INTERNATIONAL JOURNAL OF ELECTRONICS AND COMPUTER APPLICATIONS

RESEARCH ARTICLE

Article access online

OPEN ACCESS

Received: 22.02.2024

Accepted: 10.05.2024

Published: 22.05.2024

Citation: Gore PN, Tukaram Bawadane S, Sanjay Ingale S, Watve SG. (2024). Multilingual Spoken Language Recognition Using Machine Learning Algorithms. International Journal of Electronics and Computer Applications. 1(1): 15-19.

* **Corresponding author**.

pratik_gore@moderncoe.edu.in

Funding: None

Competing Interests: None

ISSN
Print: XXXX-XXXX
Electronic: XXXX-XXXX

Multilingual Spoken Language Recognition Using Machine Learning Algorithms

Pratik Nandkumar Gore[1]*, Subhash Tukaram Bawadane[1], Sahil Sanjay Ingale[1], S G Watve[1]

1 Electronic & Telecommunication, P.E.S Modern College of Engineering, Pune, India

Abstract

Machine learning algorithms are being studied to develop algorithms that can recognize and segment languages in audio recordings. This technology has great potential to improve our ability to communicate and understand different language communities. The main goal of multilingual recognition is to develop models that can accurately recognize spoken language. This is especially useful in applications such as call centers and voice assistants. Speech patterns found in online podcasts, audiobooks and its variants in Speech Corpus. This corpus contains utterances and each takes an equal time of 10 seconds. The entire corpus is divided into two parts, a large object as a training data set and a small one as a test set. Thus, an acoustic model that uses the mean values of the BFCC appears to be an appropriate method for speech recognition. The system uses Convolutional K Nearest Neighbors (KNN) to solve the multiple classification problem. The aim of the project is to know Punjabi, Hindi and Gujarati.

Keywords: Spoken Language; Indian Language; Audio Features; Machine Learning; Speech Recognition

Introduction

Language is an integral part of human communication and plays an important role in our daily communication, be it spoken or written. In our increasingly globalized world, the ability to recognize and understand multiple languages is an indispensable skill, not just for individuals but for many applications in artificial intelligence and technology also. However, one obstacle to this increase in global communication is that many people speak different languages and effectively do not have a common mode of communication. That is, effective communication requires a mutually understandable language on both sides. Language instruction provides a means to deliver this option.

Our main task is to identify spoken language parameters and features that can be used to separate languages. We will use the Bark Frequency Cepstral Coefficient (BFCC) to extract features from the audio file. So far, several methods have been used for speech recognition. Among all the methods, machine learning has the best accuracy. That's why we also used Machine learning in our project for ordinary language.

Robust machine learning algorithms that incorporate advanced audio processing and feature extraction methods are becoming increasingly effective in addressing this challenge.

Parts of a sentence can be used to illustrate different aspects of a language. Because the raw speech signal is complex, it might not be appropriate to use machine learning to feed it to a speech recognition system for spoken speech recognition. The requirement for a strong front end This front end's job is to compactly extract all pertinent acoustic information. Stated differently, pre-processing should remove any extraneous data, like background noise, and categorize the residual (meaningful) data into a more manageable set of attributes that can be fed into the classifier. A feature encompasses a variety of speech signals. These characteristics can be acoustic, prosaic, or vocal.

Using machine learning, spoken language identification (LID) entails creating algorithms that can automatically identify and differentiate between various languages spoken in audio recordings. Applications for this technology include multilingual call centers, voice assistants, and content filtering based on language.Prosody, phonetic patterns, and spectral characteristics are among the features that are taken from audio signals and used to train machine learning models for LID. Gaussian mixture models, support vector machines, and deep neural networks are examples of common methods.

The cutting-edge field of multilingual spoken language identification (MSLID) uses machine learning to create systems that can automatically detect and distinguish between multiple languages within audio recordings. MSLID has enormous potential for applications ranging from international conferencing platforms to improving the functionality of voice-enabled technologies in an era of global communication and diverse linguistic contexts. In the larger field of machine learning,

MSLID is an essential area of research and development, as this introduction explains its fundamental challenges and applications.

Literature Review

- Multiclass Spoken Language Identification for Indian Language using Deep Learning Lakshman Rao

Author: Aria Computer Science and Engineering Vishnu Institute of Tecnhonology Bhimavaram

The goal of spoken language identification (SLID) is to recognize speech in an audio file and give every sentence a speech character. The paper develops a based on concepts convolution neural network (CNN) for Telugu, Tamil, Bengali, & Gujarati language recognition. The classifier is trained by listening to data in all four languages for five hours.

