Problem

**SQuAD 2.0**

Stanford University

Given a query (question) and context pair, our models have to:
- Detect if the question is answerable
- If it is, find the answer span within the context paragraph

---

**Query**
“What president eliminate the Chiraz tax on the curriculum?”

**Context**
“Chiraz tax: consider 1884–88; eliminate the curriculum of Christianity from the curriculum. Which president is in support of this direction? While this was the most crucial figure in the association of American higher education, he was motivated not by a desire to secularize education, but by transforming the curriculum, which was then...”

**Target Answer**
“Charles W. Eliot”

---

**Base Model: QANet**

We first implemented the QANet [1]. It uses the transformer architecture and convolutions to replace RNN layers in QA models.

---

**Approach 1: Multiheaded Context-Query Attention**

We changed the basic context-query attention layer into a multihead context-query attention. We hoped to capture different types of context-query attention with multiple heads.

---

**Approach 2: Output Layer Modules**

**Single FC Layer**

**(V) Answerability Verifier**

**SEQ) Sequential Span Predictor**

**(COND) Conditioning Ending Point Predictor**

---

**Final Best Model Output Layer (Base + 1) – 2**

We use the auxiliary loss from (a) verifier and (b) sequential span predictor together with the basic boundary predictor loss.

---

**Result**

Compared to the original QANet, our model achieves a much better result for the answerability and answer span prediction. Our verifier and sequential prediction modules also help outperform the basic QANet model. We achieved an ensemble F1 score of 69.57 and EM score of 66.54.

---

**Table**

<table>
<thead>
<tr>
<th>Model</th>
<th>Embedding</th>
<th>VQA</th>
<th>F1</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLU+H</td>
<td>W + C</td>
<td>88.61</td>
<td>81.49</td>
<td>78.12</td>
</tr>
<tr>
<td>QANet</td>
<td>W + C</td>
<td>88.43</td>
<td>81.46</td>
<td>77.97</td>
</tr>
<tr>
<td>QANet + y = 1/50</td>
<td>W + C</td>
<td>79.94</td>
<td>67.93</td>
<td>66.82</td>
</tr>
<tr>
<td>Multitask QANet + y = 1/50</td>
<td>W + C</td>
<td>78.90</td>
<td>67.52</td>
<td>66.24</td>
</tr>
<tr>
<td>Multitask QANet + y = 1/2</td>
<td>W + C</td>
<td>82.19</td>
<td>74.94</td>
<td>71.18</td>
</tr>
<tr>
<td>Multitask QANet + y = 1/10</td>
<td>W + C</td>
<td>79.34</td>
<td>66.03</td>
<td>64.39</td>
</tr>
<tr>
<td>MultiQANet + y = 1/50 + COND</td>
<td>W + C</td>
<td>78.75</td>
<td>65.77</td>
<td>64.84</td>
</tr>
<tr>
<td>Ensemble (arch=H)</td>
<td>W + C</td>
<td>80.40</td>
<td>69.97</td>
<td>66.94</td>
</tr>
</tbody>
</table>

We see high true negative and positive rate on the dataset, with an ensemble Acc of 74.48%.

---

**Diagrams**

- Answer Span Length (Words)
- Answer Span Length Distribution (Words)
- Answer Span Length Distribution (%)

---

*Note: This is a sample text for demonstration purposes.*