Announcements

- A5 due today
- FP out: check-in due May 1
  final report due May 13
  (no slip days)
- Course evals

Recap Knowledge base QA

When was Samuel L Jackson born? } question

⇓ semantic parsing

$\lambda d.\ bday(SLJ, d)$ } logical form in lambda calculus (or Prolog, SQL, ...)

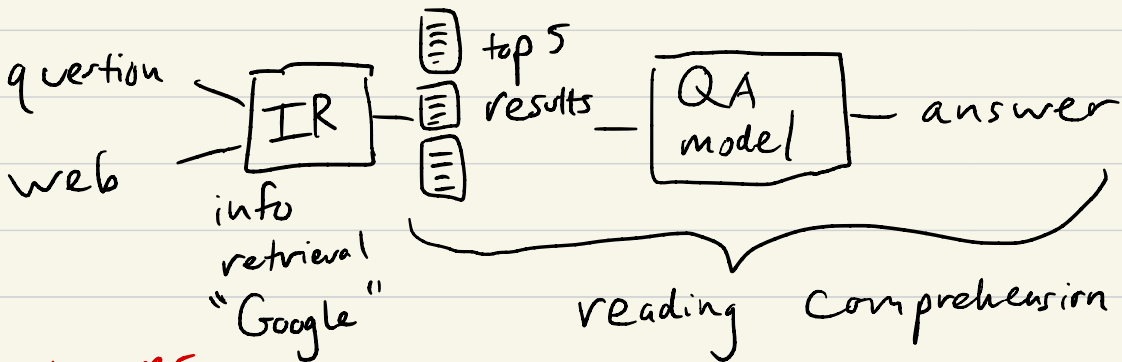
⇓ execute against KB

12/21/48

Today ① Reading comprehension (QA from raw text and not DB)
② Span extraction of answers
③ Attentive reader ④ Other domains/FP

Reading Comprehension

What temperature should I cook chicken to?
What event led to the start of WW1?



Answer

... The assassination of FF took place on < > ... It's regarded as one of the main causes of WW1.

Our focus

- Assume one document (one paragraph)
- Simple questions that don't require much "reasoning"

Span extraction

Simpler ex:

| The assassination of FF caused WW1.
| Gavrilo Princip killed him on June 28, 1914.

Q: What event led to WW1?

Baselines

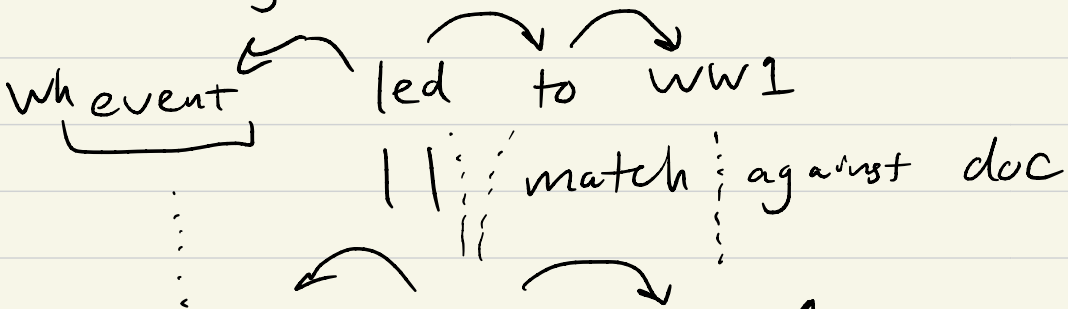
① n-gram "sliding window"
Look at spans of passage, check n-gram overlap w/ Q

[The a. of FF caused WW1] ∩ {question n-grams}

n=2: "what event" -- " to WW1"
n=1: WW1 matches doc

high word overlap ⇐⇒ span might have answer

② Parsing



One FP option: investigate combining these ideas w/ neural model

More general framework: pick a span from the document scored w/a neural net

The a. of FF caused WW1

↑ span start

↑ span end

Model: input: (doc, q)

output: (start, end) pair

start: distribution over positions in doc

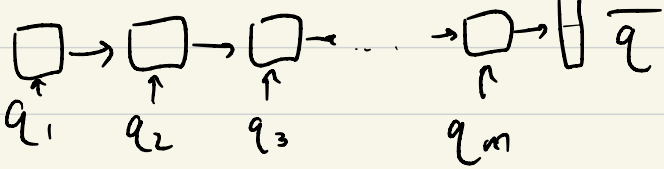
This is a bit weird! Maybe more natural to model dist over NPs

