

Sarcasm Detection in Headlines Using Domain Adaptive and Task Adaptive Pre-Training

Raghav, Elyas, Aditya, Sergio, Paul

Introduction

Subtle humor, or sarcasm, is an omnipresent component in everyday conversation, written stories, and news headlines. It can be characterized by subtle shifts in sentiment within phrases or sentences. Recognizing this subtle sarcasm goes hand-in-hand with understanding language as a whole. We chose this topic because we feel there are great strides to be made in sentiment analysis and opinion detection with robust models that can classify bodies of text as sarcasm or not sarcasm. Therefore we chose to employ the latest cutting-edge deep learning and semantics technologies to improve the current state-of-the-art language models.

Method

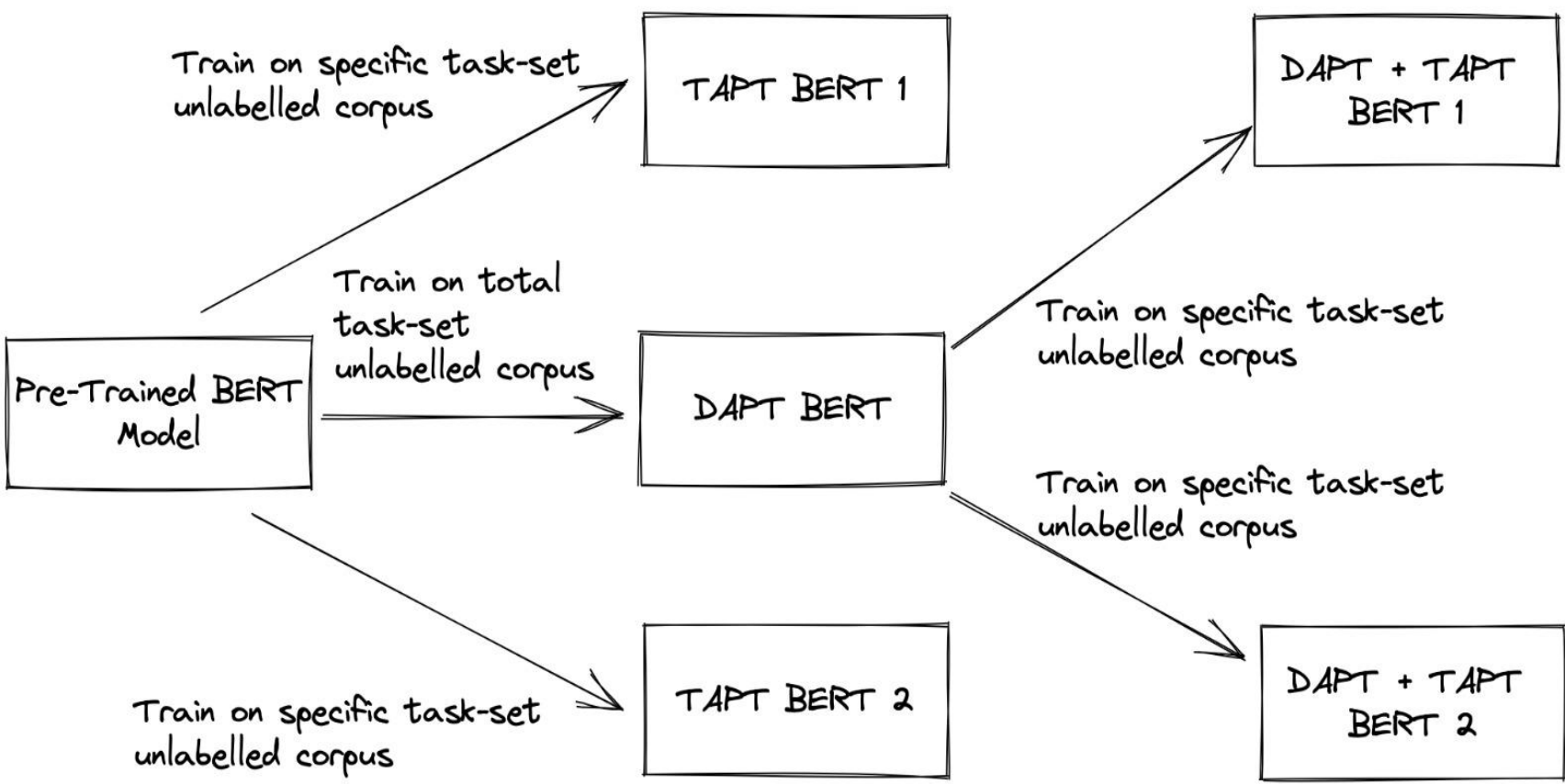
Baseline Methods

For our baseline models, we used a Logistic Regression, a Linear SVM and a RBF SVM.

Embeddings Method

To compare, we implemented two models that used embeddings. The first one had a word2vec, a LSTM, a GRU and an activation layer. And the other one had a GloVe, a CuDNNLSTM and two activation layers.

