

# Exploring Embedding and Attention Improvements to BiDAF SQuAD Model

Colin Dolese

## BACKGROUND

### Question Answering

- Cloze-style reading comprehension task: Given a question and context paragraph, predict an answer to the question using span of text from the context.
- Stanford Question-Answering Dataset (SQuAD) is recent benchmark reading comprehension data-set. SQuAD 2.0 introduces unanswerable questions.

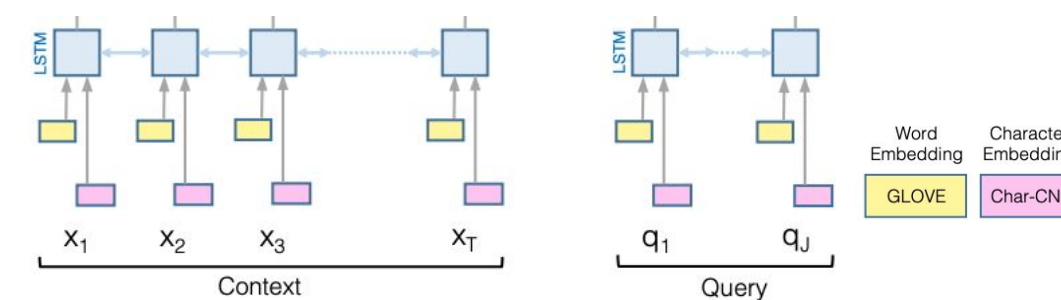
### BiDAF

- Bi-Directional Attention Flow (BiDAF) was early state-of-the-art SQuAD model.
- Attention layer allows both Context-to-Question attention and Question-to-Context attention to flow into modeling layer.

## APPROACH AND MODELS

### BiDAF + Character Embeddings

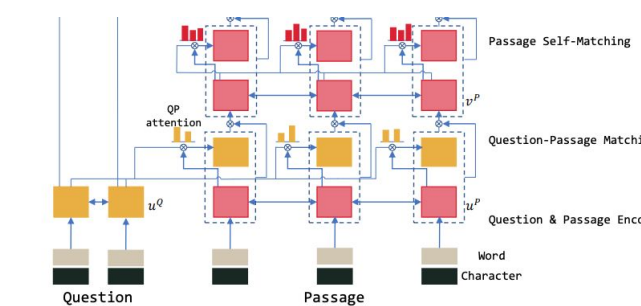
- Baseline BiDAF uses only word-level embeddings. Original BiDAF model combines word-level and character-level embeddings.
- (1) Embed character's into vectors  
(2) Feed through CNN  
(3) Max-pool output



### Self-Attention

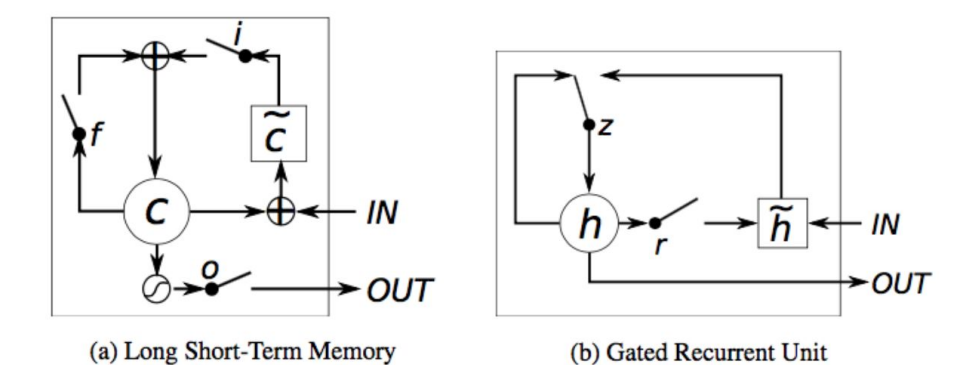
- Self-attention layer based on R-Net's "self-matching attention."
- For each question-aware context paragraph word representation compute attention pooling vector over whole context:

$$c_t = att(v^P, v_t^P)$$



### BiDAF with GRU RNN

- Baseline BiDAF and original BiDAF models both use a Bi-Directional Long-Short-Term (LSTM) RNN for encoder and modeling layer.
- R-Net uses Gated Recurrent Unit (GRU) RNN



## EXPERIMENTS

- SQuAD metrics:
  - Exact Match (EM) Baseline: 55.49
  - F1 (precision and recall) Baseline: 58.62
  - Answer vs No Answer (AvNA) Baseline: 64.68
- BiDAF + Character Embeddings tested against baseline
- Self-Attention tested varying input attention (context state, context-to-question attention, question-to-context attention)
- GRU tested with baseline and character-embeddings

## RESULTS



Character Embeddings and GRU results. Baseline (pink), Character Embeddings (blue), Baseline with GRU (orange), Character Embeddings with GRU (red)



Self-Attention results. Baseline (light blue), context-to-question (c2q) (green), c + c2q (orange), c + c2q + q2c (dark blue)

BiDAF + Character Embeddings	EM: 61.37	F1: 58.18	AvNA: 67.05
Self-Attention with context and context-to-question input	EM: 58.81	F1: 55.67	AvNA: 65.94
<b>BiDAF + Character Embeddings with GRU</b>	EM: <b>64.12</b>	F1: <b>60.85</b>	AvNA: <b>70.22</b>