- Arabic Dialect Identification in Social Media Author name - REEM AlYAMI

Even though Arabic is spoken by more than 250 million people in 22 countries, natural language processors (NLP) still consider Arabic to be a low resource language. Modern Standard Arabic (MSA), which is taught to Arabic speakers in schools and used in formal writing, is typically utilized in the composition of Arabic texts found in formal sources.

Nonetheless, informal communication between Arabic speakers uses informal local diglossic dialects. A language is considered diglossic when speakers of the same language speak distinct dialects of it.

- Function And Process in Spoken Word Recognition Author Name : William Maelsen Wilson

It is important to investigate spoken word recognition processes from both a functional and temporal standpoint. This entails considering the roles that spoken word recognition plays in the broader process of language understanding as well as the temporal order in which the listener gets access to the sensory data required for word recognition.

- Identification of Spoken Language Using Machine Learning Approach

Author Name: Md. Ashif Shahriar

The process of identifying spoken language involves figuring out the particular language that an unidentified speaker is speaking.Additionally, we will learn a number of machine learning methods for identifying spoken language. Finding parameters and characteristics in spoken language that can be utilized to distinguish between languages is our main goal. Mel Frequency Cepstral Coefficient (MFCC) will be used to extract features from an audio file. Numerous techniques have been employed thus far for language identification (LID). Machine learning has the highest accuracy of all the techniques. For this reason, machine learning was also a part of our lid project

- Spoken Language Identificati on Using Deep Learning Spoken language identification (SLID) is recognizing the language being a talk by an anonymous speaker from an audio clip . Humans are the most error-free language identification system .Here are various implementations of spoken language identification like creating front ends for multi language speech identification systems, automatic customer routing in call centers, monitoring, and web information retrieval. SLID system has three main parts, data collection, feature removal, and language classification An essential for developing and evaluating a speech recognition system is the accessibility of a technique to generate information suitable database.
- Automatic Language Identification using Machine learning Techniques

Author Name: Hariraj Venkatesan

Research on spoken language identification in regional languages contributes to both the preservation of regional languages and the expansion of technology's reach among speakers of those languages. This paper presents our research on the identification of spoken data in four regional Indian languages: Tamil, Telugu, Hindi, and Kannada. Speech signals are fed into automatic language identification systems, which then use mathematical operations to categorize the speech signals into one of the natural languages.

To learn more about the audio and speaker, mathematical operations on the characteristics of a speech signal, like frequency or amplitude, can be performed. This work presents the use of Mel Frequency Cepstral Coefficients (MFCC) to extract speech signal features for language identification.We employed Decision Tree and Support Vector Machine classifiers for classification, and the resulting accuracies were 73% and 76%, respectively.

- Sentence Level Language Identification in Gujarati , Hindi code Mixed Scripts

Author: Md. Zuberkaz.

Languages frequently converse in multiple languages when writing and speaking Code-switching and code-mixing are terms used to describe these language changes, which involve complex grammars. The increased use of social media and online discussion forums has led to an increase in textual and non-textual communication in multiple languages. This leads to the digital creation of several multilingual hybrid legal entities. To use this corpus and any Natural Language Processing (NLP)

- Automatic Language Identification Using Machine Learning Technique

Author:Hariraj Venkatesa , T. Varun Venkatasu baramania

Both the preservation of regional languages and the spread of technology among speakers of those languages benefit from research on spoken language identification in those languages. This paper presents our research on the identification of spoken data in Telugu, Tamil, Hindi, and Kannada, four regional Indian languages. Speech signals are fed into automatic language identification systems, which then use mathematical operations to categorize the speech signals into one of the natural languages. Spoken Language Identification, or SLID, is the technique of recognizing the language used in a discourse by an unidentified speaker from an audio clip. Human language is the most error-free language identification system available.Many applications have made use of spoken language identification, such as multilingual voice recognition system front ends, web information retrieval, monitoring, and automatic customer routing in call centers. The SLID system consists of three main parts: data gathering, feature elimination, and language classification. A adequate database must be available in order to create and evaluate a speech recognition system.

