

Signal Classifier for Imbalanced Classes using MATLAB

Chevon De Souza
Student (M.E. VLSI & Embedded Systems)
Sinhgad Institute of Technology and Science,
Narhe, Pune

Dr. S. N. Mali
Principal
Sinhgad Institute of Technology and Science,
Narhe, Pune

Abstract—Imbalanced classes have always been a problem in the classification process. This is because in real time applications it is almost impossible to find an effective classifier for an imbalanced class. The system that is proposed in this paper is developed using the MATLAB software and can be implemented on an FPGA board. This includes extraction of musical features from audio files. This is done using Radial Basis Function and Euclidean classifier. A database table is also created internally which can hold ‘n’ number of signals and can be expanded as per requirement. This will help improve the classification of the signals from imbalanced class thus making it an effective system.

Index Terms—Imbalanced Classes, Euclidean classifier, Field Programmable Gate Array (FPGA), Radial Basis Function (RBF).

I. INTRODUCTION

There has been a noticeable increase in data now-a-days owing to the surveillance systems, security as well as online servers. Solving the problems faced by imbalanced data is thus the need of the hour. This data explosion has made it critical to thus advance and widen our understanding of the fundamentals of knowledge discovery, collection and the analysis of the raw data thus collected. This understanding plays a key role in the decision-making process. In many real-world applications however, a relatively new challenge of learning from imbalanced data has been created and several experts in the academic field and industry are coming up with solutions to solve the problem of imbalanced learning. The imbalanced learning problem focuses on unbalanced, under-represented data and skewed distribution of classes and the performance of their learning algorithms. Learning algorithms for unbalanced classes is difficult owing to the inherent complexity of characteristics of imbalanced data sets. Thus, creation of learning algorithms from such data requires in-depth knowledge of the principles and algorithms of data engineering, keen observation of new understandings and skilled use of tools that enable conversion of raw data into clear information and knowledge representation. Thus, characterization of

imbalanced data classes can be identified as skewed representation of instances within certain classes over others. Due to the low rate of occurrence of rare instances, the classification rules predict small classes to be rare or merely ignored. As a result, prevalent classes are not as misclassified as test samples belonging to small classes.

Correct sample classification in small classes is high in certain applications as compared to the prevalent classes. For example, during the process of a disease diagnostic problem with the recognition goal as detecting people with rare diseases in comparison to those of the normal population. Thus, the ideal model of classification will be one which will provide the highest identification rate on the disease category [10]. Hence, we can conclude that imbalanced or the skewed class distribution problem is also called the small or rare class learning problem.

In several applications, it can be observed that data collection is highly skewed as the data of one class dominates the data from other classes. As a result of this, highly accurate balanced data classification systems lack severely in their performance on imbalanced data especially for data belonging to the minority class. Over-sampling, under-sampling and use of methods that modify the existing classification systems are used to improve the performance quality of classification on imbalanced data [8].

This paper presents a method to design, test and validate an effective classifier ensemble algorithm that can improve the classification accuracy by resolving the class imbalance problems faced by various applications having skewed distribution of data. Ensemble techniques have drawn considerable attention in recent years: a set of learning machines increases classification accuracy with respect to a single machine [13]. Such a combination improves generalization performance but the drawbacks include requirements of large amounts of memory and computation space, as well as problems with addressing a portable and real-time pattern recognition application.

