Discrimination of reflected sound signals

Information Technology
Machine Learning
Mou Saha
1327931
mou.saha@stud.fra-uas.de

Information Technology
Machine Learning
Riyad-Ul-Islam
1324662
riyad-ul.islam@stud.fra-uas.de

Information Technology
Machine Learning
Tanvir Hassan
1323951
tanvir.hasan@stud.fra-uas.de

***Abstract*—In pattern recognition and artificial intelligence, sound signal classification has large theoretical and functional importance. In this paper, reflected sound signals using SVM classification is being discriminated and more specifically RBF kernel is used. The results of experiments show that the audio classification system designed in the paper can classify audio signal effectively, and the average identification accuracy is about 86%.**

Keywords—Classification, Oversampling Support Vector Machine(SVM), Radial Basis Function(RBF), MFCC, Standardization

I. Introduction

Audio content is often used to make people understand the semantic content of multimedia. As the volume of audio data grows, effective digital content management becomes more important. The following are some of the reasons why audio classification is important: (a) Different kinds of audio can be handled accordingly. Each graded audio piece will be stored and indexed separately in order to facilitate compare and retrieval. During the retrieval process, the scanning space is limited to a specific subclass after classification.(b) During the retrieval process, the searching space is limited to a specific subclass after classification. Audio classification is a pattern recognition issue that is divided into two parts: feature extraction and classification based on certain features[4].

Human brain can easily discriminate task as controlling mechanism is in the human mind. The human mind can easily distinguish between large ranges of sound and correctly assign them to semantic classes, but this is not the case for computers, which receive signals in the form of a sequence of numerals with no semantic context[9].

A large number of audio classification algorithm have been proposed in literature to distinguish most audio into voice, music, and noise based on various features and characteristics. Many other studies have been conducted to improve audio classification algorithms and allow them to differentiate between more groups. SVM learns an optimal separating hyper-plan to minimize the likelihood of misclassification, as opposed to HMM, a generative model. It seems to be more appropriate for grouping than HMM. SVM have been used to classify and segment audio in a few studies. Different groups can have overlapping or interconnected areas due to the complicated attribute delivery of audio files. A kernel-based SVM is well-suited to dealing with such a case[10].

The paper organizes as follows, the section II explains the methodology for sound classification is discussed. Further Section III discusses the experiments and results obtained, in section IV conclusion and future work is discussed.

II. Methodology

Fig.1 explains the implemented methodology for classification of audio sample of 5 objects. In Section A, the data was collected from Excel file that mainly consisted of audio amplitudes and the preprocessing of data for classification. Section B describes the feature extraction procedure based on MFCC. Finally Section C describes the SVM based algorithm used for classification.

A. Dataset and Preprocessing

There are 5 objects for our experiment in 5 different excel files which represents the sample audio data of that object type. In total they have 1465 samples . These audio signals were preprocessed before feature extraction. 70% of the total collected samples were used as training data while 30% was used for testing.

B. Feature Extraction and Scaling

Mel Frequency Cepstral Coefficients (MFCCs) are the most used features for explaining the spectrum of an audio recording in a concise yet detailed manner. This feature had been mostly used for automatic speech and sound recognition. Rather than Hz, they are expressed in Mel units, which are calculated on a logarithmic scale[6].

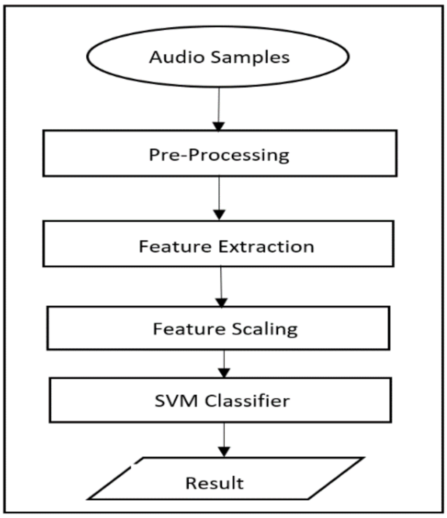

Fig. 1. Block diagram of support verctor machine classification

In the frequency domain, it's the same as convolution. The Hamming window is useful since it prevents discontinuities and minimizes mismatch. The windowed signal's Fast Fourier Transform (FFT) is then estimated. The MFCCs were calculated by grouping each frame's transformed coefficients to another set of coefficients, which were actually the filter response to a set of 20 triangular filters. Such filters are generally related to the human hearing system's frequency perception spectrum. The logarithm of the coefficients was then calculated, followed by a discrete cosine transform (DCT). The DCT method can be used to de-correlate overlapping coefficients. Since the recognition computation is degraded by rapidly increasing higher DCT coefficients, only the first thirteen coefficients were used as features to simplify the process[7].

