NLP Project Proposal

Mehdi Drissi and **Vivaswat Ojha** and **Pedro Sandoval**
Harvey Mudd College, CA
{mdrissi, vmojha, psandovalsegura}@hmc.edu

1 Project Description

Our goal for our final project is to use transfer learning to fine-tune a model for our document classification task. Recently, a language model called BERT was released that was intended to learn useful representations for a variety of tasks. As that model was able to achieve state of the art performance in multiple NLP tasks after being fine-tuned, we hope that it will also perform well on the hyper partisan sentiment classification as well. Due to the apparent mismatch in the validation and training dataset labels, we will explore training on only a subset of the validation dataset and then checking performance on the remainder.

2 Reading List

We intend to read the following publications to both familiarize ourselves with the BERT model and begin thinking about how we will tweak the model for our classification task.

The papers we will review are:

- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- Vaswani, Ashish, et al. "Attention is all you need." Advances in Neural Information Processing Systems. 2017.
- Dai, Andrew M., and Quoc V. Le. "Semi-supervised sequence learning." Advances in neural information processing systems. 2015.
- Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2010): 1345-1359.

The first paper describes the BERT model. The second paper describes an attention based architecture called the transformer. The transformer originally used for language translation, but can also be used for language modeling. It is the model that BERT is based upon. The last two papers are both about transfer learning. The survey is intended for us to be familiar with usual results obtained from transfer learning, while the semi-supervised sequence learning focuses on transfer learning in NLP. Specifically, it focuses on pre-training a language model as a way to train a better model for a specific task.

3 Dataset

We do not intend to use any additional data. We are thinking of using a portion of the validation dataset for training. More precisely we are thinking of splitting the 200k validation articles to 150k training articles and 50k validation articles. Our motivation for doing this split is that our cross-validation accuracy on the training dataset was very different from our accuracy on the validation dataset. That indicates either the way the articles were labeled or the distribution of articles is very different in the training and validation datasets. As we assume the test dataset to be more similar to the validation dataset than the training dataset, we are choosing to focus on training on the validation data instead of the test data. Our reason for this assumption is the training dataset labels did not appear hand labeled when we examined the data and instead seemed more likely to be labeled based on the source, while we would hope that the test dataset is labeled more carefully.

4 Preliminary results

In Lab 7, we built a baseline multinomial naive bayes model which had a cross validation accu-

Model	CV Accuracy (%)
Lab 7 MNB	78
Lab 8 MNB with Outlines	71

racy of nearly 80 percent. In an attempt to improve this model, we built an extension of in Lab 8. Our extension included the generation of summary outlines for our training and test articles. Our extension did not perform as well as the baseline from Lab 7.

5 Planned Methodologies

Our approach is based on using the ideas provided by the papers in our literature review. In particular, we will be relying on the "BERT" paper by Devlan et al. In this paper, the authors describe a model that they allege has a wide variety of applications for Natural Language Programming tasks including classification. We believe that making use of their model and fine-tuning it to our task is a promising approach for improving the overall performance of correctly classifying the articles in our dataset. We will also explore whether it helps to pretrain BERT on unlabeled documents in our dataset. We may choose to use the training articles as unlabeled data for fine-tuning the language model.