JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review on Speech Emotion Recognition Using Machine Learning

Prof. Swati R. Khokale, Kshitij Nehete, Bhushan Aher, Adesh Kadam, Shriraj Gosavi

***Abstract* - Emotion recognition from audio sources necessitates feature extraction and classifier training. The feature vector is made up of audio signal elements that characterise speake specific properties like tone, pitch, and energy, It's also utilised to train the classifier model to identify a specific emotion. The training and testing components of the open source dataset for North American English were manually separated. Frustration, joy, and depression were discovered using mel frequency cepstral coefficients (MFCC) to elicit speaker vocal tract data from audio recordings. These obtained feature vectors were fed into the classification model. Following the extraction of the test dataset, the classifier will identify the underlying emotion in the test audio. As training and test databases, a Northern American English acted and natural speech corpus, real-time input English speech, and regional language databases in Hindi and Marathi were used. The impacts of increasing the number of feature vectors delivered to the classifier, as well as the two methods utilised on feature vectors, are investigated. It investigates the efficiency of speech classification in Indian English, Hindi, and Marathi. For Indian English speaking, a faithfulness of 80% was achieved.**

Keywords - [Machine Learning, Speech Emotion Recognition, SVM, MFCC, Data Analysis, SER, Feature Extraction]

1. Introduction

The best Python project ever is speech emotion recognition. Call centres are the best example of this. If you've ever observed, contact centre employees never speak in the same way; their pitching/talking and videos. Once such content is posted, they live style varies depending on the customer. Now, this does occurs to regular people as well, but how does this apply to contact centres? Here's your answer: staff understand clients' emotions based on their speech, allowing them to better their service and convert more consumers. They are employing voice emotion recognition in this way. Speech Emotion Recognition, or SER is the way of attempting to interpret human emotion and affective responses from speech.This takes use of the reality that tone and pitch in the voice frequently indicate deeper emotion. This is the same mechanism that animals such as dogs and horses use to interpret human emotion. The challenge of recognising the emotional qualities of speech, regardless of the semantic content, is known as Speech Emotion Recognition (SER). While people are capable of performing this activity efficiently as a natural aspect of voice communication, the capacity to do so inevitably using programmable devices is still a work in progress. Automatic emotion detection systems are being researched in order to develop efficient, actual techniques of identifying the emotions of mobile phone users, call centre operators and consumers, car drivers, pilots, and a variety of other human-machine interface users. Introducing emotions to machines has been identified as a vital step in achieving a humanlike appearance and behaviour.

We'll utilise the libraries librosa, sound file, and sklearn (among others) to create an MLP Classifier model in this Python project. This will be able to detect emotion in audio files. The data will be loaded, features extracted, and the dataset divided into training and testing sets. The model will then be trained using an MLP Classifier that has been initialised. Finally, we'll figure out how accurate our model is. Because emotions are subjective and annotating audio is difficult, SER is difficult

2. Literature Survey

According to survey [1], To discover nonlinear systems, we employ a deep learning method called limited Boltzmann machine. A random search strategy was used to create the neural model, which has a deep architecture. This deep neural network's beginning weights come from constrained Boltzmann machines. We illustrate how to use unsupervised learning algorithms and input data to uncover nonlinear systems. The weights are learned via conventional supervised learning utilising the output data.

According to survey conducted by Arunkumar Balakrishnan [2], The multi-strategy machine learning system (MSMLS) stores information about each strategy instruction and uses it to address the problem of learning strategy picking. The student's error can be traced back to either a misapplication of knowledge or a misapplication of learned knowledge. Learned information errors can be traced back to an ineffective learning approach. The pupil is 'opened' to the trail due to a poor selection of learning strategy.

According to survey conducted by Sergiy Fefilatyev; Volha Smarodzinava; Lawrence O. Hall; Dmitry B. Goldgof [3], The multi-strategy machine learning system (MSMLS) stores information about each instructional strategy and uses it to solve the problem of learning strategy picking. The student's error can be traced back to either a misapplication of knowledge or a misapplication of learned knowledge. Learned information errors can be traced back to an ineffective learning approach. The pupil is 'opened' to the trail due to a poor selection of learning strategy.

According to survey conducted by Halil Ibrahim Bülbül; Özkan Ünsal [4], Many image-related applications, such as recognising ships on the horizon, flight control, and port security, need the detection of a horizon in an image. To detect a horizon line in an image, the majority of available systems rely solely on image processing technologies. The SVM, J48, and naive Bayes classifiers were employed to solve the problem, and their performance was evaluated. On a 20-image data set, precision of 90-99% in recognizing the horizon was obtained.

According to survey conducted by Ilia Nouretdinov, Brian Burford, Alex Gammerman [5], Not only does the technique allow us to make judgments, but it also includes metrics of predicted accuracy and reliability. By choosing an appropriate self- esteem, it is feasible to control the amount of mistakes.

