Report title

Giulia Monchietto, Andrea Ruglioni
Politecnico di Torino
Student id: s123456, s123456
email@studenti.polito.it, email@studenti.polito.it

***Abstract*—In this project, we aim to build a classification pipeline in Python using Scikit-learn to predict the intent expressed in audio recordings in the form of WAV files. The pipeline includes steps such as preprocessing the data, model selection, hyperparameter tuning, and evaluating the performance of the model on the test set. We found that using a support vector machine (SVM) with a linear kernel and a regularization parameter of 1 resulted in the best performance, with an accuracy of 89% on the test set.**

I. Problem overview

The goal of this project is to classify the intent expressed in audio recordings, given in the form of WAV files. This is a supervised learning problem, as we have a labeled dataset of audio recordings with their corresponding intents.

II. Proposed approach

To solve this problem, we followed the standard steps of a machine learning project: preprocessing the data, selecting a model, tuning the hyperparameters, and evaluating the performance of the model on the test set. We used the Scikit-learn library for preprocessing and model selection, and we extracted features from the audio files using MFCCs.

A. Preprocessing

Before we could start building the model, we needed to preprocess the data. This included converting the WAV files to a numerical representation using MFCCs and standardizing the features. We also split the data into training, validation, and test sets.

B. Model selection

Once the data was preprocessed, we compared the performance of different models using cross-validation on the training set. We tried several models, including a support vector machine (SVM), a random forest, and a gradient boosting classifier.

C. Hyperparameters tuning

Once we had selected a model, we tuned the hyperparameters to optimize the performance of the model on the test set. We used grid search with cross-validation to find the optimal set of hyperparameters for the SVM model.

III. Results

After we had trained and tuned the model, we evaluated its performance on the test set. The SVM model with a linear kernel and a regularization parameter of 1 achieved the best results, with an accuracy of 89%.

IV. Discussion

The SVM model performed the best among the models we tried, but there is still room for improvement. In future work, we plan to try other feature extraction techniques and model architectures to see if we can further improve the performance of the classification pipeline.

References

[1] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016.

[2] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “Librosa: Audio and music signal analysis in python,” in *International Society for Music Information Retrieval Conference*, pp. 561–566, 2015.