Motivation

Multilingual spoken language identification through machine learning offers a powerful tool for breaking language barriers, fostering cross-cultural communication, and promoting inclusivity. By accurately identifying and understanding spoken languages, this technology facilitates seamless interactions, transcending linguistic diversity. It opens avenues for global collaboration, enhances accessibility, and empowers individuals to connect, learn, and share knowledge across linguistic boundaries. Embracing this scope not only advances technology but also contributes to a more interconnected and harmonious world.

Furthermore, the motivational scope extends to practical applications such as improved customer service in diverse linguistic contexts, enhanced security through language-based threat detection, and efficient organization of multilingual content in a globalized digital landscape. The development of robust multilingual spoken language identification models fuels innovation, making automated language recognition an invaluable asset in a wide range of industries, from translation services to voice assistants. As we harness the potential of these machine learning algorithms, we pave the way for a future where language differences are no longer barriers but rather bridges to greater understanding and collaboration.

Machine learning-based multilingual spoken language recognition has many uses, including international communication, customer support, accessibility, cultural preservation, and scientific research. In today's connected world, it is essential because it allows technology to comprehend users' preferred language and respond to them in that language..

1. International Communication and Interaction: People speak different languages in today's globally connected world. By identifying and translating spoken languages automatically, a multilingual SLR system can promote smooth communication.
2. Multilingual Customer Service: SLR in customer service applications can be advantageous for companies that operate internationally. Calls can be automatically forwarded to agents who speak the caller's preferred language with ease.
3. Law enforcement and Security: In order to monitor and identify languages in sensitive locations like airports, government buildings, and border crossings, multilingual SLR can be used in security applications.
4. Multilingual Content Curation: SLR can be used to classify and suggest content in the user's preferred language on content platforms like social media, streaming services, and news aggregators.

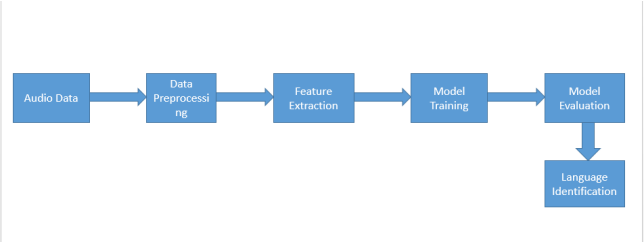

Fig 1. Block diagram

Features Extraction

Feature extraction is the process of converting input data into a feature set. The dimensionality reduction process, known as feature extraction, breaks down the initial raw data into smaller, easier-to-handle sets for processing. Several of these massive datasets involve numerous transformations that demand a lot of processing power. Techniques that combine features and/or choose variables to reduce the amount of data to be processed while accurately and fully describing the original dataset are referred to as feature extraction techniques.

After preprocessing, the incoming data goes through the following steps.

1. Select window
2. Discrete Fourier transform
3. Mail filter bank
4. Discrete cosine transformation
5. Display BFCC.

BFFC

Cepstral coefficients are obtained through the BFCC process, which combines the PLP function with the cosine transformation of spectra. To control volume control, power laws ranging from energy to MFCC and other parameters were employed, in place of the Mel filter bank, and the Bark filter bank.

- **BFCCs:** Instead of using a linear frequency scale, BFCCs use the Bark scale to capture the perceptual characteristics of sound. This makes them more aligned with human hearing. The process involves taking the cepstral coefficients calculated on the Bark scale.
- **Applications**: BFCCs are commonly employed in speech and audio processing tasks, such as speech recognition, speaker identification, and music genre classification. By utilizing a representation that aligns with human perception, BFCCs often provide better results compared to traditional MFCCs (Mel Frequency Cepstral Coefficients) in certain applications.
- **Calculation:** The calculation of BFCCs involves extracting the cepstral coefficients on the Bark scale, which typically involves filtering the magnitude spectrum of the signal with Bark-shaped filters. The resulting coefficients are then used as features for further analysis or classification tasks.