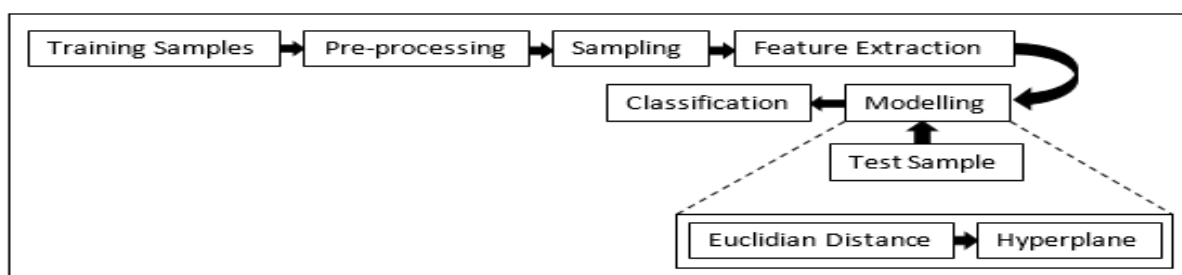


Fig.1. Block diagram of the proposed system

Fig. 1 represents the block diagram of the proposed system. The training samples are audio signals of different musical instruments. The audio varies in pitch, tone, amplitude, etc. These signals first need to be pre-processed. Pre-processing includes amplitude normalization and noise pruning as well as cleaning of the sound signal from unwanted noise. After that the sampling process is carried out. This breaks the audio signal into smaller sections. After sampling is done it is important to determine the silent samples. The silence removal is based on two important features namely signal energy and spectral centroid. A specific threshold is given to determine which sections or samples are going to be considered as the silence samples. All samples below the threshold are omitted from further processing. The next step focuses on feature extraction and is the most critical step in the classification process. Feature extraction basically checks the signal for linear as well as non-linear features. The linear feature extraction includes processes such as Zero Crossing Rate (ZCR), energy estimation, attack time, sustain time and decay time whereas Mel-Frequency Cepstral Co-efficients (MFCC) is a non-linear feature extraction process. After consideration of all the extracted features, a model is then created. This created model also includes calculation of the Euclidian distance which aids in creation of a hyperplane for classification. The final output is given after simulation of the audio signal and the different graphs can be observed in MATLAB as shown in the results below.

II. METHODS

A. Data Selection

The data selection process is the most crucial step in any classification. Here the data consists of 3 instruments namely piano, violin and the flute. The audio varies in tone, frequency and pitch. The main idea is to classify the data although it has imbalanced data in the training set. This will ensure that any sample signal later will be easily classified effectively. For this paper, training of the model was done using 5 samples of the piano, and 15 samples of violin and flute to create an imbalanced data set. The further analysis is discussed in the feature extraction section.

B. Brief Review of Different Classifiers

1. *RBF kernel*: Radial Basis Function (RBF) kernel is used in Support Vector Machine (SVM) classifications and several other classifiers. In this paper, I am using RBF kernel in association with Euclidian classifiers. Hence, two individual RBF functions will be used with unique sigma values. Following this, the weighted mean is then calculated enabling the classification of the samples into their respective individual classes. The Radial Basis Function (RBF) is a real-valued function. The RBF value is primarily dependent on the distance from the origin and secondarily dependent on the distance from the centre, labelled as point c. Thus, a function satisfying either of these cases of the property is called a radial function. The standard norm is mainly the Euclidean distance but other distance functions can also be utilised. The given functions are then approximated with the help of the sums of radial basis

functions. This process of approximation creates a perfectly linked simple neural network.

2. *Euclidian Classifier*: The Euclidean classifier is also known as the Euclidean distance classifier. This classifier is used to measure the distance between a set of samples in the N-dimensional feature space. This classifier is used especially in K-nearest neighbours. One of the standout uses of this classifier is visualised in both the hierarchical clustering and agglomerative clustering which helps in the calculation of the distance between the clusters.

C. Feature-Extraction Methods

The feature extraction method consists of two key features- the linear features and the non-linear features. Both these features provide us temporal as well as perpetual features which enable in creation of feature vectors. There are several linear feature extraction processes like Zero Crossing Rate (ZCR), energy estimation, attack time, sustain time and decay time. Similarly, Mel-Frequency Cepstral Co-efficient (MFCC) is a non-linear feature extraction processes.

1. *ZCR*: ZCR stands for Zero-Crossing Rate which is that rate of sign-changes along a signal. It can also be defined as the measure of the number of times in a given time interval/frame, the amplitude of the speech signal passes through a value of zero. It focuses on the rate at which a signal changes from positive sign to the negative sign or vice versa. The frequency content of a signal is measured as that rate at which zero crossings occur. Using representation based on the average zero-crossing rate in a short time, rough estimates of spectral properties can be determined. Estimation of spectral properties is used rampantly in speech recognition as well as music information retrieval systems where it takes the form of lead feature to classify percussive sounds.

2. *Energy Estimation*: Signal energy estimation is computed using root mean square. The amplitude of unvoiced segments falls lower than that of the voiced segments. Amplitude variation is reflected by the shorttime energy of speech signals. Certain properties in a speech signal changes or vary over time such as severe variations or peaks in the amplitude of a signal or large variations in speech signals of fundamental frequencies within the voiced regions. Thus, we can conclude that useful information of signal features, such as its pitch, its intensity or its excitation mode as well as its vocal tract parameters which include formant frequencies can be determined using simple time domain processing techniques.

3. *Attack Time*: Attack time can be defined as the time interval between the instant or point at which a signal at the circuit's input exceeds the circuit's activation threshold as well as the time interval taken by the circuit to react to the given input to a particular degree or in a specified manner. Thus, the attack time can be seen in clippers, peak limiters, compressors, and voxes.

