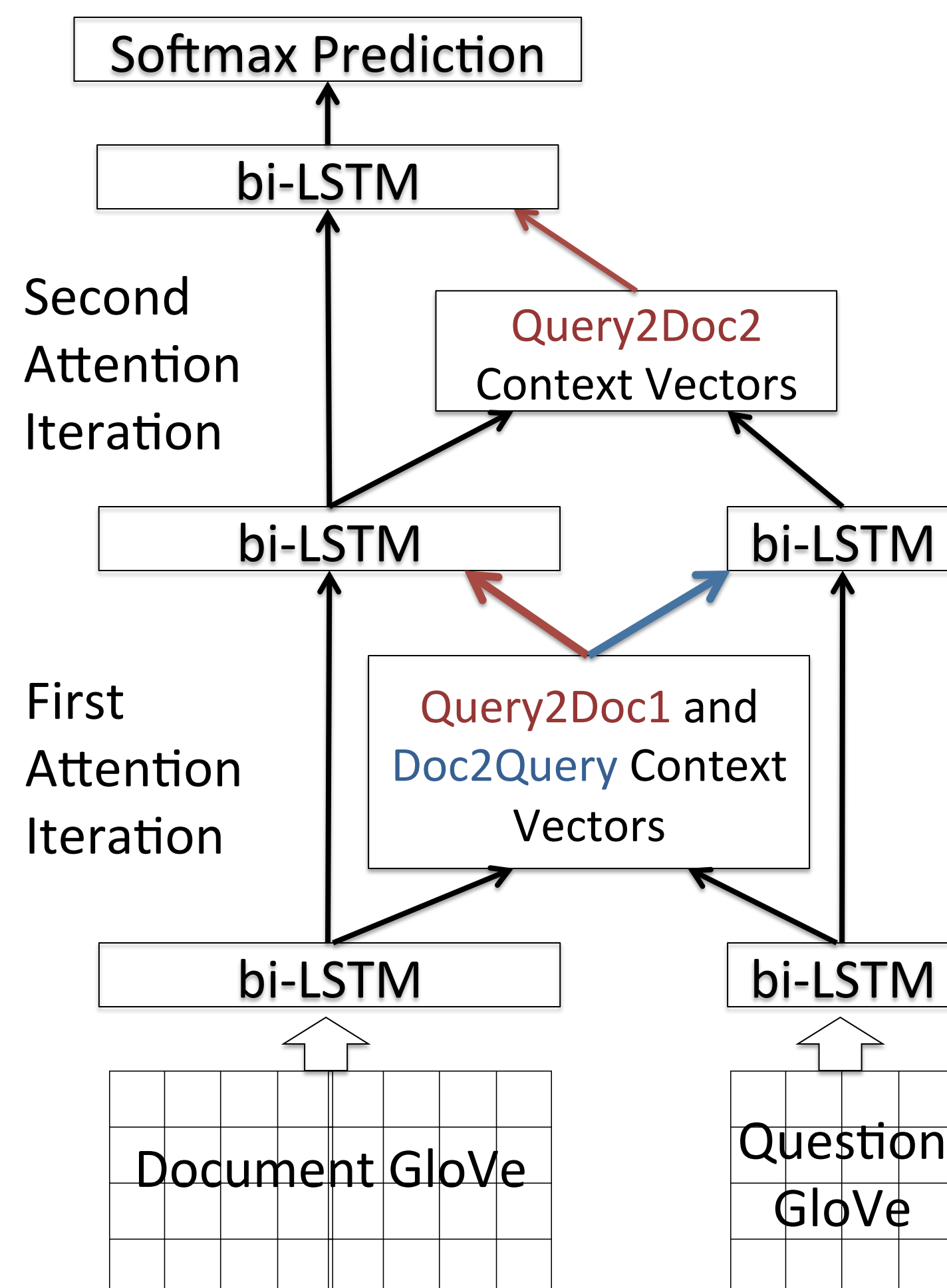
Second Attention Iteration: Interpret document yet again with knowledge (attention) of the document-aware question interpretation

Ablation using 100d GloVe embeddings trained on 6B tokens

Model	Val F1	Val EM
Single Attention w/ extra RNN	56	39
Double Attention	64	46

Currently #7 on Test Leaderboard!

