Question Answering

Shubham Gupta

February 22, 2020

1 Introduction

- For the number of massive docs present, need to retrieve only relevant information
- Question-Answer done in two parts:
  - Docs that (might) contain the answer
  - Finding answer in a paragraph or doc. Called as **reading comprehension**

2 SQuAD

- Questions have a passage and a answer.
- Answer is always a subsequence of words from the passage i.e they occur in the same order in the passage. Also called a **span**

3 Evaluation

- 3 answers
- Scored on:
  - Exact match with *gold answers*
  - F1 score
- Ignore puntuation and articles(a, an, the, etc)

4 SQuAD 2.0

- Has some questions which don't have answers in the passage
- For questions that did not have an answer, *NoAnswer* was scored 1 and anything else 0, for both exact match and F1

5 SQuAD limitations

- Only span-based answers(no yes/no, counting, implicit why, etc)
- Questions were constructed by looking at the passages
  - Not genuine info needs
  - Greater lexical and syntactic match for these answers than we would find IRL
- Multi-fact/sentence inference beyond coreference missing

6 Stanford Attentive Reader

- Neural QA system
- Simpliest system
- BiLSTM with Attention
- Deep BiLSTM works better
- Input: Word vector + One hot encoding of POS and NER tags + Term frequency + Exact match(if word occurs in the question)

7 BiDAF

- **Key idea**: Attention Flow Layer
- Attention should flow both ways - from context to question and vice version
- Make similarity matrix S_{ij}. Big concatened vector of $c_i; q_j; c_i \circ q_j$

8 FusionNet

- Attention functions:
  - MLP (Additive) form. Space O*mnk), W is kxd
  - Bilinear(product) form: Space O((m+n)k). Smaller space and used non-linearity

9 TLDR

- Most of the best solutions have bert. **Use BERT** for your solutions.