Sound Classification using Machine Learning and Neural Networks

Pooja R K.[1], Srishti Shetty[2], Suhani M[3], Mr. Janardhana D.R.[4]
[1,2,3] *Student, Department of Information Science engineering, SCEM, Mangaluru*
[4] *Asst. Professor, Department of Information Science engineering, SCEM, Mangaluru*

Abstract- **Classification of sound automatically has been a growing field in research. Researches are mostly performed on the sonic analysis on environment sounds because of its various applications to a large scale content based multimedia indexing and retrieval. These researches are mostly focused on music or speech recognition. The Urban Sound dataset created by Justin Salamon, Christopher Jacoby and Juan Pablo Bello in 2014 is one of the few free large sound dataset. Classification of sound data is done using feature extraction. The features of sound data cannot be conveyed in vector forms such as other type of data like images and texts. Hence, the feature extraction for sound data is less unequivocal. Two categories of feature extraction techniques are applied namely, signal characteristic feature extraction and Time series feature extraction. The validness of distinct models on each method, including tests of Random Forest, Naive Bayes, Support Vector Machines and Neural Network architectures which includes deep neural network, convolutional neural network and recurrent neural network. After implementing the machine learning techniques and neural networks we are able to classify different sounds.**

Index Terms- **classification, feature extraction, random forest, multi class SVM.**

I. INTRODUCTION

Machine learning is a subfield of Artificial Intelligence(AI). It plays a significant role in various applications, such as data mining, natural language processing, image recognition and expert systems. The aim of machine learning is to accept the structure of data and fit that data into models that can be analyzed and applied by people. Machine learning algorithms lets computer to train the data inputs and use statistical analysis in order to output values that fall within a specific range. Deep learning is a part of machine learning. These include networks which are capable of learning unstructured or unlabeled data. Deep learning is also known as Deep Neural Learning or Deep Neural Network. Learning can be supervised, semi-supervised or unsupervised. One of the main advantage of this technique over machine learning techniques is its capability to produce new features from limited set of features present in the training data set. Deep neural networks also play a significant role in various applications like speech recognition, image recognition, etc. Classification of sound automatically has been an emerging field of research. A major work on classification of sound resulted in the creation of UrbanSound dataset. UrbanSound dataset is not only based on sounds related to nature, human, animal, but also on the sources of sound like engine's sound, car horn, etc. Extraction of features is the greatest challenge for sound data classification. To overcome that, various feature extraction techniques have been applied and comparison of performances of the model on various feature sets is carried out. For classification, classifiers such as random forest, support vector machine, deep neural networks, convolutional networks, etc are applied. The one which yields highest accuracy is been considered. Here accuracy is in terms of identifying sounds even in noisy conditions.

II. LITERATURE REVIEW

Many methods have been implemented for preventing copyright infringements based on different technologies. In this session is discussed about few of these works.

Chih-Wei Chang et al., [1] designed a paper to overcome feature extraction which is one of the greatest challenges for classification of sounds. The problem is that the features of sound data cannot be represented in vector format like images and texts. so, the feature extraction for sound data is less unequivocal. Here they have illustrated two methods

of extracting features. The first method concentrates on maintainingthe time-series nature of the audio sequence and second method concentrates on expanding signal characteristics. These extraction techniques refines significant piece of information in each of the frames of the original signal. Here UrbanSound 8k Dataset is used. The UrbanSound8k dataset includes 8723 real-time recording samples from 10 classes of distinct sound sources. In the first category, characteristics are extracted from each sample, therefore, the number of features is constant irrespective of the shape of raw data. Isolation of Mel Frequencies Cepstral Coefficients(MFCC) is the most familiar technique used. It has become competitive baseline for bench markingnew techniques. In the second category, that is, filterbanks allows us to retain thetime-series attribute of the raw data. Later on, they tested with Recurrent Neural Networks (RNN), Deep Neural Networks(DNN) and a Convolutional Neural Networks(CNN). By seeing the results it has given that CNN and DNN has higher accuracy compared to others. CNN has better performance and is more robust compared to DNN but CNN is three times slower than the DNN. Therefore, CNN is suitable for smaller datasets so in case of larger datasets DNN can be applied.

Karol J.Piczak [2] designed a paper which aims at valuating whether convolutional neural networks can be effectively applied to environmental sound classification tasks, especially concentrating on the limited Nature of datasets present in this field. A deep model consists of 2 convolutional layers with max-pooling and 2 fully connected layers. It is trained on a low level representation of segmented spectrograms. Three public datasets are considered, they are: urbansound8k, ESC-50 and ESC-10. The accuracy of the network is valuated on these datasets. Implementation is done on mel-frequency cepstral coefficients(MFCC) feature. Experiments conducted based on manually engineered features show that a convolutional model outperforms common approaches and achieves a similar level as other methods. Results show that convolutional neural networks can be effectually applied in environmental sound classification tasks even with limited datasets.