Feature Scaling is important in machine learning as most of the times dataset will contain features that are highly varying in magnitudes, units and range. The list of features has the greatest impact on distance algorithms like SVM, KNN, and K-means. This is due to the fact that they use distances between data points to find their similarity. There are two kind of feature scaling: Normalization and Standardization. When you know that the data does not follow a Gaussian distribution, normalization is a good option. This is useful in algorithms like K-Nearest Neighbors and Neural Networks, which do not expect any data distribution. In cases where the data follows a Gaussian distribution, on the other hand, standardization can be effective. Standardization, unlike normalization, does not have a bounding range. As a result, even though the data contains outliers, standardization will have no effect on them. As SVM is used for classification and Standardization is used as feature scaling.

C. *Classification using Support Vector Machine*

In this section, the basic theory of the SVM is introduced first. SVM is a supervised machine learning algorithm that can be used to solve linear and nonlinear classification and regression problems. As a statistical learning theory implementation, the SVM was created to solve broad margin classification problems[1]. It is preferred over other classification algorithms as it needs less computing power and has high precision. It has also been shown to perform well in highly dimensional data, especially when the sample size exceeds the number of dimensions. It is also memory-efficient. In Fig. 2 , it illustrates a linear classifier in which the SVM trained samples of two groups are separated by a hyperplane[2].

One of the SVM's features is that it utilizes the kernel trick and the kernel method is the most crucial factor in SVM. The kernel's strategy is to take data and convert it into the appropriate format. Different forms of kernel functions are used by various SVM algorithms. The practical idea behind kernel function is to return the inner product between two points in a feature space. There are several kinds of functions that can be used. Linear, nonlinear, polynomial, RBF, and sigmoid are some examples[3].

Kernel function is used for sequence data, signal ,graphs, text, images as well as vectors. The most used type of kernel function is RBF as the result is localized and finite over the entire x-axis. Moreover it is a general-purpose kernel which is used when there is no prior knowledge about the data. This is one of the reason this methodology is being used for this experiment. SVM classification consists of two steps: training of all classes and testing of test data against the trained data set.

The reference resources of RBF network used in this paper is [5] Y. Tsai, Chin-Chien Tsai and Kun-Ching Wang. The

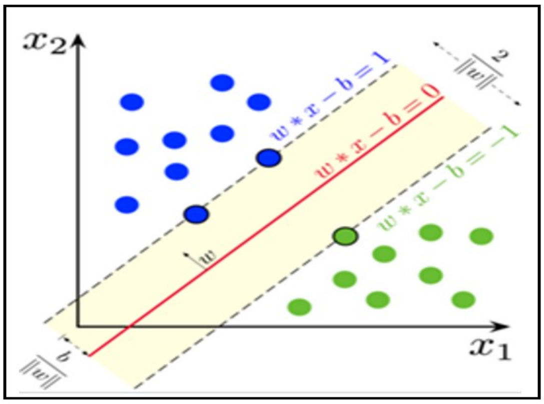

Fig. 2. Dataset representation and margin using support verctor machine.

RBF network's performance is calculated as a linear of K basis functions.The RBF kernel is a function whose value is equal to the difference between the origin and a given point.

$$G(x) = \sum_{i=1}^{k} w_i g_i(x) \qquad (1)$$

Where $\boldsymbol{w_i}$, i=1,…..,K ,denotes the weights of the output layer. The Gaussian Basis Function $\boldsymbol{g_i(x)}$ are defined as

$$g_k(x) = \exp\left(-\frac{\|x-\mu_k\|^2}{2\sigma_k^2}\right) \qquad (2)$$

Where $\boldsymbol{\mu_k}$ and $\boldsymbol{\sigma_k^2}$ denotes means and variances respectively. The Gaussian terms are integrated into a single output value by the weight coefficients $\boldsymbol{w_i}$. To create a Gaussian RBF network consists of following steps[5]-

1. Calculate the total number of Gaussian basis functions to be used for each output class and the whole process.
2. find the centers of the Gaussian basis functions;
3. determine the cluster variance for each Gaussian basis functions;
4. solve for the weight coefficients and bias in the summation term.

