NLP 244 - Assignment 1 Report

First Author
Liren Wu
lwu35@ucsc.edu

1 Introduction

For this assignment, the goal is when given a utterance, tag the sentence with IOB tags and give the relations. We are given two files: hw1_train.xlsx and hw1_test.xlsx. The train file had the following labels: utterances, IOB Slot tags, and Core Relations. The test file only had utterances. The training data has 2251 examples, and the test data has 981 utterances. We will use transformers from the HuggingFace library (Wolf et al., 2020) to build our transformer based models for this assignment. An example of the train and test data can be seen in Table 1 and Table 2. We will use supervised learning to help build our models to tag tokens and classify the relation.

2 Models

In this assignment, I was able to successfully train a BERT model (Devlin et al., 2019) and a DistilBERT model (Sanh et al., 2020). For both models, I followed and based my code on the section week 3 example. I also tried to implement an RoBERTa model, but was unsuccessful. I used the section code as a starter code. Most of the code are the same but with some slight changes. Other than changing the code to accept the assignment data, I had to make it multi-label. The original code was able to do the slot tagging and classifying single labels, but for this assignment we had to do multi-label. I essentially did a one hot encoding for the labels and changed made the model have two different loss functions. I used a CrossEntropyLoss for the slot tagging and a BCEWithLogitsLoss for the intent classifying.

2.1 BERT

I first did some pre-proccesing with the data, the utterances are tokenized by the BERT tokenizer from HuggingFace. The slot labels and intent labels are encoded and all three inputs are put in a dataset and dataloader. Other than than having two different loss functions, I did not change the BERT model from the given started code.

2.2 DistilBERT

Similiary to the BERT model, I first did some pre-proccesing with the data. The utterances are tokenized by the DistilBERT tokenizer from HuggingFace. The slot labels and intent labels are encoded and all three inputs are put in a dataset and dataloader. Along with the two different loss functions, DistilBERT does not having a pooling layer, therefore I had to slice the last hidden layer to make a pooler layer for the intent classifier.

3 Experiments

*NOTE: All scores were calculated from by the evaluation_2.py.

For both models, I did a 75/25 split for training and validation. I did some experiments regarding the amount of epochs and found that after 3 epochs, performance stopped improving much. I also experimented a few different learning rates and the experiments were ran only on the BERT model. The results from the learning rate test can be seen in Table 3. I also compared the BERT model with the DistilBERT model, the results can be seen in Table 4.

4 Results

*NOTE: All scores were calculated from by the evaluation_2.py.

From my testing, the BERT model performed the best and the configuration I used for the model can be seen in Table 5. The BERT model had a very slight performance advantage over the DistilBERT model. This seems logical as DistilBERT is essentially BERT but smaller. DistilBERT has 40% less parameters than the BERT model and runs much faster. But it is still able to preserve relaitivly the same performance, according to the benchmark on HuggingFace, over 95%.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October. Association for Computational Linguistics.

5 Appendix

Utterances	IOB Slot tags	ID
show credits for the godfather	O O O B_movie I_movie	movie.starring.actor
...	...	...
how much did the dark night generate	O O O B_movie I_movie I_movie O	movie.gross_revenue

Table 1: Training data.

Utterances
star of thor
who is in the movie the campaign
...
how many scorsese films were filmed in france

Table 2: Test data.

Learning Size	Sentence Acc.	Slot F1 Score	Intent F1 Score
5e-5	95.39%	85.52%	92.59%
1e-5	95.78%	86.14%	74.42%
5e-4	71.97%	0.00%	13.38%
5e-6	95.03%	83.36%	22.96%

Table 3: Experimenting with learning size

Models	Sentence Acc.	Slot F1 Score	Intent F1 Score
BERT	95.39%	85.52%	92.59%
DistilBERT	94.56%	82.20%	92.13%

Table 4: Model Comparison.

Settings	Value
Batch size	1
Epochs	3
Learning Rate	5e-5
Dropout	0.2
Optimizer	AdamW
Loss function	CrossEntropyLoss and BCEWithLogitsLoss

Table 5: Best Configuration.