Ian McLoughlin et al., [3] designed a paper which outlines a sound event classi

fication plan that compares the auditory image front end features with spectrogram image based front end features using support vector machine and deep neural network classifiers. Performance is estimated on a standard robust classification assignment in different levels of noxious noise and with several system improvement and shown to compare well with current state of the art classification techniques. Initially, they analyzed the use of google-style SAI features with a back-end SVM classifier and then use this baseline to check the effect of modification to extract feature and serve as process. The performing system in SVM classifier is altered with a DNN back-end,then the DNN classification performance is valuated with a number of different feature that are imitating from SIF.

Khine Zar Thwe et al., [4] designed a paper that describes classification of environmental sound event with time frequency representation such as spectrogram and Multi support vector machine is used as classifier. This paper characterizes feature extraction method for environment sound event classification contingent on representation of time frequency like spectrogram. Spectrogram is defined as a visual representation of spectrum of frequencies of sound as they fluctuate with time. Three portions are carried out to perform environmental classification. In the first portion, input signal is transformed into spectrogram image with the illustration of time frequency using short time Fourier transforms. Second portion extracts features from existing spectrogram with local binary pattern of three distinct radius and neighborhood sizes. Resulting features are integrated and taken as a single feature vector. Lastly, using multi support vector machine environmental sound event is classified. Using ESC-10 dataset valuation is performed. By using local binary pattern, more robustness acquired.

Huan Zhou et al., [5] designed a paper to classify environmental sounds using new convNet. New convNet(convolutional neutral network) works as a classifier for multiple input and adopts input's 2D mel-spectrogram low-level representation as its input layer. UrbanSound8K dataset has been used. This paper mainly aims at traversing the performances of classification by implementing multiple network inputs with discrete time resolutions. Obtained results confirmed the benefaction of time resolutions on performance of classification and excellent performance is obtained for input with moderate time resolution.

Bibek Luitel et al., [6] designed a paper to classify urban sound events such as bus engine, bus horn, car horn and whistle sounds. The above sounds are taken as they play a large role in traffic outline. A real time data is gathered from the live recordings at major areas of the urban city. Before the detection of events, the class of the events are analyzed using signal handling techniques. In addition to that features such as MFCC are extracted based on the analysis of a spectrum. Classifiers such as artificial neural networks, random forest, na¨ıve-Bayesian are used at two levels. Artificial neural networks are better to outline the non-aligned data when compared to other classifiers. Removal of noise from the signal is one of the main facet that must be carried out. The work towards the exposure of numerous and reckoning the traffic levels is must in order to minimize the human health issues. It also increases the senility of human life which is more important than anything else.

III.SYSTEM DESIGN AND IMPLEMENTATION

Architecture diagram aid us to understand, analyze and convey our point of view about the system structure and the user requirement that the system supports.

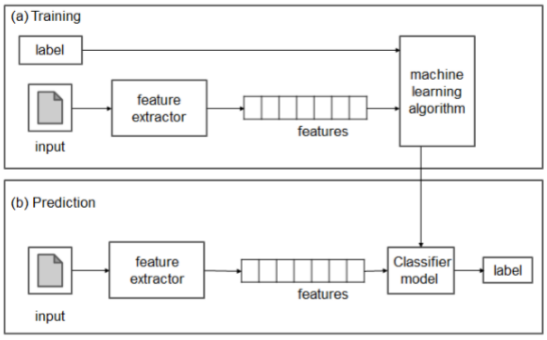

Figure 3.1:Architecture Diagram for Proposed System

The above figure shows architecture diagram of sound classification using machine learning techniques and neural networks. In this proposed software, supervised learning method is used. During training, an audio input is given to the software where the input is given to the feature extractor. Feature extractor extracts input features (such as MFCC, etc) and transforms into a feature set. This feature set is fed to the classifier model produced by the machine learning techniques and neural network algorithms. The classifier model performs classification process on feature set and labels them which later on is stored. During testing, the given input is processed in the same way as in the training phase but when it comes to classifier model, the features will be classified and compared with the labels that is stored in a file and according to that it is labeled and displayed as the output.

Initially, a single audio file is taken for processing. Normally, 193 chromatographic features are extracted from audio file. Features includes: Mel-Frequency Cepstral Coefficients(MFCC), Chroma, Melspectrogram, Spectral Contrast, and Tonnetz. Time to extract features for a single file is calculated. Further, features are extracted from all the audio files present in dataset and time required for it is calculated

Classification: Random forest and multi calss SVM classifiers are used for classification. Extracted features are classified and labeled in training phase. These labels are later on converted into class numbers. In testing phase, after extracting features it is classified and mapped with the trained data and it is labeled. Output is shown in the form of confusion matrix

IV. RESULT ANALYSIS

Result and Analysis section deals with all the output obtained from all the various modules of the project. The analysis is specially meant to explain the inference of each output obtained.