Flow chart

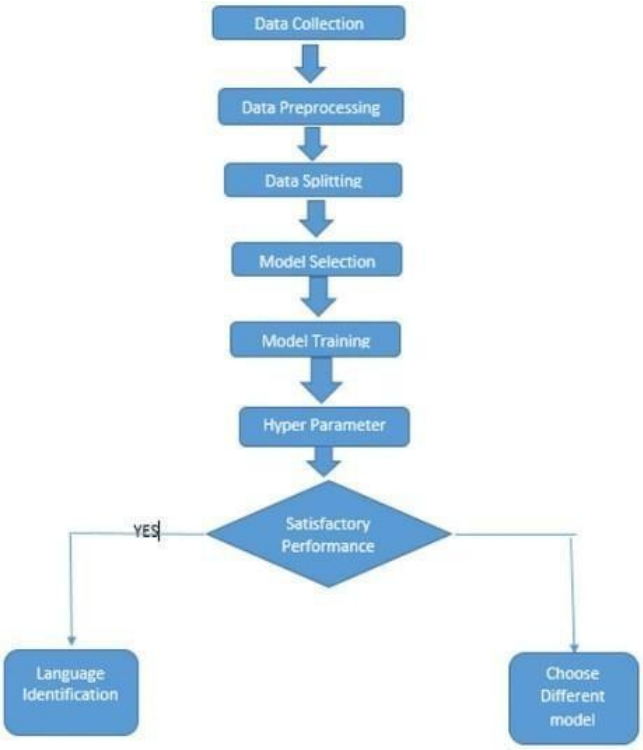

Fig 2. Flow Chart

Algorithm:

- Data Collection: In this step, a dataset of audio recordings and the corresponding language label are gathered.
- Data processing involves extracting audio properties, such as BFCCs, from unprocessed audio data to create useful representations. To enhance the performance of the model, the retrieved features might then be scaled and normalized.
- Data splitting involves dividing the dataset into subsets for testing and training. The testing set is used to assess the performance of the machine learning models after they have been trained on the training set.
- Model selection entails selecting the best machine learning algorithm, such as KNN, for spoken language identification.
- Model training involves teaching the chosen algorithm to identify and understand language patterns using the

training set.
- Optimize the model's hyperparameters to attain improved performance through hyperparameter adjustment.
- Model evaluation: Apply the relevant metrics to assess the performance of the trained models on the testing set.
- Satisfactory Performance: Assesses whether the model satisfies the intended performance standards.
- Deployment: In real-world scenarios, the model can be used to identify the language in fresh audio recordings if it performs well.

Conclusion

In this paper, we discussed and presented an experimental evaluation of our machine learning approach for language identification. The outcomes showed how well our suggested method worked to correctly identify languages from an extensive range of multilingual datasets. Our model achieved high accuracy despite some difficulties and restrictions, offering a solid basis for additional research in this area.

Expected Result

A model that can accurately and consistently identify the language spoken in a given audio segment is the expected result of a multilingual speech recognition system. Measures like accuracy, precision, recall, and F1-score are used to evaluate this, and they are based on a test set of data that the model hasn't seen during training. The overall goal is to build a robust system that can recognize spoken languages in a variety of real-world contexts and applications. The platform's intended application case and requirements would determine the exact performance expectations.

Acknowledgment

The people we need to thank the most now that the project's first phase has come to an end are those who have supported us throughout the project's creation and without whose assistance it would not have seen the light of day.

References

1) Hieronymous and Kadambe proposed a task independent spoken language identification which uses a Large Vocabulary Automatic Speech Recognition (LVASR). .
2) Rao L. Multiclass Spoken Language Identification for Indian Languages using Deep Learning . .
3) Das, Shekhar H, Roy P. A deep dive into Deep learning techniques for solving spoken language identification problems. Academic press. 2019;p. 81–100.
4) Sharma N, Jain V, Mishra A. An analysis of CNN for Image classification. *Procedia computer science*. 2018;132:377–384.
5) Kaz Z. Sentence Level Language Identification in Gujarati, hindi . .
6) Kim H, Park JS. Automatic Language Identification Using Speech Rhythm Features for Multi-Lingual Speech Recognition. *Applied Sciences*;10(7).
7) Padi B, Mohan A, Ganapathy S. Towards Relevance and Sequence Modeling in Language Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020;28:1223–1232.
8) Verma M, Buduru AB. Fine-grained Language Identification with Multilingual CapsNet Model. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM). IEEE. 2020;p. 94–102.
9) Barnard E, Cole RA. Reviewing automatic languageidentification. *IEEE Signal Processing Magazine*.
10) Waibel A, Geutner P, Tomokiyo LM, Schultz T, Woszczyna M. Multilinguality in speech and spoken language systems. *Proceedings of the IEEE*. 2000;88(8):1297–1313.