R-Net and Friends
Megan Worrel, Amrita Palaparthi, Ani Vegesana

Problem
Task: Question Answering (reading comprehension) is the task of automatically answering a question given a paragraph of relevant context
Importance: Question Answering is critical to determining how well models can understand and draw information from text
Contribution: R-net explores the effect of additional forms of attention on SQuAD performance; however, it does not combine these with meaningful additional input features. We explore the gains in performance on SQuAD 2.0 that can be achieved through the simultaneous use of R-net attention mechanisms and feature engineering as proposed by DrQA, coupled with hyperparameter tuning and ensemble techniques.

Background
Problem Setup and Notation: Given the i-th context paragraph $c_i$ and question $q_i$, predict the start and end indices of the answer within the context if it exists. These indices are represented as the logits $p_{true}$ and $p_{false}$.
Training Procedure: All models were trained for 30 epochs, using the AdaDelta optimizer with cross-entropy loss.
Evaluation: We use F1 score as the primary metric for evaluation on the SQuAD 2.0 validation and test datasets.
Baseline: We evaluate our results against a baseline Bidirectional Attention Flow (BiDAF) model using word embeddings, which achieves an F1 score of 60.65.

Methods
Additions to the baseline:
1. Character level embeddings
2. Feature Engineering
3. R-net gated and self attention
4. Hyperparameter tuning
5. Ensemble

DrQA Additional Input Features
6. Exact Match Features
7. Context vectors
8. Hyperparameter tuning
9. Ensemble

Feature Engineering Results

<table>
<thead>
<tr>
<th>Model</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiDAF</td>
<td>60.65</td>
</tr>
<tr>
<td>R-net</td>
<td>68.16</td>
</tr>
<tr>
<td>R-net + Feature Engineering</td>
<td>68.33</td>
</tr>
<tr>
<td>R-net + Hyperparameter Tuning</td>
<td>68.31</td>
</tr>
<tr>
<td>R-net + ensemble</td>
<td>68.33</td>
</tr>
</tbody>
</table>

R-net Gated and Self Attention

*Mutually Recurrent*: We implemented R-net gated and self attention as described in the R-net paper.

Non-Recurrence:
- For efficient training, we switched to a non-recurrent approach.
- Attention computation occurs as input to the LSTM.
- This eliminates the need for expensive manual iteration, leveraging much faster matrix computations and GPU processing.

Conclusions

- Implementing R-net attention mechanisms in conjunction with DrQA's additional input features results in a substantial increase in performance over our baseline model on Question Answering for SQuAD 2.0.
- Each of the four main components of our approach - DrQA additional input features, R-net attention mechanisms, hyperparameter tuning, and ensemble - built upon one another to provide an incremental increase in F1 score.
- Ensemble: The variety of models we trained based on confidence score enabled our approach to perform well on a wide range of inputs, further augmenting performance.

Analysis
Summary of Highest-Performing Improvements:

<table>
<thead>
<tr>
<th>Incremental Improvement</th>
<th>Maximum F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensembling</td>
<td>68.33</td>
</tr>
<tr>
<td>R-net + Feature Engineering</td>
<td>68.33</td>
</tr>
<tr>
<td>R-net + Hyperparameter Tuning</td>
<td>68.31</td>
</tr>
<tr>
<td>R-net + ensemble</td>
<td>68.33</td>
</tr>
</tbody>
</table>

- Exact match and ADE each individually boosted F1 score, but combining them did not result in significant improvement. This aligns with DrQA's finding that these two features play complementary roles.
- BiDAF's attention mechanism performed more effectively than R-net gated attention alone, as they play a similar role, but BiDAF was multiplicative rather than additive attention.

Out-of-vocabulary Words

- Question: What is the name of the garden for the University Library?
  - Context: Another important library is the University Library founded in 1619, it is home to over two million items. The building was designed by Athanasius Kircher.

- Question: What was the name of the garden?
  - Context: The garden was designed by Joris de Blesk and was opened on 12 June 2002. It is one of the largest and most beautiful roof gardens in Europe with a view of more than 10,000 m² (110,200 sq ft) and plants covering 8,171 m² (88,000 sq ft). At the university garden it is open to the public every day.

- Prediction (baseline): Maor Budyński and Zbigniew Badowski
- Prediction (ensemble): Maria Budyńska

Anchors
- Question: What is the name of the garden for the University Library?
  - Context: Another important library is the University Library founded in 1619, it is home to over two million items. The building was designed by Athanasius Kircher.

- Question: What was the name of the garden?
  - Context: The garden was designed by Joris de Blesk and was opened on 12 June 2002. It is one of the largest and most beautiful roof gardens in Europe with a view of more than 10,000 m² (110,200 sq ft) and plants covering 8,171 m² (88,000 sq ft). At the university garden it is open to the public every day.

- Prediction (ensemble): Maria Budyńska
- Prediction (behavior-based): Maria Budyńska

References