III. Experiment and REsult

The experiment were conducted on 5 object which are from object 1 to 5.To classify if it is object 1 or not, target class-1 is assigned for Object 1 and class-0 for others. And so all dataset is divided into two classes, either it is in Class-0 or in Class-1. For training all the objects data and then for testing data a graphical user interface has been created which is illustrated in Fig. 3. In the testing by defining a row index, a single object can be classified or all the objects can also be detected by selecting the test file.
Before any kind of pre-processing Class-0 had total of 1150 and Class-1 had total of 315 data in Table I. This is where imbalance data be can be a problem as it is difficult to convince your method to predict with better accuracy. In order to improve the performance random oversampling technique is implemented using RandomOverSampler Class and training datasets are selected randomly with replacement. Its main objective is to balance class spreading through the random repetition of minority target instances.

Oversampling is used for pre-processing the data in a way so that the number of minor class data become at least 50%.

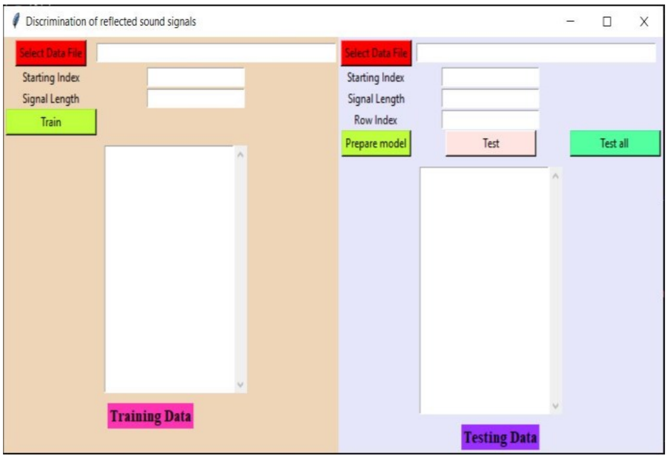

Fig. 3. Grafical user interface of discrimination of reflected sound signals.

TABLE I. CLASS TARGET DISTRIBUTION AFTER OVERSAMPLING METHOD

State	Class 0	Class 1
Original Data	1150	315
After Oversampling Method	1150	575

For feature extraction, sample rate 22050 is taken and using the librosa.feature.mfcc of MFCC coefficient ,40 feature vectors in Fig. 4 thus greatly removing the amount of calculation.

After extracting features, a large amount of complex values are found which may lead to less accurate result and also increase the complexity of the machine. Many machine learning algorithm perform better when numerical input variables are scaled to a standard range and this is called feature scalling.[8] A value is standardized as follows: y = (x – mean) / standard_deviation where mean is calculated as mean=sum(x) / count(x) and the standard_deviation is calculated as standard_deviation = sqrt(sum((x – mean)^2) / count(x)).Standardization can be achieved using the scikit-learn library and standardize is used for the data using the scikit-learn object StandardScaler. Fig. 5 shows how the standardization effect on data.

With the feature scaled data as the input have been later

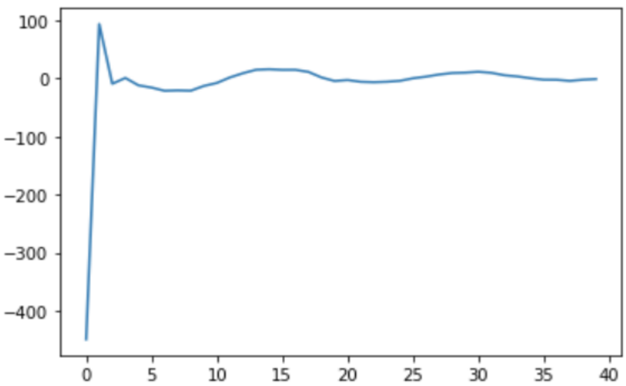

Fig. 4. Mel frequency cepstral coefficients of first audio.