Saurabh Roy, Manisha J. Nene [6] Using Machine Learning, this study proposes the notion of Station-of-Things (SoT) for detecting distinct vehicle kinds that are available in parking slots (ML). The development of an intelligent SoT environment has been made possible by advanced research in communication, processing, and sensor invention. Similarly, recognising various objects such as vehicles in parking spaces, goods in warehouses, and various furniture in furniture showrooms, all of which have different dimensions, but the same system can be used to estimate objects by changing the aspects of the objects in the ML toolkit database.

Halil Ibrahim Bulbul, Nese Usta, Musa Yildiz [7] The effort will analyze MLP (Multi-Layer Perceptron) and SVM (Support Vector Machine) classification approaches that have never been compared using these signals. The literature will help to find a more efficient mechanism with new applications in the study. Its goal is to design a strategy for enhancing MLP and SVM calculation time and standard classification performance, as well as to contribute to the informed awareness of this effort.

A. Smailagic [8] We discover the following in this paper: Context-aware computing refers to when a mobile computer is aware of its user's current condition and environment and adjusts its behaviour accordingly. We have proved the ability of our technology to generate a valid user context model automatically, as well as conducted experimental measurements and evaluations. to locate This allows us to use a completely unsupervised machine learning approach. Unsupervised learning refers to the process of determining a user's context without having to manually annotate current user states. Using this discretization allows for learning from user experiences. Each observed user interaction is regarded as a labelled example that may be used to build a statistical model for context-dependent preferences.

According to survey conducted by Shadi Banitaan; Mohammad Azzeh; Ali Bou Nassif [9], The goal of this study is to evaluate how machine learning approaches influence the accuracy of user movement prediction. F-measure and Precision recall are measured utilizing five base classifiers and two ensemble learning algorithms. Precision recall is used to assess the results and F-measure using five base classifiers and two ensemble learning techniques.

According to survey conducted by Jan Outrata [10], For machine learning, we discovered two approaches for preparing input data (ML). The first entails adding new attributes to the set of attributes characterising

items in the input data table, while the second entails changing existing attributes with new attributes. The methods make use of formal concept analysis (FCA) and boolean factor analysis, both of which were recently described by FCA, in which the new characteristics are specified by so-called factor concepts computed from the input data table. For data pre-processing, we use Boolean Factor Analysis, and we are familiar with data processing.

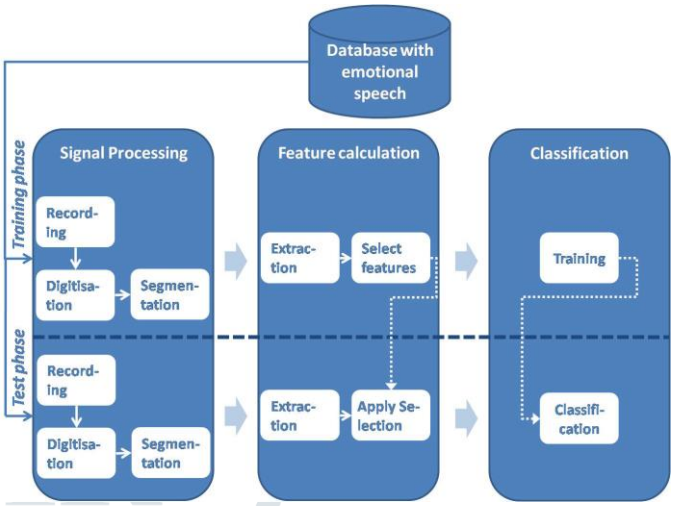
According to survey conducted by Ion C. Fre, man, Ashley J. Haigler [11], Through the prism of multiple machine learning methods, this paper explores industry-based doctorate dissertation research in a professional computer doctoral programme for full-time working professionals. To gain a better understanding of issues investigated by full-time working industry specialists. The IBM Watson and machine learning techniques are the subjects of this research report. The study sheds light on the distinctions in machine learning algorithm categorisation.

According to survey conducted by Corey A. Graves; John W. Campbell; Jung H. Kim [12] There have been numerous studies that have sought to categorise pupil attentiveness. Many of these methods relied solely on qualitative data and did not include any quantifiable information. The goal of this project is to close the gap between qualitative and quantitative techniques to classifying student alertness. Instructors can use the data from these algorithms to get vital feedback on the effectiveness of their instructional tactics and approaches.

According to survey conducted by Jian Zhang [13], Transfer learning aims to improve a specific learning task by combining it with data and other related learning problems. The connection between the auxiliary and goal tasks can be remote in many circumstances. Only a handful of the auxiliary data's attributes may be useful for target learning. To find a collection of hierarchical features from the auxiliary data, the method use a constrained Boltzmann machine.

Shamsuddin S. Khan, Rohit Barve, Umesh Kulkarni [14], Shamsuddin S. Khan, Rohit Barve, and Umesh Kulkarni are the authors of this study. We discover that The goal of machine learning is to develop algorithms that can learn on their own and improve over time, and that can be used for prediction Algorithms that can engage with agents and optimize their learning behaviour as a result of these interactions.The consequence of pseudo intelligence increases the effectiveness in delivering improved survivability of humans, according to our analysis.

