Comparative analysis of Intent recognition models

Shreya S Adiga
PES1UG20CS401
Department of CSE, PES University
Bangalore, India
shreyasharada.adiga@gmail.com

Shubangi Saxena
PES1UG20CS418
Department of CSE, PES University
Bangalore, India
shubangisaxena12@gmail.com

Srushti S Hampannavar
PES1UG20CS437
Department of CSE, PES University
Bangalore, India
srushtihampannavar@gmail.com

Abstract—**Intent Recognition (IR) is considered a key area in Natural Language Processing (NLP). Interpreting the context of text searched by user improves the response time and helps the search engine to give appropriate outputs. It has crucial usage in various applications. For comparison purposes, we have applied primarily used Machine Learning techniques, namely Naive Bayes, Support Vector Machine, and Random Forest, as well as Deep Learning Techniques used for intent recognition like Bidirectional Long Short Term Memory Network on the same dataset and evaluated the accuracy. It is found out that BERT accuracy is far better than that of the other models implemented.**

Keywords— BERT, Intent recognition, deep learning, deep neural network, natural language processing, intent detection

I. Introduction

Intent recognition is finding out context corresponding to the user's text. It is a classification task in which the user's text is classified based on what he/she wants to achieve using predefined intents. Human language contains several constructs which are very complicated to handle, which makes this task of intent recognition a highly complex problem.

The key contributions of this paper are the following:

- A comparative analysis of different models such as BERT(Bidirectional Encoder Representations from Transformers), LSTM, SVM, naive bayes, random forest.
- The proposed models attained state-of-the-art results for the task of user's spoken utterance classification.

II. Related Work

This section summarizes the work that has been done using the dataset for various purposes. The assumptions made, a summary of the approach used and the results obtained have been mentioned. Furthermore any lacuna inferred from the following approaches along with any limitations have also been appropriately mentioned.

Jacob et.al [1] mainly focus on a new language representation model called BERT. It includes sentence-level tasks which aim to predict the relationships between sentences and token-level tasks where models are required to produce fine-grained output at the token level.

Tulika Saha et.al [3] have used BERT model to classify tweet acts. The model is based on calculating attention weights over the representations of tokens of a sequence to identify tweet acts. The main objective of this paper was to assign the most appropriate tweet act (say y) among a set of tags (Y = {y1, y2, ...yi} where i is the number of tweet acts) which is a multi-class classification problem.

III. Analysis Of The Dataset

For our intent recognition model, we have used the Snips dataset, which was collected through crowdsourcing for the Snips personal voice assistant. There are 7 unique intent classes for the training set, on a variety of topics including playing music, restaurant reservations, and getting the weather (e.g. 'Book an Italian place with a parking for my grandfather and I' and 'Which movie theater is playing The Good Will Hunting nearby?'). The training set contains 13,084 utterances, and separate validation and test sets that contain 700 utterances each.

Table 1| Dataset Detail

Dataset	"SNIPS-NLU"
Vocab Size	30,522
Maximum position embeddings	512
Intents	7
Training samples	13,084
Validation samples	700
Test Samples	700

A. Exploratory Data Analysis and Data Processing

The dataset has 3 columns in total and all of them were of type char. On performing basic data preprocessing, no missing values or duplicate values we found in the dataset.