4. *Sustain Time*: Sustain is a parameter of musical sound over time which enables denoting that length or period of time for which a sound

remains audible before becoming silent. It is one of the four segments in an Attack Decay Sustain Release (ASDR). In ASDR, sustain is initiated when both the attack as well as the decay portions have run their course and does not end until the key is released. Thus, the level at which an envelope will remain at is determined by the sustain control which is a level control unlike the attack, decay, and release controls which are time (rate) controls. In pianos, the sustain pedal prolongs the notes even after the keys are lifted increasing the sustain time of any tone while the sustain phase in violins and other string-instruments varies according to the force with which the strings are manipulated by an external force as well as the hollowness of the instrument as in the flute.

5. *Decay Time*: Decay time is that time interval which is required for the amplitude of a vibrating system to decrease to approximately 37% (or $1/e$) of its initial value. The decay time is exponential and independent for a system with a constant rate of damping from the initial amplitude.

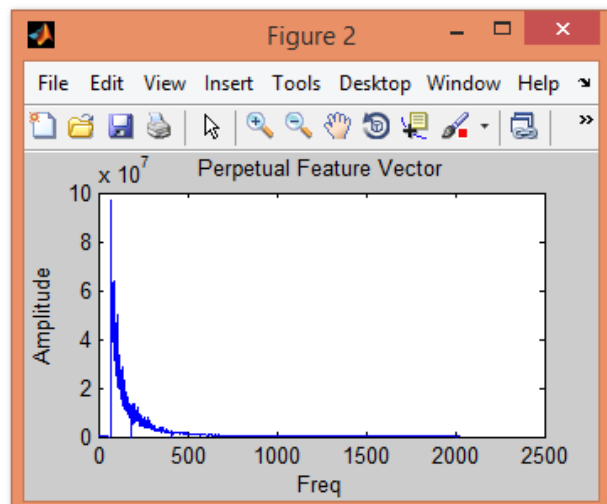
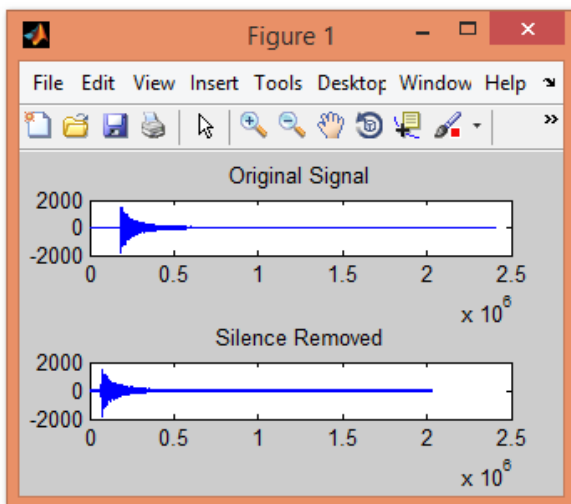
6. *MFCC*: MFCC stands for Mel-Frequency Cepstrum Coefficient (MFCC). MFC's are individual coefficients that collectively form a network of MFCC's. The M in MFC denotes melody which represents the short-term power spectrum of a sound. This spectrum can be based on either a linear cosine transform of a log power spectrum or a non-linear mel scale of frequency. MFCC's are derived from a type of cepstral representation of the audio clip. The frequency warping in MFCC allows better representation of sound i.e. in audio compression. MFCCs are commonly used as features in speech recognition systems as well as in music information retrieval applications such as genre classification.