	0	1	2	3	4
0	-448.983001	93.270492	-9.513580	0.499695	-12.208366
1	-421.785270	103.923287	-25.749424	5.594627	-7.550054
2	-440.907435	104.951627	-4.362396	4.410302	-14.791782
3	-434.229532	106.931547	-5.603647	0.986572	-11.855906
4	-472.460745	75.232255	-13.047862	12.218632	-7.651166

Before Feature Scaling

	0	1	2	3	4
0	-0.502617	0.401376	0.955999	1.211350	0.566761
1	-1.006163	-1.160768	1.007307	0.835870	0.347900
2	-0.322701	0.212380	-0.393886	0.513937	0.638852
3	-0.763162	0.236119	-0.408360	-0.567177	0.670501
4	1.184691	1.720789	-1.310005	-0.579501	-0.834683

After Feature Scaling

Fig. 5. How feature scaling affected the data.

used in SVM classifier and kernel chose for the SVM was RBF kernel. Moreover, RBF kernel give better accuracy hence RBF kernel is selected. The dataset is split in 70%-30% for training and testing respectively. The accuracy using RBF kernel is 90.92% . The results acquired are given below in Fig. 6 and Table II in the way of confusion matrix.

In Table II the confusion matrix shows 518 instances from two classes used for testing out of 369 correctly classified and 149 are incorrectly classified giving the 90.92% accuracy. The confusion matrix's results are used to measure different classification-related metrics. From Table II the true positives are when the input sample belongs to class and correctly classified false positives are when an input sample doesn't belong to class but it is classified as belonging to that class, true negatives are when an input sample doesn't belong to the class and it is correctly classified as not belonging to that class, false negatives are when input sample belongs to class but it is classified as not belonging to that class).

Various Classification metrics such as precision, recall, are calculated and given in Table III and Fig. 7 , the terms are defined as follows-

1. Precision: It is defined as the ratio of no. of true positives to the addition number of true positives and false positives.

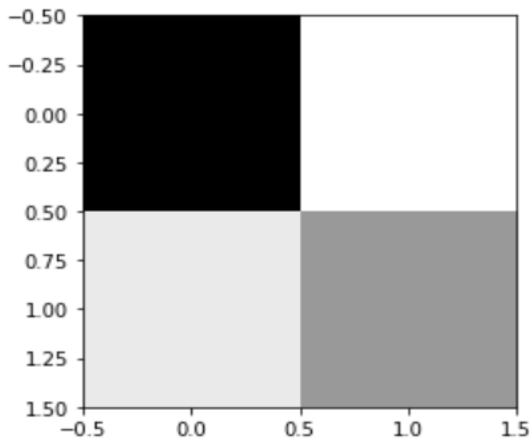

Fig. 6. Confusion matrix for linear predictor with radial basis function.

TABLE II. CLASSIFICATION RESULTS FOR TWO DIFFERENT CLASSES

	Class-0	Class-1
Class-0	332	10
Class-1	37	139

2. True Positive Rate(TPR)/Recall: It is defined as the ratio of no. of true positives to the addition number of true positives and false negatives.
3. False Positive Rate: It is defined as the ratio of number of false positives to the total number of input samples that do not belong to that class.
4. False Discovery Rate(FDR):It is the number of false discoveries in an experiment divided by total number of discoveries in that experiment. A test that crosses the approval level is referred to as a "discovery."
5. Negative Predictive Value(NPV): True negatives / True negatives + False negatives. The negative predictive value estimates the probability that a negative evaluation would result in a true negative.
6. F1-score: It is calculated as follows-

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall}$$

To see the performance of classifier the Receiver Operating Characteristics (ROC) curve is plotted for the class metal which is shown in Fig. 8.

In Fig. 8, it can be seen the classifier is perfect because when comparing two tests, the more accurate test is the one with an ROC curve further to the top left corner of the graph, with a higher AUC(Area Under the Curve).

IV. CONCLUSION

The proposed technique was used for the recognition of the following five objects. Mel cepstral coefficients were used as extracted features for the classification of test samples in this technique. SVM technique is used for classification which lead to 90.92% of accuracy. This technique allows for a greater range of classes to be worked on, as well as more accurate results.

	precision	recall	f1-score	support
0	0.90	0.97	0.93	342
1	0.93	0.79	0.86	176
accuracy			0.91	518
macro avg	0.92	0.88	0.89	518
weighted avg	0.91	0.91	0.91	518

Fig. 7. Classification Metrics.

