Lab 10: Sequence Labeling and Text Classification Part – II: Sequence Neural Network with Enhanced Model IAS Using BERT

Adnan Irshad
University of Trento, Italy
adnan.irshad@studenti.unitn.it

1. Introduction:
This lab was focused on sequence labeling (specifically, Slot Filling) and text classification (Intent Classification) using Recurrent Neural Network. The Model IAS is a baseline architecture that has been modified to improve its performance. by taking BERT (Bidirectional Encoder Representations from Transformers) as base model and fine tune it for the sequence labbeling and slot filling tasks on ATIS dataset by following paper: BERT for Joint Intent Classification and Slot Filling

2. Analysis IAS Model:
The Model IAS is a sequential neural network architecture used for intent classification and slot filling tasks. It consists of an utterance encoder, slot output layer, and intent output layer. In the enhanced model, the LSTM component has been replaced with a BERT model to leverage its powerful language understanding capabilities.

3. BERT Model:
Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based model that has achieved state-of-the-art performance in various natural language processing tasks. It utilizes a bidirectional architecture and pre-training on large amounts of unlabeled text data to capture deep contextual representations.

4. Enhanced ModelIAS Class:
The ModelIAS class has been modified to incorporate the BERT model instead of the LSTM component. The changes include:

- Initializing a BERT model for language understanding.
- Incorporating the BERT model in the forward pass to process the utterance and generate slot and intent predictions.
- Adjusting the input size and dimensions accordingly to match the BERT model requirements.

5. Training Setup and Results:
- The training setup involves defining hyperparameters such as learning rate, batch size, optimizer, and loss function.
- The model is trained on the ATIS dataset for a specific number of epochs.
- I performed different experiments and testing to analyze the output of the model. The results from best three experiments are presented, including average training loss, intent accuracy, and slot F1-score.

- Experiment #9 (Best result with AdamW)
Training Parameters:
Train Batch Size: 128, Eval Batch Size: 64, no. of epochs: 40.0
Learning Rate: 0.0001, Adam Epsilon: 1e-08, Warmup Steps: 0, Max Grad Norm: 1.0
Optimizer: AdamW, Loss Function: Cross Entropy Loss

Average Train loss: 0.2671438528584655
Average Intent Accuracy: 0.9761104889884286
Average Slot F1-Score: 0.9542520301192089

- Experiment #7 (Best Result with Adam)
Training Parameters:
Train Batch Size: 128, Eval Batch Size: 64, no. of epochs: 40.0
Learning Rate: 0.0001, Adam Epsilon: 1e-08, Warmup Steps: 0, Max Grad Norm: 1.0
Optimizer: Adam, Loss Function: Cross Entropy Loss

Average Train loss: 0.6333494229665534
Average Intent Accuracy: 0.973124300111982
Average Slot F1-Score: 0.9029772329246935

6. Key Takeaway:
The enhanced Model IAS, incorporating a BERT model instead of an LSTM, showcases the potential of leveraging powerful language understanding capabilities for intent classification and slot filling tasks. However, further experiments and analyses can be conducted to fine-tune hyperparameters and explore the full potential of the enhanced model.

Note: For a more detailed understanding and implementation of the enhanced Model IAS using BERT, please refer to the provided code snippets and the accompanying research paper linked in the lab description.

7. References
https://arxiv.org/abs/1902.10909