Evaluating and Developing Methods of Generating Code-Switched Data
07-400, Spring 2022

Abhishek Vijayakumar,
under Prof. Alan W. Black, Language Technologies Institute
https://inkyubeytor.github.io/category/code-switching.html

March 21, 2022

1 Major Changes

The primary goal of this project has not changed since the previous milestone.

2 What You Have Accomplished Since Your Last Meeting

Since our last meeting, I have determined the learning rate range in which I can train models effectively. I have trained several models with different learning rates and looked at their performance on part of speech tagging to determine at what learning rates training breaks down.

I have also identified a problem with the sentiment analysis task (see Surprises) from the GLUECoS benchmark, while trying to adapt the sentiment analysis task into a usable benchmark for my evaluation

3 Meeting Your Milestone

My primary goal was to identify the hyperparameters I should be using for efficient pre-training to actually have an impact on model parameters. I have achieved this goal in terms of training parameters such as learning rate by observing the thresholds above which such parameters degrade model quality, but I have not determined a set of parameters that consistently gives above-baseline performance when training on any dataset.

4 Surprises

On the sentiment analysis task, it seems that regardless of training scheme, the finetuned model has most outputs belonging to the neutral class. I believe this applies to the published benchmark numbers as well, indicating that the benchmark number is not a good measure of performance. I have not been able to train a model to overcome this problem. We believe this problem is due to the existence of a neutral class in the sentiment analysis task, which is a common cause of such behavior.

5 Looking Ahead

My first goal for the next two weeks is to achieve a statistically significant performance improvement on an evaluation task via some form of pretraining from a baseline model, such as DistilBERT. This is the first step towards providing a metric for how useful generated code-switched data is for pretraining models.

My second goal is to refine the sentiment analysis task into one where models can achieve appreciable improvements over random guessing (or outputting a single label all the time). This will likely be achieved with the removal of the neutral class entirely. While this may make the sentiment models less usable on real-world tasks, they may be able to provide a better picture of dataset quality by focusing on the easier task of discriminating positive and negative sentiment.

6 Revisions to Your Future Milestones

My future milestones may undergo revisions based on whether I can meet my new next milestone. The changes to the sentiment analysis task are my primary addition to my next milestone.

7 Resources Needed

No further resources are needed for this project at this time.