

# CS 378 Lecture 23

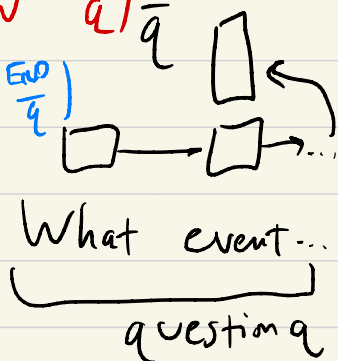
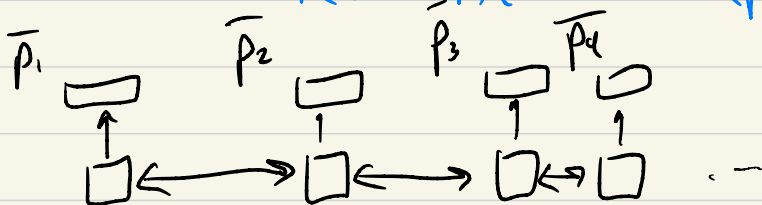
- Today
- ① Improvements to attentive reader
  - ② State-of-the-art reading comprehension
  - ③ Pre-training + ELMo

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

Attentive reader

$$P(\text{start} | p, q) = \text{softmax}(\bar{p}_i^T W^{\text{START}} \bar{q})$$

$$P(\text{end} | p, q) = \text{softmax}(\bar{p}_i^T W^{\text{END}} \bar{q})$$



The assassin. of F F caused... } doc  
ANS } p

## 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?

```
graph LR; q["q: who is Barack Obama?"] --> LSTM[LSTM]; LSTM --> q_bar["q̄"]; q_bar -.-> passage["Barack Obama is the pres..."]
```

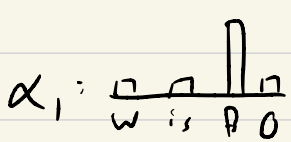
q may not "remember" Barack Obama

We want to explicitly find word matches between q and p.

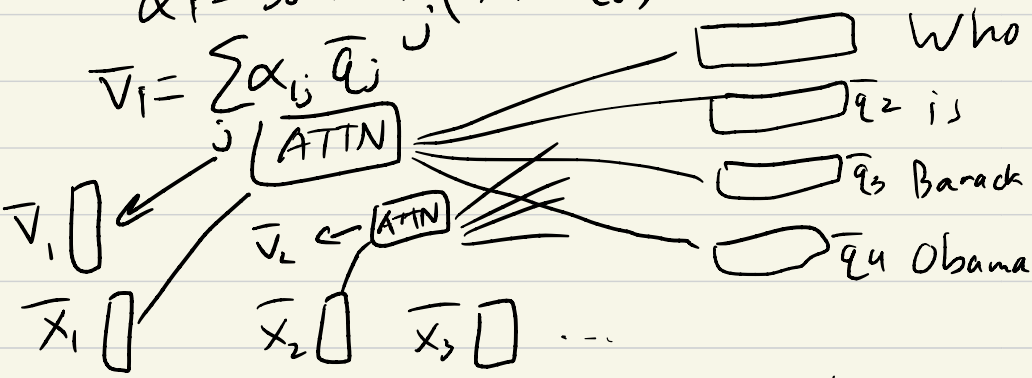
# "Context-to-query attention"

vector of same length as question

$$\alpha_i = \text{softmax}_j(\bar{x}_i^T V \bar{q}_j)$$



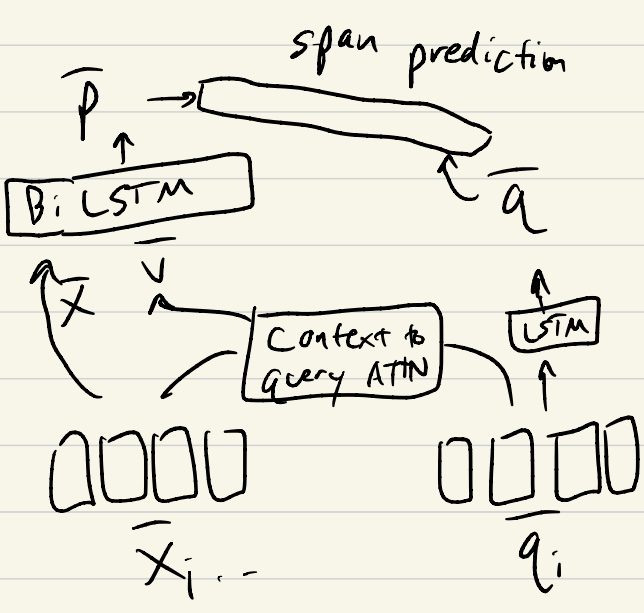
$$\bar{v}_i = \sum_j \alpha_{ij} \bar{q}_j$$



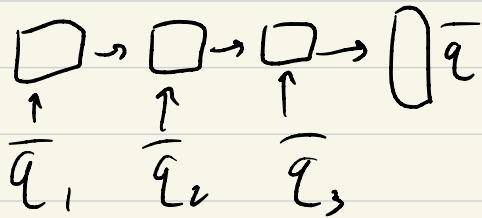
Barack Obama is...

$\bar{v}$  "knows about" relevant question stuff for this passage word

$\bar{v}_i$  "knows" Barack is in  $q$



② Question encoding: before we had:



Instead of this, now we compute a weighted sum of  $\bar{q}_i$ s

$\beta_j$  attention weights over the question

$$\beta_j = \text{softmax}_j(\bar{u} - \bar{q}_j)$$

↑  
weight vector, parameter

$$\bar{q} = \sum_j \beta_j \bar{q}_j$$

weighted bag-of-words  
encoding

These tricks: DrQA (Chen et al.)