Project Report: Natural Language Understanding (NLU) System for Intent Classification

1. Introduction

In this project, we developed a Natural Language Understanding (NLU) system for intent classification. The system is designed to classify intents from text inputs and provide confidence scores based on the classification results. Additionally, a fallback mechanism is implemented to handle cases where the predicted intent does not meet a predefined confidence threshold.

2. Features of the System

- Classify intents based on given text inputs.
- Provide confidence scores for the predicted intents.
- Implement a fallback response if the predicted intent does not meet the predefined confidence threshold.

3. Data Gathering & Understanding

We gathered data containing examples of different intents and their corresponding labels. The data was preprocessed to extract intents and examples, and then stored in a structured format. Exploratory data analysis techniques were employed to understand the distribution of intents in the dataset.

4. Data Pre-processing

Text pre-processing techniques such as lowercasing, punctuation removal, tokenization, stop word removal, and lemmatization were applied to clean the text data. Spelling correction was also performed using an autocorrect library. The pre-processing steps were aimed at standardizing the text data and improving the quality of features for model training.

5. Data Visualization

Word clouds and frequency distribution plots were generated to visualize the most common words and their frequencies in the dataset. These visualizations helped in understanding the key themes and patterns present in the text data.

6. Model Building

Two different models were built for intent classification: a custom LSTM-based model and a BERT-based model. The custom model architecture consisted of embedding,

bidirectional LSTM, and dense layers. The model was trained using the intent-labelled data and optimized using various callbacks such as model checkpointing and early stopping. The BERT-based model utilized the pre-trained BERT architecture for sequence classification and fine-tuned it on the intent classification task.

7. Model Optimization with Pre-trained Models

The BERT-based model was initialized with pre-trained weights and fine-tuned on the intent classification task. The model was trained using the encoded input text sequences and evaluated for accuracy and loss metrics.

8. Model Testing

A function was implemented to classify intents from new text inputs using the trained model. The function preprocesses the input text, tokenizes it using the BERT tokenizer, and makes predictions using the fine-tuned BERT-based model.

9. Fallback Mechanism

A fallback mechanism was implemented to handle cases where the confidence score for the predicted intent is below a predefined threshold. If the confidence score is below the threshold, the system returns a fallback response indicating that the intent could not be confidently determined.

10. Conclusion

In conclusion, the developed NLU system effectively classifies intents from text inputs and provides confidence scores for the predictions. The system's performance was evaluated using various evaluation metrics and was found to achieve satisfactory results. Additionally, the implemented fallback mechanism enhances the system's robustness by handling cases of uncertain intent classification. Overall, the project demonstrates the effective application of NLU techniques for intent classification tasks.