XCS224U - Natural Language Understanding

Experiment Protocol

Natural Language Inference

Ankur Dhoot

March 2021

Contents

1 Introduction

This experiment investigates the application of neural attention mechanisms to the field of NLI. Natural language inference (NLI) is a subfield of NLU that focuses on whether a machine can reasonably infer a natural language hypothesis h from a given premise p (MacCartney and Manning, 2009).

2 Hypothesis

Attention mechanisms were introduced in the context of neural machine translation to mitigate the difficulty of trying to encode an entire sentence into a fixed length vector (Bahdanau et al., 2015). Attention mechanisms have also been effectively used in the context of encoder-decoder systems for NLI (Rocktaschel et al., 2015). This experiment aims to investigate the effectiveness of attention on the SNLI dataset *across the relation classes*.

Hypothesis: Attention helps neural models predict entailment / contradiction classes better than the neutral class.

3 Data

This experiment will utilize the SNLI dataset which contains over half a million examples evenly distributed across the three relation classes. (Bowman et al. 2015). The dataset also contains 10K dev and 10K test examples.

4 Metrics

Most papers that compare models on the SNLI dataset use accuracy as their primary metric. While accuracy can be misleading in the case of imbalanced datasets, accuracy is generally a valid metric on the balanced SNLI dataset.

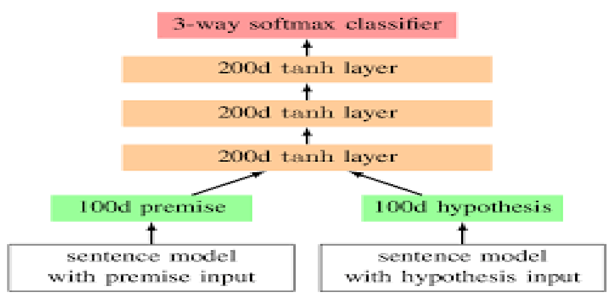
However, this experiment aims to investigate the effect of attention on each relation class. Since accuracy isn't a per-class metric, accuracy isn't as well-suited for this experiment. Therefore, this experiment will use the combination of precision and recall, the F-1 score.

Time permitting, it could be illuminating to look qualitatively look at a sample of examples that the classifier correctly predicts for each class and compare those with examples the classifier incorrectly predicts. Looking at the attention weights for these examples could provide insight into what the classifier is focusing on between the premise and hypothesis.

5 Models

5.1 Baseline

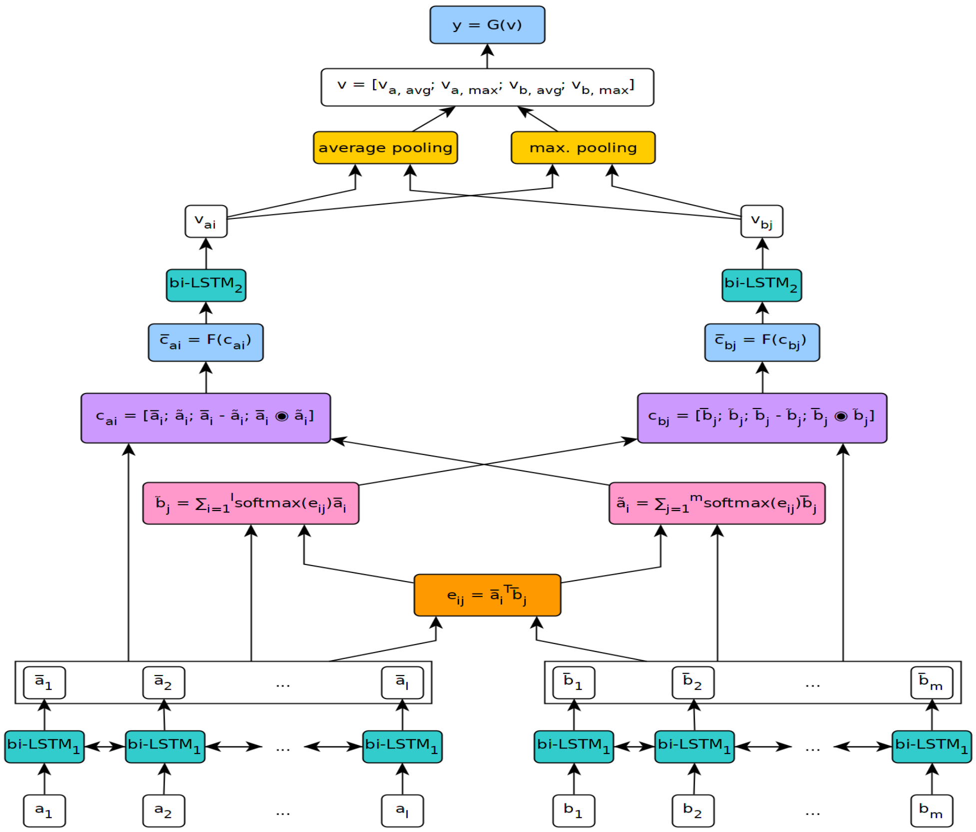
The baseline model will be the original neural model from Bowman et. al (2015). This model runs the premise and hypothesis through an RNN and uses the RNN output vector as the representation for the premise and hypothesis. The vectors are then concatenated and fed into a deep network with a softmax classifier at the top.



5.2 Attention Model

The attention model will be ESIM. (Chen et al., 2017)

This model runs the premise and hypothesis through an RNN. Attention weights are then computed between all the hidden states in the premise and hypothesis using simple dot product attention. The attention vectors are then calculated using softmax weighting. A combination of the hidden states and attention vectors is passed through a projection layer. Finally, another RNN is run over the output of the projection. The output of the RNN is passed through a deep layer with a softmax classifier on top.



6 General Reasoning

Attention works by allowing the hypothesis sentence to focus in on parts of the premise (and vice-versa). In the case of entailment or contradiction, there are likely to be specific alignments that attention can focus on that either entail or contradict each other. In the case of a neutral relation, such alignments may be less likely to exist. Thus, by building models with and without attention, this experiment aims to investigate whether attention helps models correctly predict entailment / contradiction examples more than neutral examples.

7 Progress Summary

So far, I have finished implementing the baseline model using Glove embedding vectors. This involved getting setup with Google Cloud to train on GPUs and debugging issues with model implementation.

Work yet to be completed includes building the ESIM model and comparing the performance across the three classes. Time permitting, a qualitative evaluation may also be done.

8 References

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced LSTM for natural language inference. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pages 1657–1668.

Ankur P Parikh, Oscar Tackstrom, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016.

Tim Rocktaschel, Edward Grefenstette, Karl Moritz Hermann, Tomas Kocisky, and Phil Blunsom. 2016. Reasoning about entailmentwith neural attention. In Proceedings of ICLR.

Reza Ghaeini, Xiaoli Fern, and Prasad Tadepalli. 2018. Interpreting Recurrent and Attention-Based Neural Models: a Case Study on Natural Language Inference. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 4952–4957