horizontal line

Sprint 2 Results

Sprint End: 10.27.2019

Sprint Captain: Yirou Ge

Sprint Co-captain: Hongying Liu

**─**

# Sprint Goal

## Get the baseline model:

1. Using RoBERTa to encode gold\_standard dataset with label generated by snorkel, and trying shallow RNN to get the baseline model for this project.

## Find more features:

1. Adding more features to Snorkel to see if we can improve its performance. The features we consider include:

N-Grams: Unigrams, bigrams, and trigrams

Word Frequency & Similarity Coefficient

## BabbleLabble implementation first iteration

1. Run first iteration of BabbleLabble implementation

# Results

Main: We have finished the proof-of-concept and generated the baseline model on Google Cloud. However. As we are testing the model and trying to run first iteration of BabbleLabble implementation, some important issues emerged.

* Baseline Model:

We have run the codes on the Google Cloud platform to encode the sentences in gold standard dataset and noisy dataset, which gave us two sets of embedding data. After that, we trained the shallow RNN with the embedding data of gold standard dataset to generate the baseline model.

During this process, we modified the code for calculating training loss and validation loss for better illustration and calculated F1 score (0.3481). The current result shows that the baseline model is clearly overfitted. Later we plan to add early stopping method and more features to avoid overfitting and improve the model performance.

* Feature Engineering

We first looked at the unigrams, bigrams, and trigrams generated from the gold standard dataset and then added three additional label functions based on the most frequent ngram words. As a result ,the labeling accuracy was increased by 0.024 and F1 score was improved by 0.081



* BabbleLabble implementation

We found out that BabbleLabble actually provides an easier way to create useful labeling functions than Snorkel by allowing the programmers to write natural language explanations rather than explicit label functions. The label model that BabbleLabble uses is actually the same as the one in Snorkel-MeTaL. Therefore, the results generated by BabbleLabble don’t seem to be different from Snorkel if the labeling functions are the same and we will need to discuss with WRI on tomorrow’s meeting to make further plans about this.