CSCI 6364 – Machine Learning

Spring 2022

Final Project Proposal

Jesse Robles-Alcaraz

# Comparison of Intent and Named Entity Classifiers

## Background

User expectations regarding their interactions with search systems and dialog systems (e.g., chatbots, Siri, Cortana) have significantly increased in the last decades beyond simple keyword matching and rule-based responses. A main driver for this has been large corporations such as Google, Amazon, and Microsoft’s significant investment in research and development of machine learning algorithms and hardware. This has led to the development of embedding techniques and language models, followed by the development of the Transformer architecture which further pushed the state of the art in natural language understanding. Notable among these is the Bidirectional Encoder Representations from Transformers (BERT) model. These models are trained on general corpora of text data (e.g., new articles, Wikipedia) and can be fine-tuned to specific use cases to achieve better performance.

Some recent research indicates that for specific domains where the language style differs from the data used to train BERT, training a transformer architecture from scratch can approach the performance of these larger architectures, some instances outperforming them. This project is based on one of the papers (Bunk, et al. 2020) reviewed, which showed that training a custom model, named the Dual Intent and Entity Transformer (DIET) from scratch outperforms BERT with fine tuning on intent classification and entity recognition. Unfortunately, the authors of these papers do not appear to provide access to their training data. Thus, part of this project was to identify datasets that can be used to train and compare these two models. The BERT model used was trained using the Huggingface library, and the DIET model was trained using the Rasa python conversational AI library. The datasets are described in the section and table below.

## Data Sets and Data Preprocessing

The datasets were sourced from the GitHub repository at <https://github.com/jianguoz/Few-Shot-Intent-Detection> and from Kaggle at <https://www.kaggle.com/datasets/joydeb28/nlp-benchmarking-data-for-intent-and-entity>. These datasets are somewhat similar in that they are intended to be used in a conversational assistant or search setting.

Table : Descriptions of Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset Name** | **Has Entities** | **Original Format** | **Description** |
| ATIS | Y | CoNLL | Airline travel information systems. Contains intents and entities. Examples: "Find cheapest flights from BWI to MCO" |
| BANKING77 | N | Sample/Labels | Intents in the banking domain. No entities. Examples: "Still waiting on my new card" |
| BENCHMARKING\_DATA | Y | JSON | Kaggle dataset for benchmarking intents and entities. Examples: "Book restaurant", "Add song to playlist" |
| CLINC150 | N | Sample/Labels | A dataset with 10 domains and 150 intents. No entities. Style is similar to personal assistant. Examples "how do you say dog in spanish" |
| HWU64 | N | Sample/Labels | Personal assistant data with 64 intents and several domains. No entities. Examples: "remind me about the meeting" |
| SNIPS | Y | CoNLL | Personal assistant data with 7 intents. Includes entities. Examples: rate this current novel 1 stars" |

Most of these datasets were in CoNLL format, and one was in JSON format. The first challenge was to preprocess the datasets so that each model type can accept the data. It made more sense to handle a single format, so the JSON parser processes the JSON files and outputs in the same CoNLL format as the other datasets. Additionally, all but the benchmarking\_data dataset were split into train, test, validate datasets. That dataset was split into the same format as it was processed into CoNLL format.

The Huggingface library data input format for Named Entity Recognition is like the CoNLL format in that it takes in tokenized text and the NER label for each token. The main preprocessing step is to encode the CoNLL format into numeric values. The input format for the Rasa DIET model is a YAML file that contains examples grouped by intent and with entities annotated in the text example. The CoNLL parser in the code base handles preprocessing the input files and outputting the proper format for each model type. For DIET, new YAML files are created.

## Model Training and Model Comparison

Since the Rasa DIET model performs both query intent classification and named entity recognition, two BERT models were trained on the datasets that contained both intents and entities. In order to reduce model size and training time, the distilbert-base-uncased model was used as the basis for fine tuning, which is a smaller and faster version of BERT. The BERT models were trained for 5 epochs with a learning rate of 2e-5 and weight decay of 0.01, which is a form of regularization. These parameters were selected based on the Huggingface documentation. For the DIET models, the model was trained on 100 epochs with the other parameters set with default values. In terms of training time, the DIET models take longer because of the larger number of epochs. These epochs are necessary because the models are trained from scratch.

The model sizes are also notable. The average size of the DIET models is approximately 86MB, while the BERT models

## Comparison of Results

Below are two charts showing the precision, recall, F1 score and accuracy of each model in intent classification and named entity recognition. The seqeval python library was used on the NER evaluation and the sklearn classification report was used on the query intent evaluation. On these datasets, the models were generally comparable to each other. For query intent classification, the BERT model slightly outperformed the DIET model on 5 of the 6 datasets on all metrics but F1 score.

## Conclusions and Future Work

Below is a list of papers and other references that I have reviewed thus far:

# References

Bunk, Tanja, Daksh Varshneya, Vladimir Vlasov, and Alan Nichol. 2020. "DIET: Lightweight Language Understanding for Dialogue Systems."

Guo, Weiwei, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, et al. 2020. "DeText: A Deep Text Ranking Framework with BERT." *CoRR* abs/2008.02460. https://arxiv.org/abs/2008.02460.

Larson, Stefan, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, et al. 2019. "An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction."

Tejaswini Mallavarapu, Ying Xie, and Simon Hughes. 2022. "CatBERT: An Incrementally Trained Language Representation Model for E-Commerce Applications." *Proceedings of the International Workshop on Interactive and Scalable Information Retrieval methods for eCommerce (ISIR-eCom).*