Today

1. Improvements to attentive reader
2. State-of-the-art reading comprehension
3. Pre-training + ELMo

Recap

Reading comprehension: find an answer to a question by picking a span from a document (also called context, passage)

Attentive reader

$$P(\text{start} \mid p, q) = \text{softmax}(\tilde{\beta}_i^T W_{\text{start}} \frac{\alpha}{q})$$
$$P(\text{end} \mid p, q) = \text{softmax}(\tilde{\beta}_i^T W_{\text{end}} \frac{\alpha}{q})$$

The assassin of F F F caused...
Announcements

① FP links fixed
② FP check-in due Nov 24

---

Improvements to Attentive Reader

① We want to capture more interaction between the question + passage

Barack Obama is the pres...

q: Who is Barack Obama?

LSTM

q

q may not “remember” Barack Obama

We want to explicitly find word matches between q and p.
Context-to-query attention

\[
\alpha_i = \text{softmax}\left( \bar{x}_i^T \bar{q}_j \right) \quad \bar{q}_j
\]

\[
\bar{v}_i = \sum_j \alpha_i q_j
\]

Barack Obama is... \( \bar{v} \) "knows about" relevant question span prediction stuff for this passage word

\( \bar{v}_i \) "knows" Barack is in \( q \)
2. Question encoding: before we had:

\[ q \rightarrow q \rightarrow q \rightarrow q \]

\[ q_1, q_2, q_3 \]

Instead of this, now we compute a weighted sum of \( q_i \)s

\[ \beta_j, \text{ attention weights over the question} \]

\[ \beta_j = \text{softmax}_j (u - q_j) \]

\[ \bar{q} = \sum_j \beta_j \bar{q}_j \]

weighted bag-of-words encoding

These tricks: DrQA (Chen et al.)