Formally: model $P(start \mid doc, q)$, $P(end \mid doc, q)$ } share most NN params

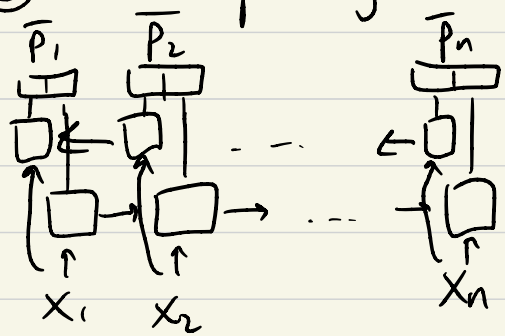
To extract answer: find highest-scoring span (start, end) such that the span is not more than 5–10 words. $P(start \mid \ldots) P(end \mid \ldots)$

Attentive Reader

① encode question into a vector $\bar{q}$



② encode passage into $\bar{p}_1 \cdots \bar{p}_n$ w/ biLSTM



bilinear layer "weighted dot product"

③ compute start/end probs

$$P(\text{start} \mid p, q) = \underset{n \text{ values}}{\text{softmax}} \begin{bmatrix} \bar{q}^T W^{start} \bar{p}_1 \\ \bar{q}^T W^{start} \bar{p}_2 \\ \vdots \end{bmatrix}$$

$$P(\text{end} \mid p, q) = \text{softmax}_i \left[\bar{q}^T W^{end} \bar{p}_i \right]$$

Compare $\bar{q}$ to each position in the passage, score how good that position is

Params 1x LSTM for q
1x biLSTM for $\bar{p}$
W^{start}, W^{end} ← need to be different

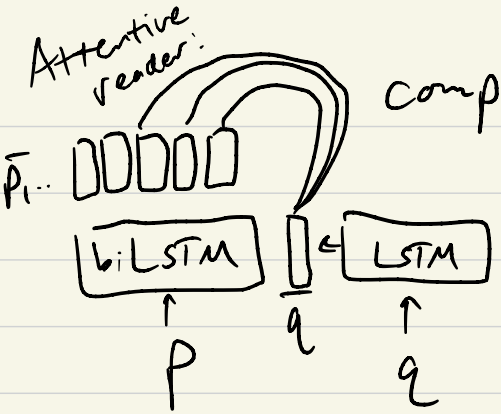
word embeddings (GloVe)

Example from SQuAD

Stanford Question Answering Dataset:
paragraphs from Wikipedia
annotators pick spans + write questions

Model learns:
1) Answer type
2) Match nearby passage words w/question

Attentive reader is a good baseline but not powerful enough
Next time: two modules in our model to make it better



compare $\bar{p}_i$ w/ $\bar{q}$ to determine start+end

looks like attention

BioASQ dataset

Train on (non-bio) Wikipedia examples from SQuAD

Test on BioASQ. Will the model do well?

Recent evidence suggests that abnormal activation of cyclin-dependent kinase 5 (cdk5) is a critical prodeath signal in stroke. However, the mechanism(s) by which cdk5 promotes death is unclear. Complicating the role of cdk5 are the observations that cdk5 can exist in multiple cellular regions and possess both prosurvival and prodeath characteristics. In particular, the critical role of cytoplasmic or nuclear cdk5 in neuronal jury, in vivo, is unclear. Therefore, we determined where cdk5 was activated in models of ischemia and how manipulation of cdk5 in differing compartments may affect neuronal death. Here, we show a critical function for cytoplasmic cdk5 in both focal and global models of stroke, in vivo. Cdk5 is activated in the cytoplasm and expression of DNcdk5 localized to the cytoplasm is protective. Importantly, we also demonstrate the antioxidant enzyme Prx2 (**peroxiredoxin** 2) as a critical cytoplasmic target of cdk5. In contrast, the role of cdk5 in the nucleus is context-dependent. Following focal ischemia, nuclear cdk5 is activated and functionally relevant while there is no evidence for such activation following global ischemia. Importantly, myocyte enhancer factor 2D (MEF2D), a previously described nuclear target of cdk5 in vitro, is also phosphorylated by cdk5 following focal ischemia. In addition, MEF2D expression in this paradigm ameliorates death. Together, our results address the critical issue of cdk5 activity compartmentalization, as well as define critical substrates for both cytoplasmic and nuclear cdk5 activity in adult models of stroke.

What type of enzyme is **peroxiredoxin**?

May do poorly because:

1. UNK words in GloVe
2. Rare words have poor embeddings
3. Answer type not seen on Wiki
4. Diff. sentence structure