Problem Description: The SQuAD 2.0 challenge, a Question-Answering task. One of the most important Natural Language Processing challenges.

Approach:

- Also trained a simpler model that used only a Context-Query attention layer after BERT to reduce overfitting.

Analysis:

- The original BERT computes interactions between all words in the input. However, for Q&A, the interactions between context and query as separate groups may require more emphasis.
- CQ-BERT and QANet + BERT outperform BERT-small.
- QANet + BERT could outperform CQ-BERT with data augmentation.
- Failed to surpass BERT-large, but with more resources to perform further hyperparameter tuning it is likely that this result could be overturned.
- Overfitting on BERT-large variants may also contribute to the inferior performance.

Model Architectures

- QANet + BERT
- Context-Query Attention + BERT

Conclusion

- QANet and BERT can be combined to achieve near state-of-the-art results on the SQuAD 2.0.
- Using Context-Query Attention as the output layer for BERT may prevent overfitting and performs better, with an F1-score of 80.23.
- Our final model places 5th on the CS224N 2019 Winter leaderboard.

Future work:
- Hyperparameter tuning on BERT-large
- QANet Data augmentation

References