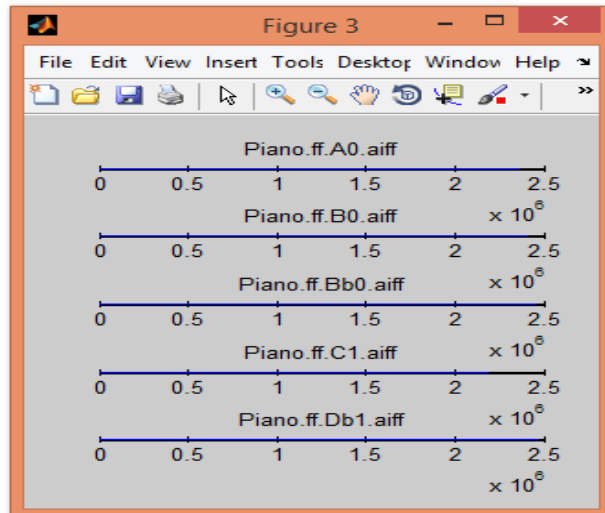
D. DataAnalysis

Once the feature vectors are created, they are plotted in a multi-dimensional space. A hyperplane needs to be created to separate the imbalanced data received. This hyperplane is calculated with the help of the Euclidean distance. Origin shifting helps in creating the hyperplane. Using Radial Basis Function Kernel as well as Back Propagating Neural Networks, the hyperplane is created which calculates the gradient of a loss function with respect to all the weights in the network. It is used as a supervised learning technique in most cases. After this is complete, the retrieval result of each class is compared and the percentage of retrieval is calculated. The final result of classification is based on this retrieval result.

III. RESULTS

The simulation in MATLAB clearly shows all the processing as well as the vector features and classification of the training samples. Figure 1 shows the original signal as well as the same signal with the silence removed. The silence removed signal helps to reduce computation time as the silenced samples are not processed further. Figure 2 gives a graph of the frequency versus amplitude of the perpetual feature vector that is obtained as a by-product of the MFCC feature extraction. Figure 3 shows a clear comparison of the different training samples that match the sample being processed. The last figure shows the values of the temporal extracted features, the perpetual extracted features and the audio retrieval results clearly classifying the sample signal as that of a piano audio file.





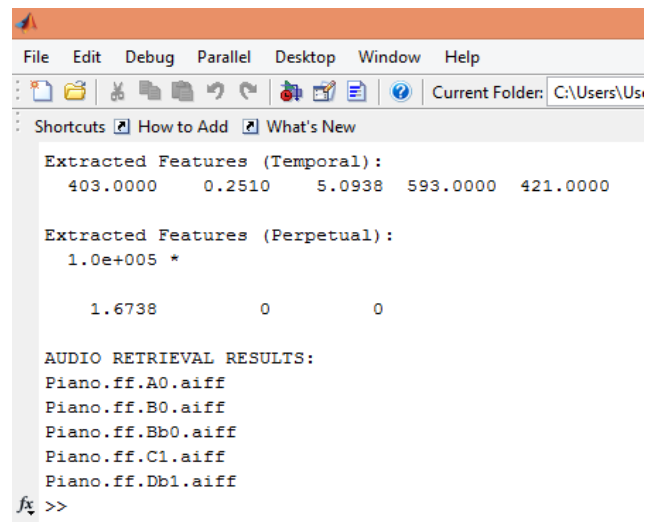
IV. DISCUSSION

Based on the results of the present paper and problems in classification of imbalanced classes, I would like to highlight the following points.

1. The audio retrieval gives the weighted mean of the two classifiers thus working as a voting system.
2. With the silence removal technique, the computation time is greatly reduced.
3. After considering various classifiers, the RBF kernel usage along with the Euclidian classifier gives a better result as compared to using only the SVM method.
4. The classification is almost 100% accurate and yields the desirable results despite the imbalanced data set.

REFERENCES

- [1] D. S. Shete, Prof. S. B. Patil and Prof. S.B. Patil, "Zero crossing rate and Energy of the Speech Signal of Devanagari Script", IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 1, Ver. I (Jan. 2014), PP 01-05 e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197.
- [2] Mehdi Neggazi, Messaoud Bengherabiand Z. Boulkenafet, "An Efficient FPGA Implementation of Gaussian Mixture Models Based Classifier: Application to Face Recognition", The 8th International Workshop on Systems, Signal Processing and their Applications, 2013.
- [3] Y. Xu, X. Cao, and H. Qiao, "An efficient tree classifier ensemble-based approach for pedestrian detection," IEEE Transactions on System, Man, Cybernetics-Part B: Cybernetics, vol. 41, no. 1, pp. 107 to 117, Feb. 2011.
- [4] Z. Salcic and W. Abdulla, "Speech recognition system for embedded real time applications", IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pages 118 to 122, 2009.
- [5] Haibo He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263 to 1284, Sep. 2009.
- [6] Y. Sun, A. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, no. 4, pp. 687 to 719, 2009.



V. CONCLUSION

The multiclass Euclidean classifier along with RBF kernel showed great performances and maps the vector features to a multi-dimensional space. Besides this, the audio retrieval gives 100% efficiency in classification. To maximize the usage of target class data, an efficient implementation is carried out. Thus this paper presents a method to design, test and validate an effective classifier ensemble algorithm that can improve the classification accuracy by resolving the class imbalance problems faced by various applications having skewed distribution of data.