Semi-Supervised Question Answering:

Generative Augmentation in SQuAD2.0

Alexander Lu
alu2@stanford.edu

Problem —

Extractive question-answering has been extensively investigated, but remains heavily dependent on extensive hand labeling to construct datasets. We investigate semi-supervised question-answering methods to leverage unlabeled data in one-shot approaches to SQuAD2.0.

Approach —

The SQuAD2.0 dataset is used with imposed partitions of 50%, 75%, 100% of total question, answer labels. The BiDAF model (without character embeddings) serves as the baseline. A variety of data augmentation methods and model adjustments are then explored to improve model performance under these conditions.

Model Simplification
Model size is reduced by replacing LSTMs with GRUs.

Naïve data augmentation
NER and POS-based heuristic extracts candidate answers from text. Surrounding context is imputed as the corresponding question.

Neural question generation
A generative model trained on the labeled dataset partition is used to generate new questions from the unlabeled fraction.

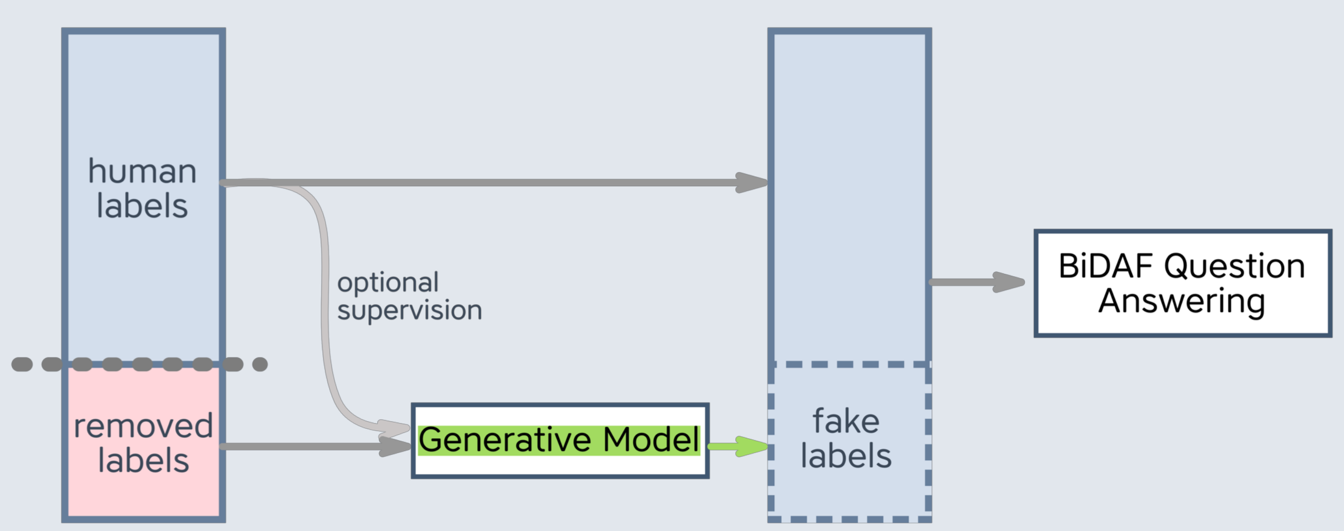

Figure 2: Semi-supervised Training Scheme

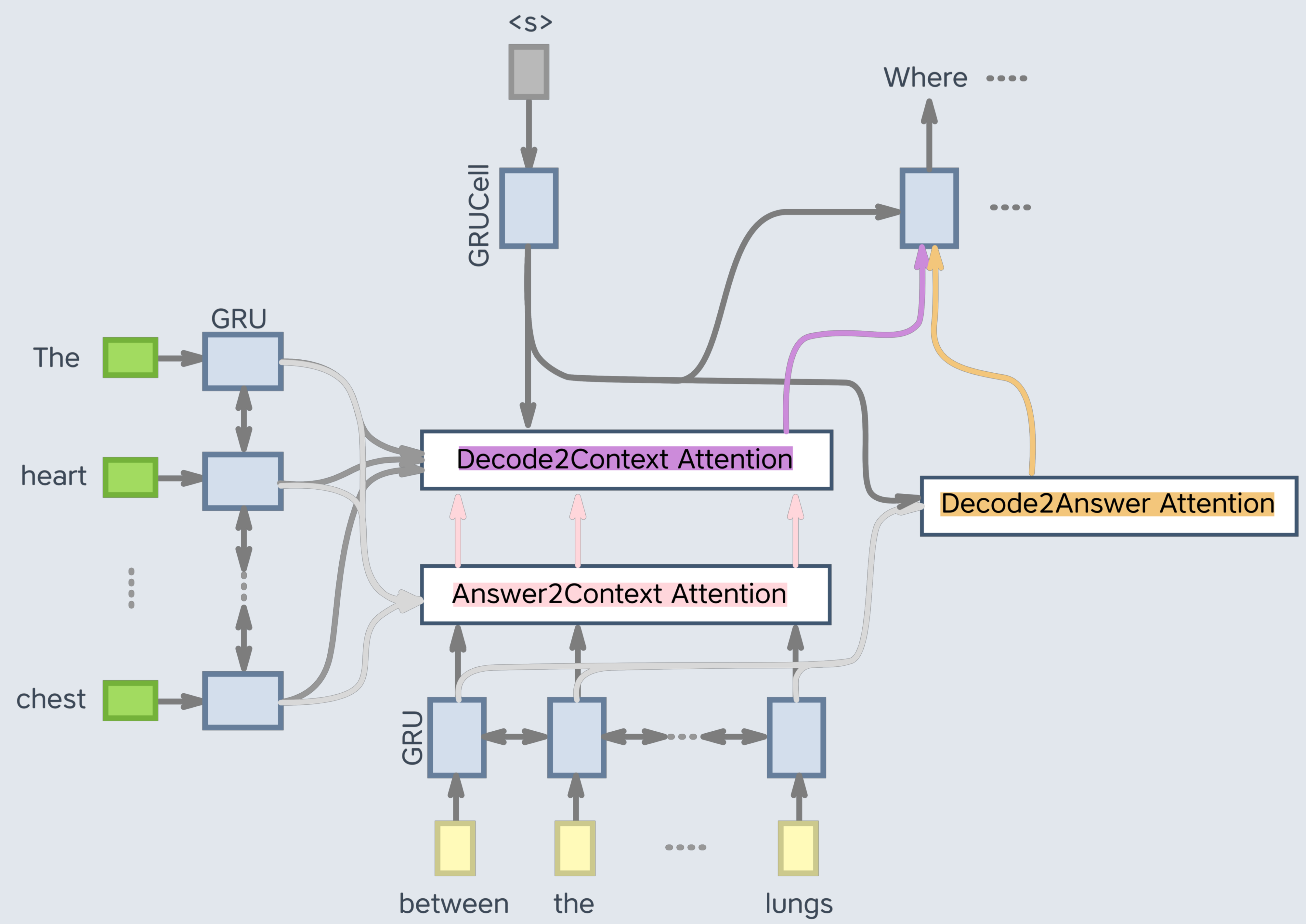

Figure 1: Generative Model Architecture.