BERT + Domain Adaptive Pre training

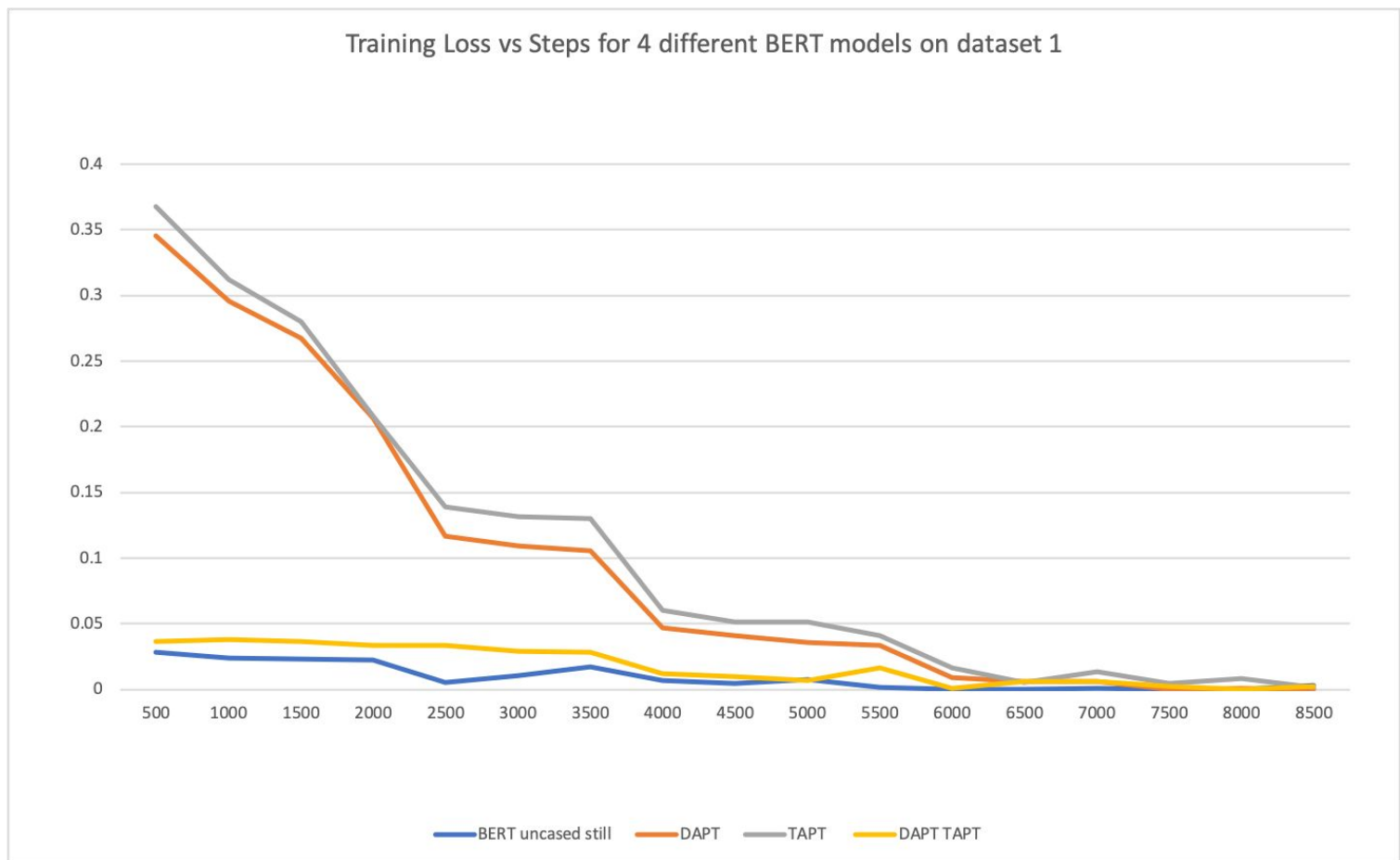
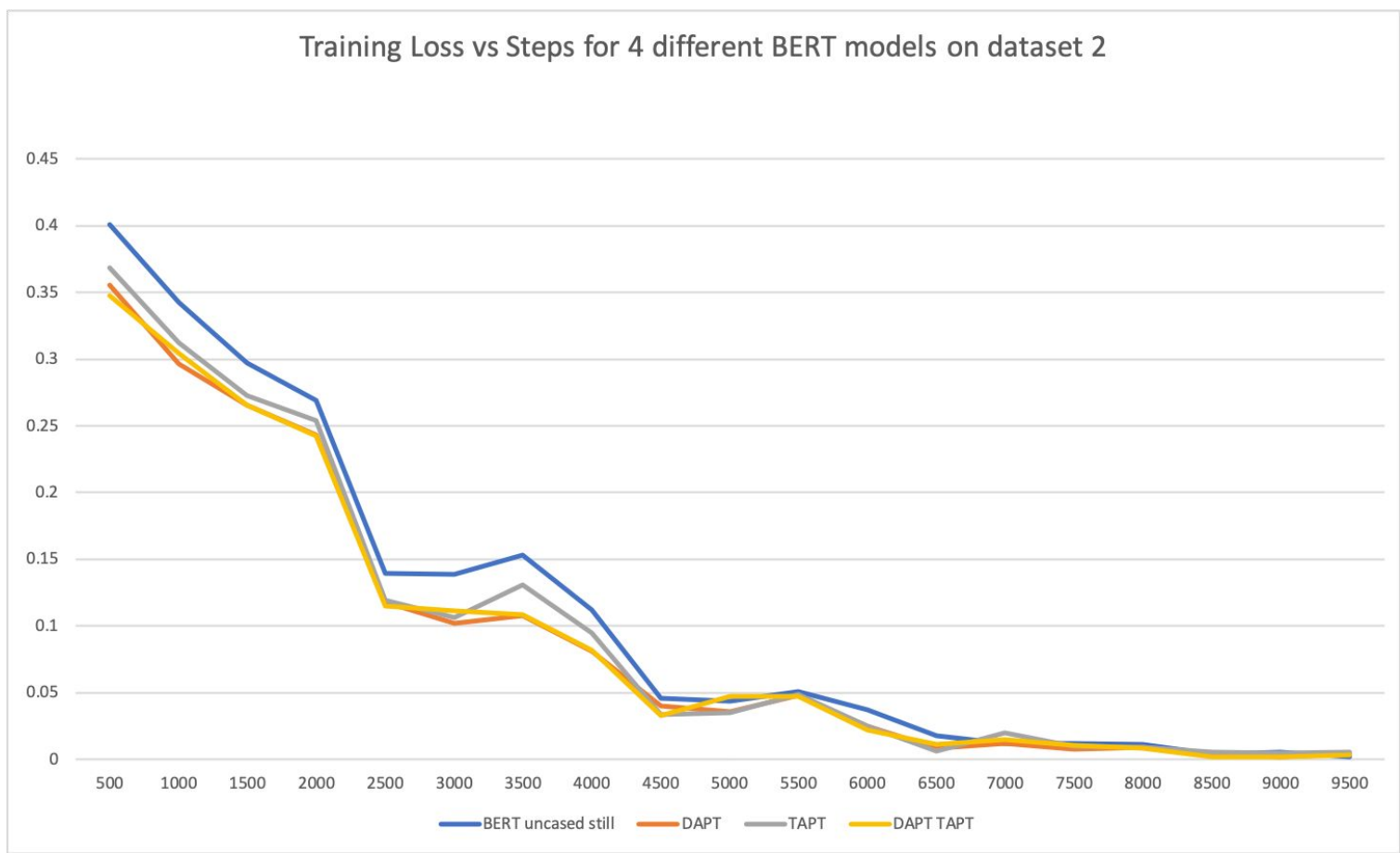


