Language Identification

Team Information

Person.1

Ambati. Padma Sri

Padmasriambati2004@gmail.com

<https://github.com/PADMASRIAMBATI>

Person.2

Maddula. Rushika Sritha

[rushikasrithamaddula2005@gmail.com](mailto:rushikasrithamaddula2005@gmail.com)

<https://github.com/Rushikasrithamaddula11>

Person.3

Dulam. Bhavya Tulasi

dulambavyatulasi@gmail.com

<https://github.com/bavya1309>

**Problem Statement**

*Language identification*

*Description*: Identify the language of the speech Dataset.

Language identification is the process of automatically

detecting the language of a given text. It involves analyzing

various linguistic features, such as character sequences, word

frequency distributions, and grammatical patterns, to determine

the most probable language. This task is crucial for numerous

applications, including text processing, machine translation,

and content filtering. Common techniques for language

identification include n-gram models, statistical analysis, and

machine learning algorithms such as Naive Bayes and neural

networks. By accurately identifying the language of a

document, systems can provide better user experiences,

enable multilingual support, and facilitate efficient information

retrieval.

**Dataset Details**

The dataset contains speech samples of English, German,

Spanish and French languages. Samples are equally balanced

between languages, genders and speakers. The ready to use

dataset can be downloaded from Kaggle.

**DataSet Link**

https://www.kaggle.com/datasets/toponowicz/spoken-language-identification

**Method or Experimental Setup** –

*Data Preprocessing*: - Audio files are loaded and converted into Mel Frequency Cepstral Coefficients (MFCCs), which serve as features for the model.

- MFCCs are extracted using the librosa library.

- Model Architecture: - A Convolutional Neural Network (CNN) model is employed for

language classification.

- The model architecture consists of two convolutional layers followed

by max-pooling layers, flattening, and fully connected layers.

- ReLU activation functions are used in the convolutional layers to introduce non-linearity.

- Batch normalization is applied to stabilize and accelerate the training process.

- Dropout regularization is incorporated to mitigate overfitting.

- Training Configuration:

- The model is compiled with the Adam optimizer, which is known for its robustness and efficiency.

- Sparse categorical cross-entropy loss function is chosen as it is

suitable for multi-class classification tasks.

- The training process utilizes early stopping to prevent overfitting and improve generalization.

- Hyperparameters such as learning rate are optimized using

techniques like GridSearchCV.

- Model Evaluation: -

The model is evaluated on a separate testing dataset to assess its performance in language classification.

- Evaluation metrics include accuracy, precision, recall, and F1-score.

- Confusion matrix analysis provides insights into the model's behavior across different language classes.

Results and Observations✨

|  |  |
| --- | --- |
| **Model Type** | **Accuracy** |
| CNN | 0.916 |
| CNN (Grid Search) | 0.912 |

**Observations**

1.Both the basic CNN model and the CNN model optimized with

grid search perform well in classifying the languages in the audio

files.

2. The accuracy achieved by both models is above 90%,

indicating strong performance in language classification.

3. The slight difference in accuracy between the basic CNN and

the optimized CNN through grid search is negligible and could be

attributed to random variations in the dataset split or model

initialization.

4. The CNN architecture effectively captures the spatial

dependencies in the MFCCs features extracted from the audio

files, enabling accurate language classification.

5. The model's performance could further improve with additional data augmentation techniques or fine-tuning of hyperparameters.

**Conclusion**

In the language identification project conducted by our team at

IIIT Hyderabad, utilizing Natural Language Processing

techniques, we developed a Convolutional Neural Network

(CNN) model to classify languages in audio files. With a team

of three members, we meticulously crafted the model

architecture, optimized hyperparameters through grid search,

and trained the model using TensorFlow and Keras libraries.

The CNN model achieved impressive accuracy exceeding 90%,

demonstrating robust language classification capabilities. This

project underscores the efficacy of CNNs in processing audio

data for language identification tasks and highlights the

collaborative efforts of our team in delivering successful

outcomes in NLP-based projects.