Figure 4.1 gives the detailed representation of total number of audio samples present in each of the category of sounds say either a siren or the sound of the air conditioner or any other sound category that is denoted alonf the Y-Axis shown in Figure 4.1

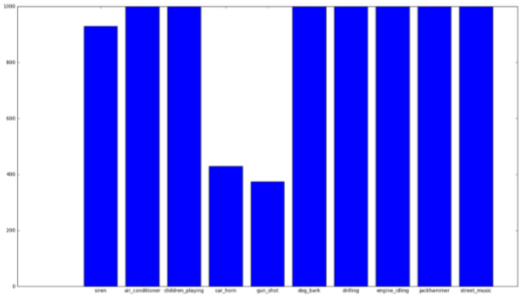

Figure 4.1: A graph showing count of audio samples in each of the categories of sounds in Sound Classification

Figure 4.2 is the graphical representation of the confusion matrix for the sound classification by the Random Forest classifier. The different colours present in the graph denote different match cases. The major four cases present in the confusion matrix is true negative, true positive, false negative and false positive. The brighter colors along the diagonal indicate the perfect matches, i.e. true positives and true negatives. The dark colors or to be precise the black colour along the graph indicates the error in matching, i.e. false poitives and false negatives. The mild colors on either side of the diagonal is the matches for the background sound in the given audio files. The Accuracy of the Random Forest Classifier based on the confusion matrix is found to be 62%.

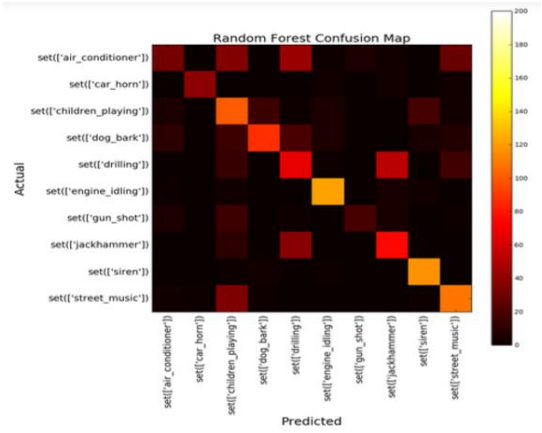

Figure 4.3: Random Forest confusion map of Sound Classification

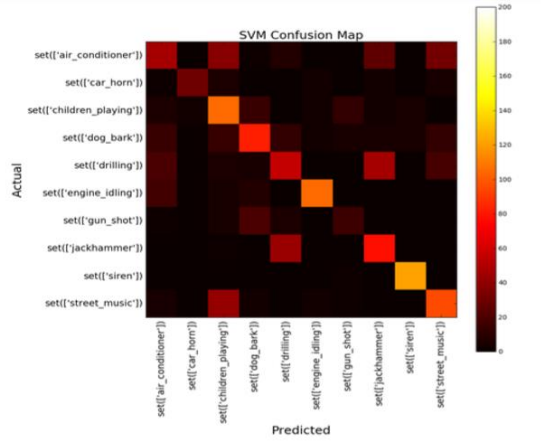

Figure 4.4: SVM confusion map of Sound Classification

Figure 4.4 is the graphical representation of the confusion matrix for the sound classification by the Support Vector Machine classifier. The cases generated in this confusion matrix is exactly as described for Figure 4.3. The Accuracy of the Support Vector Machine based on the confusion matrix is found to be 60%.

V. CONCLUSION AND FUTURE WORK

The project "Sound Classification using Machine Learning Techniques and Neural Networks" successfully classifies and labels sounds even in noisy conditions. Random forest algorithm yields highest accuracy of 62%. This can be further enhanced in terms of accuracy along with live recording and classification of that recording without any constraints.

REFERENCES

[1] Chih-Wei Chang and Benjamin Doran, "Urban Sound Classification: With Random Forest, SVM, DNN, RNN and CNN Classifiers"CSCI E-81 Machine Learning Final Project Fall 2016.

[2] Karol J. Piczak," Environmental Sound Classification With Convolutional Neural Networks", MLSP 2015 IEEE 25th International Workshop, ISBN(E):978-14673-7454-5.

[3] Ian McLoughlin, Haomin Zhang, Zhipeng Xie, Yan Song and Wei Xiao,"Robust Sound Event Classification Using Deep Neural Networks" ,IEEE/ACM Transactions on Audio, Speech and Language Processing(Volume:23, Issue:3, March 2015), ISSN(E): 2329-9304.

[4] Khine Zar Thwe and Nu War, "Environmental Sound Classification Based On Time-Frequency Representation",SNPD 2017 18th IEEE/ACIS International Conference, ISBN(E):978-1-5090-5504-3.

[5] Huan Zhou, Ying Song and Haiyan Shu, "Using Deep Convolutional Neural Net work to Classify Urban Sounds", Region 10 Conference, TENCON 2017 – 2017 IEEE, ISSN(E): 2159-3450.

[6] Bibek Luitel, Y.V.Srinivasa Murthy and Shashidhar G. Koolagudi,"Sound Event Detection in Urban Soundscape using Two-level Classification", Distributed Computing, VLSI, Electrical Circuits and Robotics(DISCOVER), IEEE, ISBN(E):978-1-5090-1623-5.

[7] Michael Blaha and James Rumbaugh, "Object-Oriented Modelling and Design with UML", 2nd Edition, Pearson Education, 2005, pp 21--157.