

# CS224n Assignment 4

March 20, 2017

## Summary

The implemented model is based on paper [1]. For this model, most of heavy lifting is done by the encoder while the decoder is a simple two layer feed-forward network (Prediction Layer). The encoder consists of four layers: a Filter Layer, a Context Representation Layer, a Multi-Perspective Context Matching Layer and an Aggregation Layer. I find the idea of first filter the context paragraphs very appealing, since not every position in the context matters equally for question answering. This is somewhat similar to the attention idea but the paper does not use such terminology. The matching mechanism is simply a way to weight the different dimensions of the representations. Although the authors of [1] claim we need to use three matching layers, Full-Matching, Maxpooling-Matching, Meanpooling-Matching, and support the claim with an ablation study. But I did not find significant differences between using one of the three matching layers and using all three of them. I only used a Full-Matching layer in my submission. The major difference between my implementation and what's described in the paper is that I only used the GloVe word vectors while paper in addition include character level representations as well.

The network is trained using ADAM with exponentially decaying learning rate for 10 Epochs. On average, the F1 score on the validation set fluctuates within 55% to 65%(based on 100 samples), while the EM score fluctuates within 0.41 to 0.52.

It is unfortunate that I ran out of time to explore more ideas. Nevertheless, I learned a lot by doing this project. Thanks for the joint effort of the CS224n teaching team.

## References

- [1] Zhiguo Wang, Haitao Mi, Wael Hamza, and Radu Florian. Multi-perspective context matching for machine comprehension. *CoRR*, abs/1612.04211, 2016.