classification model to predict the gender (male or female) based on different acoustic parameters *

Veena Madhuri Ayyannagari,Ishita Pallempati[1] and T. Satyanarayana Murthy[2][0000−0002−6320−8496]

[1] CBIT , India veenaayyannagari.az@gmail.com,pallempatiishita@gmail.com
[2] Chaitanya Bharathi Institute of Technology , Hyderabad, India
murthyteki@gmail.com, tsmurthy_it@cbit.ac.in

Abstract. The recognition of a person by his voice is one of the forms of biometric authentication, which makes it possible to identify a person by a combination of unique voice characteristics and refers to dynamic methods of bio metrics. Speaker recognition is a technology that can automatically identify the speaker based on the speech waveform, that reflects the physiological and behavioral characteristics of speech parameters from the speaker. Like traditional speaker recognition systems, there are two stages, namely, training and testing. These are the main stages of speaker recognition. Learning is the process of extracting phonetic characteristics from a speaker that has already been recorded or saved as a sample, storing them in a database, and familiarizing the system with the characteristics of the speaker's voice. Testing is the process of comparing questionable sound and phonetic characteristics from a speaker recognition database.

Keywords: First keyword · Second keyword · Another keyword.

1 Introdution

In the present day we have seen a rise in personalised products and a product with higher compatibility with the user. There have been many devices/software which have used voice as a medium of communication. There are voice assistants such as Siri and Alexa which are being used in many devices to help the users improve their user experience with the products. A major issue arises when the voice systems are unable to recognise the user properly. This can be in terms of gender, race, many more factors. A study by Dr. Tatman has proven that there has been a 13 percent higher accuracy in the recognition of a male voice over a female.

The uses of voice in the present day are limited. Soon there will be technology where we can involve voice as a form of security (i.e., Voice controlled

* Corresponding Author: T.Satyanarayana Murthy

locking systems, Voice unlocked services, etc.). In such situations one must keep in mind the accuracy of the system used for voice recognition. Inaccuracy of the recognition systems can cause many problems to the users. To avoid such a situation the team has worked on different classification methods using a vast data-frame with multiple acoustic parameters to produce an accurate model for voice recognition in terms of gender.

Voice recognition, also known as speech recognition, is a fascinating area of research and application within the field of machine learning. It involves converting spoken language into written text or interpreting voice commands for various purposes. With recent advancements in machine learning algorithms and the availability of large speech datasets, voice recognition systems have become increasingly accurate and widely used in applications such as virtual assistants, transcription services, and voice-controlled devices. This project overview provides a high-level understanding of a voice recognition system developed using machine learning techniques.

Classification is a fundamental task in machine learning and data analysis, aiming to predict the class or category to which an instance belongs based on a set of input features. Numerous classification algorithms, known as classifiers, have been developed to tackle this task, each with its own strengths and weaknesses. Understanding the comparative performance of classifiers is crucial for selecting the most suitable algorithm for a given dataset and problem domain. This study aims to provide a comprehensive analysis and comparison of various classifiers, shedding light on their performance across different evaluation metrics.

This project focuses on the development of a voice recognition system using machine learning. The goal is to accurately transcribe spoken language and interpret voice commands, enabling seamless human-computer interaction. The project employs a dataset of recorded speech samples and uses machine learning algorithms to train a model capable of recognizing and interpreting various speech patterns.

The project follows a systematic approach, involving several key stages. Initially, the dataset is preprocessed, which includes steps such as noise reduction, feature extraction, and normalization. Feature extraction techniques such as Mel-frequency cepstral coefficients (MFCCs) or spectrograms are commonly used to represent the acoustic characteristics of the voice samples.

Next, a suitable machine learning algorithm is chosen for training the voice recognition model. Popular choices include recurrent neural networks (RNNs) such as long short-term memory (LSTM) networks or convolutional neural networks (CNNs). These algorithms are capable of learning complex patterns and

extracting meaningful features from the speech data.

The dataset is split into training and testing sets, and the model is trained on the training set using labeled speech samples. The model is optimized by adjusting hyperparameters and employing techniques such as regularization and dropout to prevent overfitting. The performance of the model is evaluated on the testing set using metrics such as accuracy and precision.

Once the model is trained and validated, it can be used for voice recognition tasks. In this project, the system focuses on two main tasks: speech-to-text transcription and voice command interpretation. For speech-to-text transcription, the model predicts the written text corresponding to the spoken input. For voice command interpretation, the model maps specific voice commands to corresponding actions or responses.

The project also explores potential challenges in voice recognition, such as dealing with variations in accent, background noise, and speaker-independent recognition. Techniques like data augmentation, multi-speaker training, and adaptation methods may be employed to address these challenges and improve the overall performance of the system.
The results and insights gained from this project contribute to the ongoing research and development of voice recognition systems. The project aims to enhance the accuracy and usability of voice-controlled applications, enabling more efficient and intuitive human-computer interaction.

The field of machine learning has witnessed significant advancements in classification algorithms, leading to a plethora of classifiers available for solving diverse problems. However, the performance of these classifiers can vary depending on the characteristics of the dataset and the specific task at hand. Therefore, it is essential to conduct a comparative analysis to identify the strengths and weaknesses of different classifiers and their suitability for specific scenarios.

This study presents a comprehensive evaluation of classifiers, considering various performance metrics such as accuracy, precision, recall, and area under the receiver operating characteristic curve. A diverse collection of benchmark datasets spanning different domains is utilized to ensure the generalizability of the findings.

Several popular classifiers, including but not limited to decision trees, logistic regression, support vector machines (SVM), k-nearest neighbors (KNN), naive Bayes, and random forests, are examined in this comparative analysis. The classifiers are trained and tested using standardized protocols, ensuring fair and unbiased performance assessment.