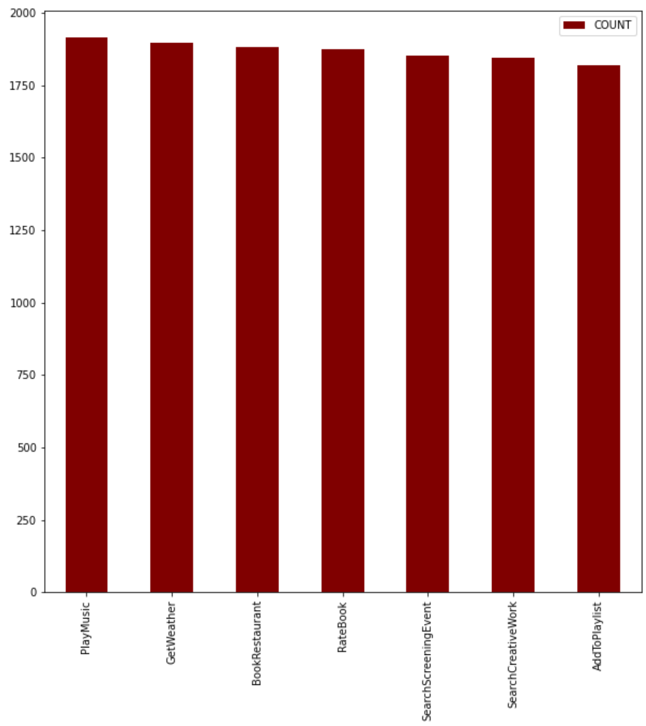


Fig 1. Number of examples per intent of train data

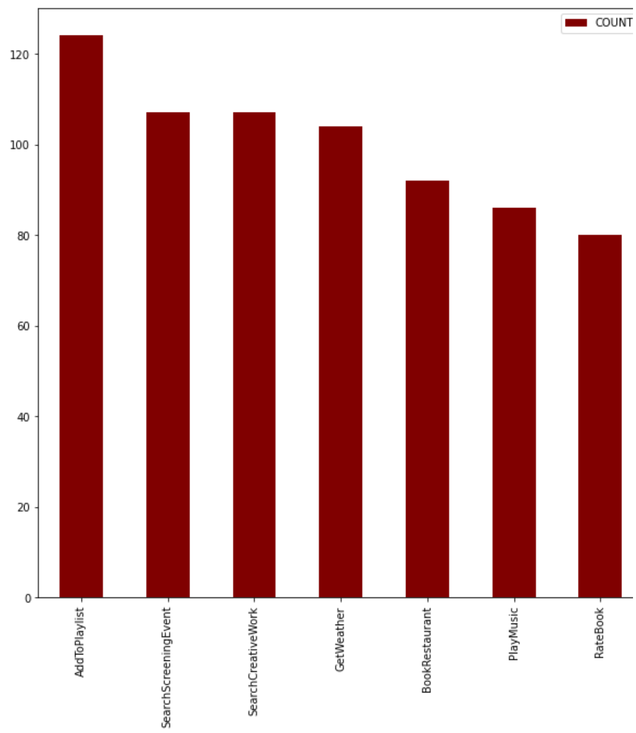


Fig.2. Number of examples per intent of test data

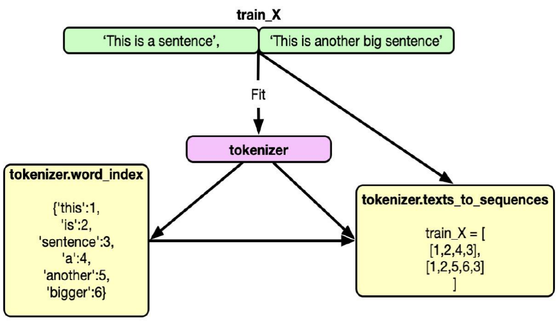
IV. Proposed Methodology

The dataset contains various user spoken utterances. The dataset has already been divided into testing and training sections. The problem at hand is to determine the intent behind the user's utterance on the given data.

In order to predict the intent, we intend to use a number of different models, the explanation of the models and the experimental results of the same have been recorded.

A. *Support Vector Machine*

We have used support vector machine as there is a clear margin of separation between classes.To separate the classes, there are many hyperplanes that can be chosen by the SVM. It finalizes on a plane which is at a maximum distance from the data points of either of the categories/classes. Maximizing the marginal distance gives less scope for misclassification so that when test data points arrive, classification error is the least.



The dataset is split into test and train datasets. They are cleaned and predictors are separated from the outcome variables. The train and test have 700 and 13084 samples respectively.

B. *Naive Bayes*

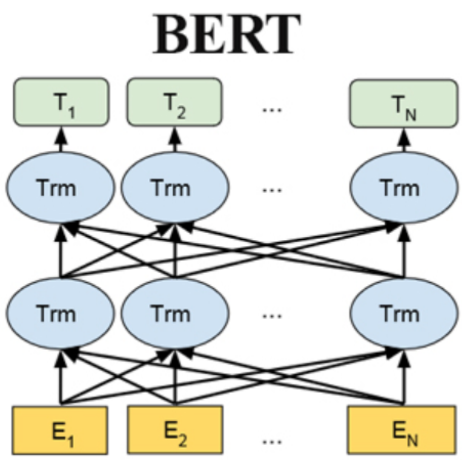
In our case we have naive bayes for multiclass classification.The data is classified by maximizing $P(O|C_i)P(C_i)$ using Bayes theorem of posterior probability (where, O is the object or tuple in a dataset and i is an index of the class.)

C. *Random Forest*

Random first is based on bagging algorithm and uses ensemble learning technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it reduces overfitting problem in decision trees and also reduces the variance and therefore improves the accuracy.

D. *BERT*

BERT is a transformer based model that has been pre-trained on an enormous amount of English data.It learns to represent language efficiently, and can then be fine-tuned for a specific task.
BERT outputs a matrix tensor of shape batch size by number of intents, which consists of the intent probabilities for each utterance in the batch. BERT is a multi-layer bidirectional Transformer encoder. There are two models introduced in the paper. BERT denote the number of layers(i.e., Transformer blocks) as L, the hidden size as H, and the number of self-attention heads as A.