Data

For this project, we found two clean datasets that contained news headlines from the Onion (sarcastic headlines), as well as the Huffington Post (serious headlines).

Experiment and results

For all our models, we trained them and evaluated them separately on the two datasets we obtained. All training and evaluating was run on Google Colaboratory.



	Accuracy
Logistic Regression	0.81
Linear SVM	0.81
RBF SVM	0.79
word2vec	0.79
GloVe	0.86
BERT uncased still	0.99
DAPT	0.981
TAPT	0.929
DAPT TAPT	0.92

	Accuracy
Logistic Regression	0.8
Linear SVM	0.8
RBF SVM	0.79
word2vec	0.79
GloVe	0.86
BERT uncased still	0.918
DAPT	0.932
TAPT	0.931
DAPT TAPT	0.929

Testing Accuracy of each models on datasets 1 (left) and 2 (right)

Related Works

Some of the existing models that we came across in our initial analysis explored the sentiment analysis method of predicting sarcasm, by noticing shifts in sentiment within the conversation (context and reply). BERT has been used to predict sarcasm primarily from variations of reply/context datasets. Many existing models utilize Google’s Word2Vec representations or Global Vector representation (GloVe) to construct contextual embeddings to be fed into NNs. Thus we tested both representations within our project.

From the last couple years, some of the models we encountered first pre-train a sentiment model using unlabeled data (based on RNNs or CNNs) for learning contextual embeddings and sentiment-specific feature extraction. We attempted both domain- and task-adaptive pretraining; additionally we investigated the effect of using both adaptation techniques together.

Conclusion and future work

Overall, our we can see an improvement in performance between our baselines (traditional machine learning and word embeddings) and our BERT task and domain adaptive pretrained models.

Indeed, while the best model between baselines and embeddings had an accuracy of 86% on both datasets (GloVe), the worst performing BERT model had an accuracy of 92% on dataset 1 (DAPT TAPT) and 91.8% on dataset 2 (BERT uncased still).

However, it would be interesting to expand our domain to sentences other than news headlines. For a more general approach, we realize that current news is very important. For example, “Elon musk bought twitter for \$44B while I got it for free on the app store” is a clear sarcastic joke based on current events. A potential avenue to explore would be OpenAI’s GPT-3 model, that having been trained on the internet in 2020 could have the contextual knowledge necessary for broader sarcasm detection