The results of the comparative analysis provide insights into the strengths and weaknesses of each classifier. Factors such as dataset size, dimensionality, class imbalance, and noise levels are taken into account to understand the impact on classifier performance. Additionally, the study investigates the influence of parameter tuning and feature selection techniques on the performance of different classifiers.

Furthermore, this research aims to identify the scenarios in which certain classifiers outperform others, enabling practitioners and researchers to make informed decisions regarding classifier selection. The findings of this study contribute to the development of guidelines and best practices for classifier selection, helping to improve the efficiency and effectiveness of classification tasks in real-world applications.

This study offers a comprehensive analysis of the comparative performance of classifiers. By examining a wide range of classifiers across various evaluation metrics and datasets, it provides valuable insights into the strengths and weaknesses of different algorithms. The results can guide researchers and practitioners in selecting the most appropriate classifier for their specific classification tasks, leading to improved performance and more accurate predictions in practical applications.

2 Literature Survey - comparative performance of classifiers

ARTICLE 1:
Ashraf Tahseen Al (2021), Elsevier: This study used experimental research design to propose a biometric-based approach for recognizing voice using machine learning systems. The advantage of this approach is that it is non-invasive and does not require an individual to provide physical samples. However, there are some disadvantages to this approach, including accuracy issues, privacy concerns, and the potential collection of sensitive information such as voice patterns, which could be misused.

ARTICLE 2:
Orken Mamyrbayev (2019), Ministry of Education and Science of the Republic of Kazakhstan, IRN AP05131207: This study used experimental research design to propose a method for scaling speech signals using the scalar method and a multilayer perceptron. The advantage of this method is that it can be scaled to handle large amounts of data and complex models. However, one disadvantage of this approach is that multilayer perceptions can become too specialized to the training data and may not generalize well to new data.

ARTICLE 3:
Jurgen Arias (2020), Youness Mansar, Audio Classification: A Convolutional Neural Network Approach: This experimental study proposed combining neural network (NN) and convolutional neural network (CNN) models together in a voting classifier by joining their probability of the predictions, achieving a 95accuracy. However, a disadvantage of this approach is the difficulty in hyperparameter tuning, which can be time-consuming. The advantage of this approach is that by better feature extraction, the strengths of both CNN and NN can be leveraged to extract better features.

ARTICLE 4:
Piero Paialunga (2021), Trimmed.mp3: This statistical study proposed using principal component analysis (PCA) reduction to reduce the dimensions of speech data. One disadvantage of this approach is that the reduced dimensions in PCA may not have any meaningful interpretation in the original feature space.

However, an advantage of this approach is that it helps to reduce the computational complexity of algorithms.

ARTICLE 5:
Melissa Stolar (2020), Speech Emotion Recognition (SER): This experimental study proposed enriching low-level features with higher-level derivatives and statistical functionals of the low-level parameters. One disadvantage of low-level features is that they typically provide limited information, and statistical information relies solely on parameters. However, statistical functions can understand the data by summarizing the key features of the data distribution.

3 Conclusion

There are four basic types of voice: soprano, alto, tenor and bass

Men's voices are deeper than women's because their vocal chords are longer. As a boy gets older, his vocal cords grow and his voice changes from high to low.

Classification is a fundamental task in machine learning and data analysis, aiming to predict the class or category to which an instance belongs based on a set of

input features

This project focuses on the development of a voice recognition system using machine learning

In conclusion, the choice of classifier should be based on a careful evaluation of the specific characteristics of the dataset, the complexity of the problem, the desired interpretability, and the available computational resources

By considering these factors, practitioners can select the most suitable classifier to achieve optimal performance in their classification task

References

1. Gender and Dialect Bias in YouTube's Automatic Captions
Rachael Tatman Department of Lingusitics University of Washington rctatman@uw.edu
2. Voice based gender classification using machine learning
A Raahul, R Sapthagiri, K Pankaj and V Vijayarajan
School of Computer Science and Engineering, VIT University, Vellore- 632014, India
E-mail: vijayarajan.v@vit.ac.in
3. Voice Based Gender Classification Using Machine
Learning
1K.Geethika Anasuya, 2 D. Soujanya,
3 Dr. K. Venkata Ramana
1Department of Computer Science and Systems Engineering, Andhra University College of Engineering
(A), Andhra University, Visakhapatnam - 530003, India
2Andhra University College of Engineering (A), Andhra University, Visakhapatnam - 530003, India
4. GENDER RECOGNITION USING VOICE
Neenu PA, Reesha PU MSC Scholar , Assistant Professor Department of Computer Science,
St. Joseph's College (Autonomous) Irinjalakuda, Thrissur, India
Gender recognition is main problems in the society. Analyzing the gender from sound properties such as mean, wavelength, frequencies, skewers etc.Machine learning is the advanced method in IT field which helps in identifying the gender.We use SVM algorithm to classify the gender whether it is male or female. This paper is helpful to transgender because to enhance voice therapy.
5. JOURNAL OF CRITICAL REVIEWS ISSN- 2394-5125 VOL 6, ISSUE 06 , 2019
2595 Gender Voice Classification using Deep Learning Convolutional
Neural Networks D. Aravinda1
, N. Arpitha1
, M. Mamatha1
1Department of CSE, Sree Dattha Institute of Engineering and Science Hyderabad, India.
6. Speech Based Gender Identification UsingEmpirical Mode Decomposition (EMD) A Thesis submitted in partial fulfillment of the requirement for the degreerequirement for the degree of Bachelor of Science in Computer Science andEngineering of BRAC UniversityByMahpara Hyder Chowdhury