E. *BiLSTM*

BiLSTM layer learns bi-directional long-term dependencies between time steps of time series or sequence data. It is the process of making any neural network o have the sequence information in both directions backwards(future to present) or forward(past to future).

V. Experiment Results

V.I. Metrics Used

Accuracy is the percentage of times we are predicting correctly.

Accuracy = (TP + TN)/(TP + TN + FP + FN).

Precision(P) is the percentage of correct predictions out of total positive predictions.

Precision = TP/(TP + FP).

Recall(R) is the percentage of correct predictions out of total actual positive predictions.

Recall = TP/(TP + FN).

F1 Score (F1) is the measure of the balance between precision and recall.

F1 Score = 2 ∗ (Precision ∗ Recall)/(Precision + Recall)
where, TP = True Positive

TN = True Negative

A. *Support Vector Machine*

We obtained an accuracy score of 96.14%. Best C is 1.0, gamma is auto and the kernel is linear.

B. *Naive Bayes*

We obtained an accuracy score of 96%.

C. *Random Forest*

We obtained an accuracy score of 93%.

D. *BERT*

We obtained an accuracy score of 95.86%
Summary of the model

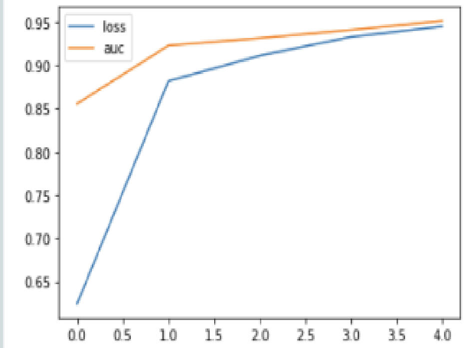
```
Layer (type:depth-idx)                        Param #
=================================================================
├─BertModel: 1-1                              --
|    └─BertEmbeddings: 2-1                    --
|    |    └─Embedding: 3-1                    23,440,896
|    |    └─Embedding: 3-2                    393,216
|    |    └─Embedding: 3-3                    1,536
|    |    └─LayerNorm: 3-4                    1,536
|    |    └─Dropout: 3-5                      --
|    └─BertEncoder: 2-2                       --
|    |    └─ModuleList: 3-6                   85,054,464
|    └─BertPooler: 2-3                        --
|    |    └─Linear: 3-7                       590,592
|    |    └─Tanh: 3-8                         --
├─Dropout: 1-2                                --
├─Linear: 1-3                                 5,383
=================================================================
Total params: 109,487,623
Trainable params: 109,487,623
Non-trainable params: 0
=================================================================
```

E. *BiLSTM*

We obtained an accuracy score of 95.6%
Summary of the model

```
Model: "sequential_9"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_9 (Embedding)     (None, 35, 128)           1415552

 bidirectional_11 (Bidirecti  (None, 35, 256)          263168
 onal)

 bidirectional_12 (Bidirecti  (None, 128)              164352
 onal)

 dense_17 (Dense)            (None, 32)                4128

 dropout_11 (Dropout)        (None, 32)                0

 dense_18 (Dense)            (None, 7)                 231

=================================================================
Total params: 1,847,431
Trainable params: 431,879
Non-trainable params: 1,415,552
_________________________________________________________________
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f7f73012690>



VI. Conclusions and Future Work

After comparing all the models that we tried for intent detection, we observed that although most of the models give us very high accuracy, using a BERT model to accomplish this task gave us the best results and consistently performed well on the cross-validation and test datasets.

In future we plan to use BERT model and add more functionalities so that we will be able to classify open intents(intents that belong to two or more classes or that cannot be classified at all).

VII. Contributions Of Team Members

Each of us was involved in the data processing and exploratory data analysis section. Further for stage 2,each of us worked on a model for classification of the activities, the exact details of which are given below -
Shreya Adiga - BERT
Shubangi - Bi-LSTM
Srushti - SVM, Random forest, Naive bayes
All of the team members were involved in the writing of the report and creation of the video and all decisions regarding the project were taken on call. All the code has been documented and all the references used have been mentioned appropriately

VIII. References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding .

[2] Vasima Khan, Tariq Azfar Meenai, “Pretrained Natural Language processing model for Intent Recognition (BERT_IR).

[3] Tulika Saha, Aditya Prakash Patra, Sriparna Saha, Pushpak Bhattacharyya, “A transformer based approach for identification of tweet acts”.

[4] Sreelakshmi K, Rafeeque PC, Sreetha S , Gayathri, “ Deep Bi-Directional LSTM Network for query intent detection”.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion

[6] Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, “Attention is all you need.”