3. Proposed System



Human emotion is mostly a product of mixed feelings, according to the proposed system. As a result, they proposed an upgrade to the SVM algorithm that takes into account mixed signals and selects the most dominant one. A ranking SVM algorithm was chosen for this objective. The ranking SVM applies all of the results from the individual binary classification SVM classifiers, also known as rankers, to the final multiclass issue. Their system attained a 44.40 percent accuracy using the ranking SVM algorithm.

To figure out what the audio is about, we first translated it back into test mode using the voice Emotion api. Second, plot the waveform and spectrogram to observe sample audio files in order to test the audio files. The next step is to clean audio files by decreasing the sampling frequency and removing any unwanted noise from the raw audio.Extraction of Features from Audio Files: The extraction of features from audio files is the next step. We use the LiBROSa Library in Python for extracting features, which is one of the libraries that may be used to analyse audio. Each audio is classified into one of the specified labels when a fresh dataset is loaded and the feature extraction process is performed.

4. Conclusion

In this paper, we suggested a method for extracting the emotional characteristics parameter from an emotional speech signal using deep belief networks (DBNs), a type of deep neural network. We suggested a classifier model based on deep belief networks (DBNs) and support vector machines (SVMs) by combining deep belief networks (DBNs) and support vector machines (SVMs) (SVM).In the practical training phase, the model has a low degree of complexity and a final recognition rate that is 7% higher than the conventional artificial extract, and this method can accurately obtain emotion characteristic parameters, resulting in a significant increase in the recognition rate of emotional speech recognition. However, training DBN's feature extraction model took 136 hours, which was considerably longer than other feature extraction techniques. We will continue to investigate speech emotion recognition using DBNs in the future and expand the training data set. Our ultimate goal is to figure out how to boost speech emotion recognition rates.

References

[1] Erick De La Rosa; Wen Yu, "Restricted Boltzmann Machine for Nonlinear System Modelling" 2016, IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA doi : 10.1109/ICMLA.2015.24

[2] Arunkumar Balakrishnan "Development of an Ontology of Learning Strategies and its Application to Generate Open Learner Models." 2014, International Conference on Machine Learning and Applications Miami, FL, USA doi: 10.1109/ICMLA.2009.58

[3] Lawrence O. Hall; Dmitry B. Goldgof "Horizon Detection Using Machine Learning Techniques" 2017, 5th International Conference on Machine Learning and Applications (ICMLA'06), Orlando, FL, USA doi : 10.1109/ICMLA.2006.25

[4] Halil Ibrahim Bülbül; Özkan Ünsal "Determination of Vocational Fields with Machine Learning Algorithm" 2016, Ninth International Conference on Machine Learning and Applications, Washington, DC, USA doi : 10.1109/ICMLA.2010.109

[5] Ilia Nouretdinov, Brian Burford, Alex Gammerman "Application of Inductive Confidence Machine to ICMLA Competition Data" 2013, International Conference on Machine Learning and Applications, Miami, FL, USA doi : 10.1109/ICMLA.2009.24

[6] Saurabh Roy, Manisha J. Nene, "Machine Learning for Object Prediction in Station-of-Things Environment" 2020, 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India doi : 10.1109/ICISC47916.2020.9171098

[7] Halil Ibrahim Bulbul, Nese Usta, Musa Yildiz, "Classification of ECG Arrhythmia with Machine Learning Techniques" 2017, 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico doi : 10.1109/ICMLA.2017.0-104

[8] A. Smailagic, "Unsupervised machine learning and cognitive systems in learning user state for context-aware computing" International Conference on Machine Learning and Applications, 2004. Proceedings, Louisville, KY, USA doi : 10.1109/ICMLA.2004.1383485

[9] Shadi Banitaan; Mohammad Azzeh; Ali Bou Nassif, "User Movement Prediction: The Contribution of Machine Learning Techniques" 2016, 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA doi : 10.1109/ICMLA.2016.0100

[10] Jan Outrata, "Boolean Factor Analysis for Data Pre-processing in Machine Learning" 2015, Ninth International Conference on Machine Learning and Applications, Washington, DC, USA doi : 10.1109/ICMLA.2010.141

[11] C. Fre, man, Ashley J. Haigler, "What are they Researching? Examining Industry-Based Doctoral Dissertation Research through the Lens of Machine Learning" 2018, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA doi : 10.1109/ICMLA.2018.00217

[12] Corey A. Graves; John W. Campbell; Jung H. Kim. "Using Support Vector Machines to Classify Student Attentiveness for the Development of Personalized Learning Systems" 2013, 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA doi : 10.1109/ICMLA.2013.66

[13] Jian Zhang, "Deep Transfer Learning via Restricted Boltzmann Machine for Document Classification" 2019, International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, USA doi : 10.1109/ICMLA.2011.51

[14] Shamsuddin S. Khan, Rohit Barve, Umesh Kulkarni, "Proposed model on Prediction and Analysis using application of Health care" 2018, 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India doi : 10.1109/CESYS.2018.8723946