1. COMP 5600 Project Proposal Report

Laura Wilson ([lgw0020@auburn.edu](mailto:lgw0020@auburn.edu))

Coordinator: Daniel Harrison([dah0052@auburn.edu](mailto:dah0052@auburn.edu))

Taylor Cross([tac0062@auburn.edu](mailto:tac0062@auburn.edu))

Dataset: <https://huggingface.co/datasets/renumics/speech_commands_enriched>

Title: Speech Command Classification Using Machine Learning Techniques

**Introduction**

The advancement in speech recognition technology has led to a growing interest in developing more sophisticated and responsive systems capable of understanding and executing spoken commands. This report outlines a project to create a machine-learning model capable of classifying audio files into different speech commands. The primary application domain for this project includes robotics and large language models, where the system could interpret complex tasks through voice commands, focusing on keyword classification.

**Dataset Overview**

The dataset for this project consists of audio speech commands designed for training speech recognition systems. It includes a variety of words, as well as audio files containing silence and static, to aid in distinguishing words from non-speech sounds. This dataset is pre-existing and will be utilized to train our model.

**Machine Learning Approach**

To address the classification task, we will employ multi-layer perceptrons (MLPs), a class of feedforward artificial neural networks. MLPs are well-suited for pattern recognition tasks such as this, where the input is an audio file and the output is a specific speech command.

**Challenges**

The primary challenges anticipated in this project include:

The complexity and computational demands of processing a large dataset of audio files.

Adapting the dataset for use with an MLP may require significant preprocessing to ensure effective training.

Technology Stack

**Our project will leverage the following technologies:**

TensorFlow or PyTorch: For building and training the neural network model.

Librosa: For audio processing tasks, such as feature extraction.

Renumics Spotlight: For data exploration and understanding dataset characteristics.

**Hyper-parameters and Fine-tuning**

Key hyper-parameters for our MLP include layer sizes, learning rates, and activation functions. We will fine-tune these parameters using techniques such as grid search and cross-validation, with a focus on optimization algorithms and data splits for validation.

Model Evaluation

The effectiveness of our classifier will be demonstrated through its ability to accurately detect spoken words from the audio inputs. We will compare the model's predictions against the actual labels in the dataset to assess accuracy. We hope to get an accuracy of up to 80%.

Project Timeline and Milestones

March 29th - 31st: Setup of the development environment and installation of necessary libraries and tools.

March 31st - April 2nd: Exploration and preprocessing of the SpeechCommands dataset, including feature extraction.

April 2nd - 9th: Initial model development and baseline training.

April 9th - 11th: Experimentation with basic hyperparameters.

April 11th - 16th: Evaluation and refinement of the model, including progress reporting.

April 16th - 18th: Advanced hyperparameter tuning.

April 18th - 20th: Extended training and in-depth evaluation.

April 20th - 22nd: Incorporation of dataset enrichments and further experimentation.

April 25th - 26th: Final documentation and reporting, project review, and submission.

Conclusion

This project aims to push the boundaries of speech recognition technology by developing a robust classifier for speech commands. By leveraging advanced machine learning techniques and a comprehensive dataset, we hope to create a model that significantly enhances human-computer interaction, particularly in robotics and large language models. The anticipated challenges, particularly in dataset complexity and computational requirements, will require a thoughtful approach to model development and refinement. Still, the potential applications of a successful project are vast and compelling.