#7 62.015 50.309 success henighan

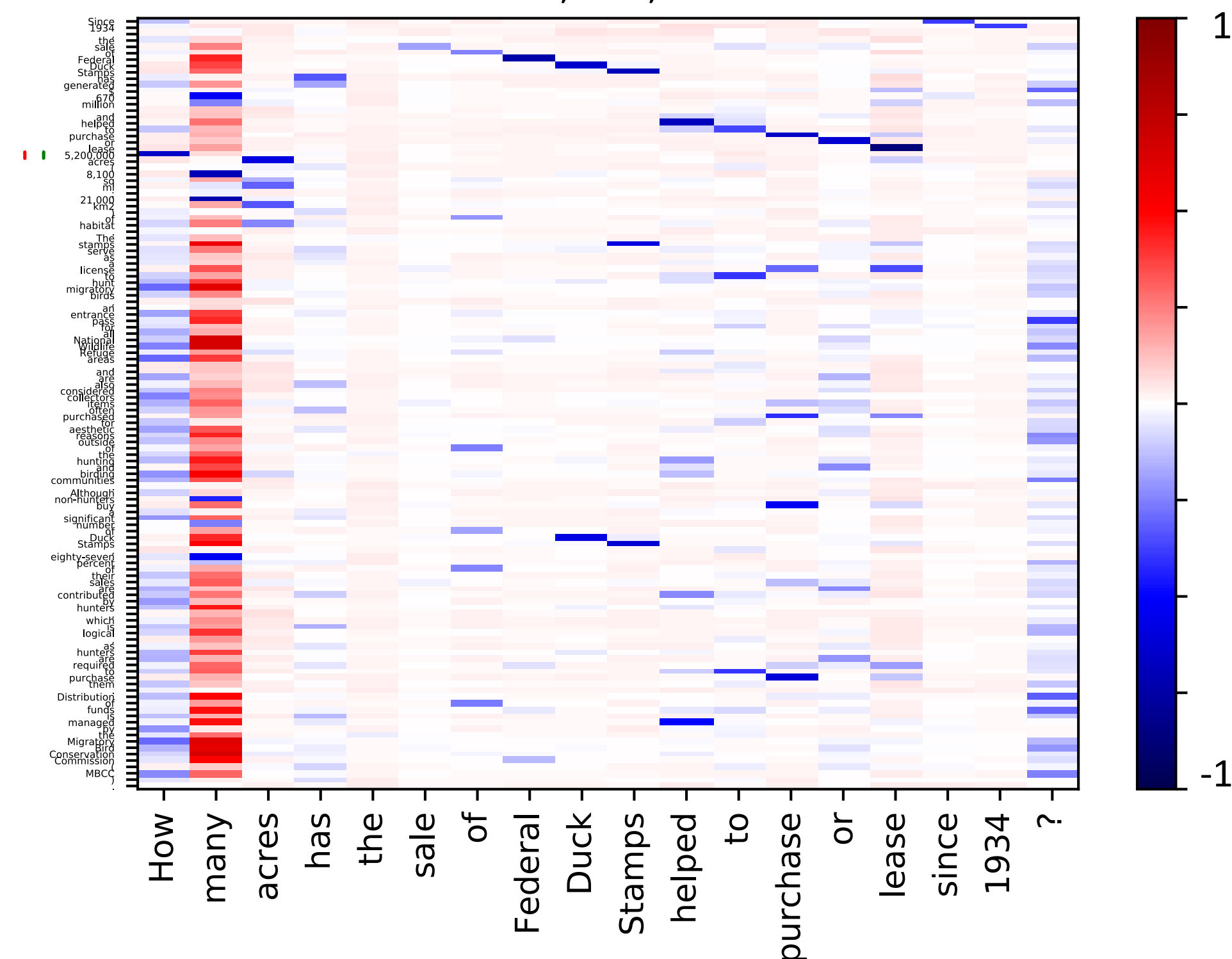


Analysis: How can we tell how the system's attention is changing in the first and second iteration?

Answer: Compare Query2Doc2 Context Vectors!

Below is map of (Query2Doc2 - Query2Doc1)

Answer: 5,200,000 acres



System shifts its attention towards question word "many" and away for nearly all document words. For what query words did attention change most in second iteration?

Shifted Attention Away from: Shifted Attention Towards:

Why
How
What
It

Year
Date
Many
Color