SPOKEN LANGUAGE CLASSIFICATION

Abhishek Agarwal Raghav Sood Surabhi S. Nath
IIIT DELHI

Motivation and Problem Statement

We aim to perform **Language Identification in audio samples** and classify them into multiple languages spoken in India. Most of us are well versed with more than one language and our brain is able to identify and respond in the appropriate language in conversations. Can this learning be transferred to a machine? Further, we will also attempt to detect multiple languages in the same audio sample using segmentation.

Technology developed to date can successfully convert speech to text only if the language is given. Our attempt for automatic language identification can ease this process and prevent the user from explicitly specifying the choice of language.

Data Acquisition Effort

We will use publicly available datasets for our models:

OpenLRS dataset for US Accented Spoken English language, which contains 1000 samples (.flac format) with an average length of 15s,

TopCoder Spoken Languages Dataset for Hindi Language, which contains 150 samples (.mp3 format) with an average length of 10s.

We will use web scraping on **Youtube** and **All India Radio** for other regional languages and split the samples into smaller audio files of 10-15s each to keep the data consistent across both languages.

Preprocessing Techniques

- Data Filtering
- Down Sampling
- Audio compression
- Fourier Transformation
- RMS value normalization
- Data Equalization
- MFCC

Strategy for Model Selection

Cross Validation Techniques:

- K Fold Cross Validation, LOOCV
- Bootstrapping
- Monte Carlo / Random Sampling

Tuning Hyperparameters:

- Grid Search CV

Evaluation Metrics

Confusion Matrix - True Positives, False Positives, etc. will be used to calculate precision, recall, accuracy, and F1 score, Accuracy, ROC

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$F\ Score = \frac{1}{\frac{\alpha}{Precision} + \frac{1-\alpha}{Recall}} \qquad 0 \le \alpha \le 1$$

Learning Techniques

Baseline Technique: Support Vector Machine - One one the most diverse classifiers hence a baseline for our problem.

Advanced Techniques:

I. Gaussian Mixture Model - Suitable since signals, including audio, are mostly Gaussian in nature (Probabilistic)

II. Hidden Markov Model - Suitable for time series data and is often used for speech recognition (Probabilistic)

III. Neural Network - Suitable due to its feature learning capabilities and discriminative training (Backpropogation)

IV. Clustering methods - Unsupervised learning techniques (eg. K Means) are suitable to see the natural tendency of how the data gets grouped (true labels are ignored) (Euclidean Distance)

Timeline

Proposal Deadline, Data Collection, Pre- processing **25th September**	Feature Engineering **10th October**	Model Training, English vs Hindi Classification **25th October**	Advanced techniques, Multiclass Classification **10th November**	Analysis, Evaluation, Final Report **25th November**

Individual Contribution

	Data	Feature Extraction	Model Training	Validation	Testing and Evaluation
Abhishek	Data Acquisition	Feature Identification	Support Vector Machines, Gaussian Mixture Model, Neural Networks	K fold Cross Validation, LOOCV	Confusion Matrix, Accuracy
Raghav	Data Acquisition	Feature Extraction	Hidden Markov Model, Clustering, Neural Network	Bootstrapping, Monte Carlo	ROC, AUC, MCC
Surabhi	Data Preprocessing	Feature Selection	Hidden Markov Model, Gaussian Mixture Model, Neural Network	Disjoint set, Three-way Cross Validation	Precision, Recall, F1 Score