Robust Feature Selection Method for Music Classification

P.RAMESHKUMAR[1], School of Computing, SASTRA University, Thanjavur-613401. rameshkumar6410@gmail.com.

M.MONISHA[2], School of Computing, SASTRA University, Thanjavur-613401. monisha2990@gmail.com.

B.SANTHI[3], School of Computing, SASTRA University, Thanjavur-613401. shanthi@cse.sastra.edu.

T.VIGNESHWARAN[4], Department of Computer Science, SHIVANI Engineering College, Trichy-620009. 5075vicky@gmail.com.

Abstract: **Human latent of distinguish varied music natures and cluster those into classes of categories are so incredible which expert in music will achieve such categorisation using their logical judgement and hearing senses. Till now, the technical society have concerned in delve into computerize the human way of distinguish the music in view of each necessary factor of the music tune, songs from voice of the artists are all based on the instrument types. The outcomes of those works up to now have been significant however still it can be enhanced. The main intend of this paper is to categorize the western and carnatic songs process automation by extracting the selected features which gives more accuracy of categorization. Inspite of the several features of audio signals (Method-1) the proposed method identifies the important robust features (Method-2) (RMS, Low Energy, Rolloff, Brightness, Pitch, Inharmonicity, Flatness and Entropy) with respect to carnatic/western music categorization. The identified features for carnatic/western music classification were tested using several classifiers and their accuracy of classification was discussed.**

Keywords: Music categorization, Music classification, Feature selection and Feature extraction.

1. INTRODUCTION

Signal processing deals with operation or analysis of signals (analog or digital). Its operations include signal acquisition and reconstruction, quality improvement, signal compression, features extraction etc [4]. Groups of processing the signals consist of analog, discrete time and digital signal processing. In this paper we have discussed about the musical signal processing. Music signal processing methods used for synthesis, coding/decoding content-indexing and composition of music signals. Music signal processing has been divided into two branches: music content reaction and music signal modelling [7].

Best mechanism to represent audio signals is used to symbolize low-level sound characteristics for recognizing, relating and distinguishing meticulous sounds. Audio signals are identified using the genre of music. Classifying the music genre manually is time consuming and labour task. So many researchers have moved their work towards the automatic music genre classification task, representing the music signals based on spectral characteristics, is extracted. Generally, the computerized classification task of music genre includes two stages: extraction of features and design of classifier [5]. Here, in this paper these music genres are classified based on some of the features as shown below.

1.1 Audio Features Taxonomy

Taxonomy is an association of entities based on various principles. There is unambiguous, single and generally applicable taxonomy of AF as it is manifold in nature. The taxonomy follows a method-oriented loom that reveals the interior structure of diverse features and their similarities also it facilitates the assortment of features for a meticulous task [14]. Assortment of features is determined by some factors such as computational constraints for example feature withdrawal on devices with restricted capabilities or semantic issues such as features describing rhythm. The domain a feature allows to roughly estimating the computationally complex nature [2]. The structure of temporal domain based on 3 distinguished groups of features: amplitude-based, power-based and zero crossing-based features.

For the frequency domain, a semantic layer that divides the set of features into two distinct groups: Perceptual and Physical features. Perceptual features are the meaningful aspects of sounds that include brightness, chroma, harmonicity, loudness, pitch, and tonality, whereas physical features are grouped based on their extraction process. They are distinguished based on features such as auto regression, Short-Time Fourier Transform, Adaptive time-frequency decomposition [18]. Feature that is derived from Short-Time Fourier Transform that can be further alienated into features which take the complex part into description and features that function on the real part of the spectrum.

978-1-4799-2352-6/14/$31.00 ©2014 IEEE

SVM fit into broad class of kernel methods, which is an algorithm based on the information that are only dot products. Usually, SVM are an example of a linear two-class classifier. The information for two-class knowledge crisis consists of items labeled with either one of those two labels similar to the two classes; For example the labels are -1 (negatives) or +1 (positives) are considered. Neural networks were primarily urbanized based on the basic principle of the process of the neural (human) system. From then an extremely huge diversity of various network designs which have been established. Each and every one is comprises of (neurons) units, and links among those, which collectively resolves the network performance. The option of the network nature based on the solving the problem.

2. RELATED WORKS

Noor Azilah Draman et.al has proposed the method of modified AIS-based classifier to resolve the problems in the classification of music genre area where concentrate on censoring and monitoring modules and the feature classification, feature extraction and feature selection were discussed [12]. Mehdi Banitalebi-Dehkordi and Amin Banitalebi-Dehkordi discussed a sparse FFT based feature extraction method which combines long-term features and short-term features. Their method of extracting the features is better than the other compressive sampling based classifiers [10].

