**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.

**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

**APPRAOH 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:
  \[ a_j = \text{att}(v^p, v^q) \]

**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

**RESULTS**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>EM</th>
<th>F1</th>
<th>AvNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiDAF + Character Embeddings</td>
<td>61.37</td>
<td>58.18</td>
<td>67.05</td>
</tr>
<tr>
<td>Self-Attention with context and context-to-question input</td>
<td>58.81</td>
<td>55.67</td>
<td>65.94</td>
</tr>
<tr>
<td>BiDAF + Character Embeddings with GRU</td>
<td><strong>64.12</strong></td>
<td><strong>60.85</strong></td>
<td><strong>70.22</strong></td>
</tr>
</tbody>
</table>

Image Credits: Seo et al. (Bi-Directional Attention Flow for Machine Comprehension, Microsoft Research) (R-Net: Machine Reading Comprehension with Self-Matching Networks)