

- Barely any multi-fact/sentence
- The standard Attentive Reader

Bidirectional-LSTM, attention

$$a_i = \underset{i}{\text{softmax}} (q^T W_s \tilde{p}_i)$$

→ predict start token

$$a_i' = \underset{i}{\text{softmax}} (q'^T W_{s'} \tilde{p}_i)$$

→ predict end token

## • BiDAF

- attention flow layer: attention should flow through both ways. From context to question, and from question to context.
- (C2Q)  
Context-to-Question attention

attention functions:

- MLP (additive) form:

$$S_{ij} = s^T \tanh(W_i c_i + U_j q_j)$$

- Bilinear (Product) form:

$$S_{ij} = c_i^T W q_j$$

$$S_{ij} = c_i^T U^T V q_j$$

$$S_{ij} = c_i^T W^T D W q_j$$

much efficient  
attention