TABLE III. VARIOUS METRICS CALCULATED FROM CONFUSION METRICS

Parameter	**Value**
True Positive Rate(TPR)	0.7898
True Negative Rate(TNR)	0.9708
False Discovery Rate(FDR)	0.0671
Negative Predictive Value(NPV)	0.8998

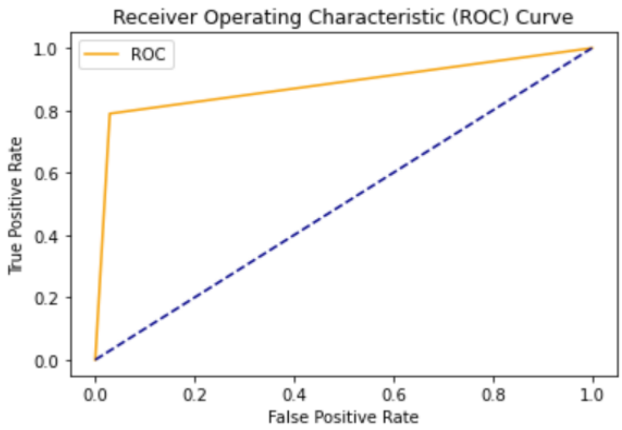

Fig. 8. Graph showing the performance of a classification model at all classification thresholds.

Expansion of audio groups and research into content-based audio segmentation are among the projects planned for the future.

REFERENCES

[1] A. C. Braun, U. Weidner and S. Hinz, "Support vector machines, import vector machines and relevance vector machines for hyperspectral classification — A comparison," 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 2011, pp. 1-4, doi: 10.1109/WHISPERS.2011.6080861.

[2] P. Rai, V. Golchha, A. Srivastava, G. Vyas and S. Mishra, "An automatic classification of bird species using audio feature extraction and support vector machines," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7823241.

[3] Bo Liu, Zhi-Feng Hao and Xiao-Wei Yang, "Nesting support vector machinte for muti-classification [machinte read machine]," *2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 2005, pp. 4220-4225 Vol. 7, doi: 10.1109/ICMLC.2005.1527678.

[4] P. M. B. Mahale, A. Sayadiyan and K. Faez, "Mixed Type Audio Classification using Sinusoidal Parameters," 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 2008, pp. 1-5, doi: 10.1109/ICTTA.2008.4530061.

[5] Y. Tsai, Chin-Chien Tsai and Kun-Ching Wang, "Automatic content-based classification of MP3 objects using radial basis function network in surveillance system," 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, Yantai, China, 2010, pp. 1506-1510, doi: 10.1109/FSKD.2010.5569329.

[6] S. Sharma, P. Fulzele and I. Sreedevi, "Novel hybrid model for music genre classification based on support vector machine," 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2018, pp. 396, doi: 10.1109/ISCAIE.2018.8405505.

[7] P. Rai, V. Golchha, A. Srivastava, G. Vyas and S. Mishra, "An automatic classification of bird species using audio feature extraction and support vector machines," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 2, doi: 10.1109/INVENTIVE.2016.7823241.

[8] A. Bhandari, "Feature scaling," Analyticsvidhya.com, 03-Apr-2020. [Online].Available: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/.

[9] G. Mali and S. P. Mahajan, "Audio Genre Classification Employing Support Vector Machine," 2018 4th International Conference for Convergence in Technology (I2CT), Mangalore, India, 2018, pp. 1-5, doi: 10.1109/I2CT42659.2018.9058184.

[10] L. Lu, H.-J. Zhang, and S. Z. Li, "Content-based audio classification and segmentation by using support vector machines," Multimed. Syst., vol. 8, no. 6, pp. 482, doi: https://doi.org/10.1007/s00530-002-0065-0

Name	Matriculation Number	Contribution on Project
Mou Saha	1327931	Research, Feature extraction, Feature Scaling and Documentation
Riyad-Ul- Islam	1324662	Research, Graphical User Interface, Data Preprocessing and Documentation
Tanvir Hasan	1323951	Research, SVM Classification, Model Analysis, Documentation