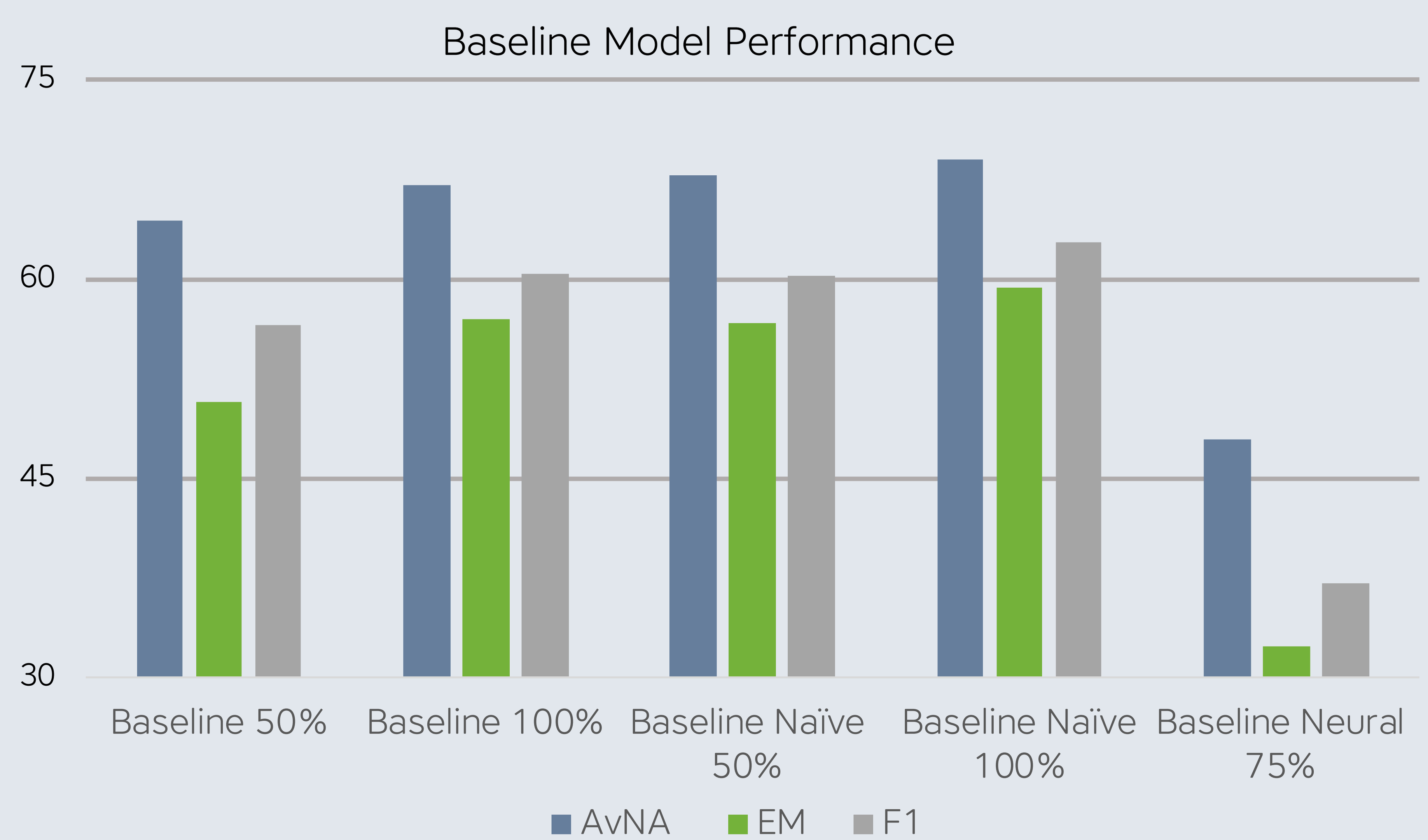

Figure 3: Model Performance Overview

Analysis —

Naïve data augmentation proves impressive in ameliorating effects of training data restriction. This one-shot formulation resembles the unsupervised stage of training BERT, confirming that context neighborhoods may enable effective representations of textual meaning.

Additionally, utilizing the generative model to include fluent questions fails to improve performance relative to naïve data augmentation baselines 😪. This suggests difficulty in capturing relationships between answer and paragraph content.

Conclusions —

Through various data augmentation methods across different extents of training data restriction, we show that model performance in extractive question-answering is dependent both on exposure to a wide range of context inputs, as well as the quality of associated questions. In the future, it can be investigated whether such these methods can have similar impacts on more complex SOTA models.

References

Zhilin Yang, Junjie Hu, Ruslan Salakhutdinov, and William W. Cohen. Semi-Supervised QA with Generative Domain-Adaptive Nets. arXiv preprint arXiv:1702.02206v2, 2017.
Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603, 2016.
Tong Wang, Xingdi Yuan, and Adam Trischler. A Joint Model for Question Answering and Question Generation. https://arxiv.org/pdf/1706.01450.pdf

Acknowledgements

I would like to thank the CS224n teaching staff for their support throughout the quarter and for this opportunity. I would also like to thank the TAs for their excellent mentorship over the duration of this project.