PAWS
Paraphrase Adversaries from Word Scrambling

Ishani Bhatt
Daniel Mathew
Mohammed Sauban
Vaibhavi Kardale
Shantanu Purandare

Introduction

Task :

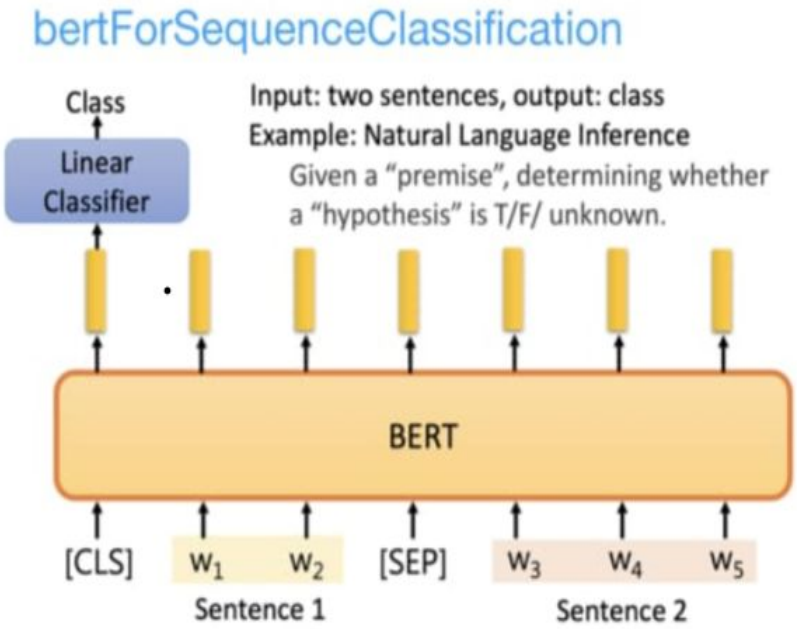
Given a pair of sentences, determine if one sentence is a paraphrase of the other.

PAWS (Paraphrase Adversaries from Word Scrambling) - dataset that has sentence pairs with high word overlap but with equal distribution of negative and positive example pairs.

Dataset sentences are pooled from Quora and Wikipedia.

Progress so far

- Used BERT for training on PAWS dataset due to good results on MRPC and QQP.
- Using BertForSequenceClassification, which has a classification layer on top of the Bert model, classified the sentence pairs as Paraphrases or non-Paraphrases.
- Configurations: ‘bert-base-cased’,’max-length’:128 (due to GPU limitations), pad to max length, learning rate: 2e-5
- Analysed the effect of change in training dataset size on the accuracy of our model and concluded that data augmentation could provide us with better results.



Error Analysis

- Training on 49k PAWS labeled dataset.
- Fairly good accuracy of 91% after 4 epochs.
- Relatively less Recall score of 87% might be due to higher bias in the dataset towards negative examples.
- Model rightly classifies pairs with confidence of 98%. Relatively high confidence of 92% on misclassified pairs.
- As confidence of misclassified pairs is high, there isn't an ideal margin below which we can assert uncertainty.
- Confidence is based of softmax output.

Accuracy	91%
Precision	94%
Recall	87%
F1-score	0.91

Confidence	Rightly Classified	Misclassified
Paraphrase	97.69%	92.43%
Non Paraphrase	98.74%	90.99%
Both	98.30%	91.83%

Future Work

- 2 ways to improve the accuracy of the model.
  1. Data augmentation
  2. Vary the architecture
- Data augmentation :
  1. Construct a sequence to sequence LSTM trained on existing dataset PAWS to generate new paraphrase sentences.
  2. Perform clustering on the sentence embeddings from a pool of sentences. Then intra-cluster connections are paraphrase(1) and inter-cluster connections are not paraphrase (0).
- Vary the architecture:
  1. Here we could use Microsoft MT-DNN since it has shown to outperform BERT. However we need to draw out the implementation to fine tune it to PAWS dataset.