Geoffory Peeters provided the method for rhythm classification based on periodicity representations. In this the use of periodicity, shared with rhythm set features, tempo information and comparison is carried out using estimated and annotated tempo [3]. Pasi Saari et.al work shows the classification in the field of MIR can be enhanced by wrapper selection when the functions were estimated by considering the generalizability and simplicity of the formed models. Their result shows that simple classifiers perform well in comparison to the more composite SVM classifier [13]. Shin-cheol Lim et.al recommended the novel music-genre categorization system that consists of spectrotemporal features basis on the timber features, SVM ranker for selecting feature and RBF kernel evaluation for SVM classifier. Their system affords greater classification precision for both GTZAN and ISMIR2004 databases and the proposed method have need of lower dimensions [16].

George Tzanetakis et.al identified the sets of features for demonstrating music surface and rhythm information employ to build computerized genre classification algorithms and they tested the performance by training numerical pattern recognition classifiers on data sets of real-world [6].

3. PROPOSED METHOD

In the previous section, we introduce some of the basic concepts regarding music, its signals, and some features of audio signals. Here there is a brief discussion about our proposed features [6]. Some of the features that have taken for the comparison are Root Mean Square Energy (RMS), Rolloff, Pitch, Inharmonicity, Flatness, Entropy, Low Energy, and Brightness.

Root Mean Square Energy: Computation of the total energy of the signal x can be simplified by taking the average root for the square of the amplitude, basically called as Root-Mean Square Energy (RMS).

Low Energy: To get the estimation for distribution of energy in temporal by using energy curve, to see either it remains the same throughout the signal, or if frames are more constraint than others. One way is computing the Lower Energy rate, i.e., the percentage of frames showing less than average energy.

Rolloff: One way to approximate high frequency amount in the signal which consists of analysing the frequency such that the total energy fraction is controlled below the approximated frequency [8].

Brightness: A dual method which consists in fixing is that cut-off frequency and measuring the amount of energy above that frequency [9].

3.1 Workflow Model:

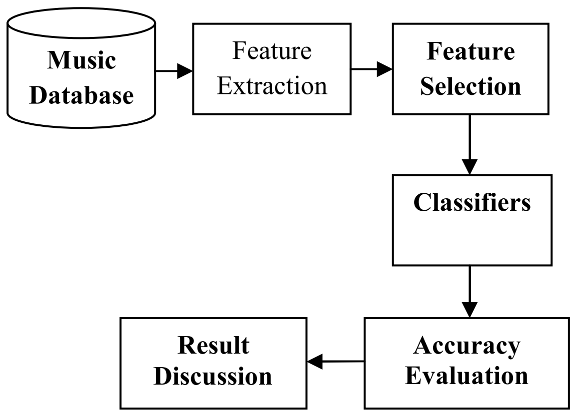

Fig.1: Proposed feature selection framework for music classification

Pitch: Extra pitches returned either as discretized note events or as continuous pitch curves.

In-Harmonicity: The quantity of partials that are not multiples of the fundamental frequencies are called as Inharmonicity which values between 0 and 1 [1].

Flatness: The flatness indicates whether the distribution is spiky or smooth and simple ratio results between the arithmetic mean and the geometric mean [15].

Entropy: Entropy returns the relative entropy of the input using the equation [11].

In our proposed method, the above features of audio signal are used for classification and compared with different classifiers and their performance is evaluated. Fig1 shows SVM and neural network framework and the procedure of our method. The procedure starts with formation of music database with several collections of Western and Carnatic music's. The next step is to extract the features form music's available in the database. After extracting the features select only the essential and critical feature which plays a major role in classification. In our paper, we have selected the features such as RMS Energy, Inharmonicity, Pitch, Entropy, Flatness, Low Energy, Brightness and Rolloff. Once the features are selected then it is given as input to classifiers (SVM, Neural Network). The classification performance such as confusion matrix for each classifier is identified and evaluated. Finally the results of each classifier are compared to get higher accuracy.

4. EXPERIMENTAL RESULTS

As shown above our experimental work starts with the collection of western and carnatic songs. In our experimental work we have collected totally 150 songs which consist of 75 western songs and 75 carnatic songs. Once the songs were collected our next step is to extract the features from the music stored in a database. Initially we have taken fourteen features which were closely related to carnatic and western music classification named as Method-1. After that we have identified eight robust features among the initially selected fourteen features (RMS, Energy, Rolloff, Pitch, Inharmonicity, Flatness, Entropy, Low Energy and Brightness) named as Method-2 (proposed method) which play a major role in categorization of music especially western or carnatic songs. All these features were extracted using MIR-Music Information Retrieval [13] Toolbox on Matlab and stored in database. The sample database for five carnatic and five western songs were as shown below in Table 1.

The next step is to perform classification or categorization of western and carnatic songs using the above extracted features on well known classifiers such as SVM classifier and Neural Networks (network pattern recognition tool-nprtool) classifier then analyzing the performance measures.

4.1 Neural Network Classifier

Neural network contains more than three neuron layers: 1 output layer, 1 input layer and as a minimum of 1 hidden layer. Generally the network which has only 1 hidden layer, as shown in Fig.2 limits time of calculation, particularly at the time of satisfactory results.

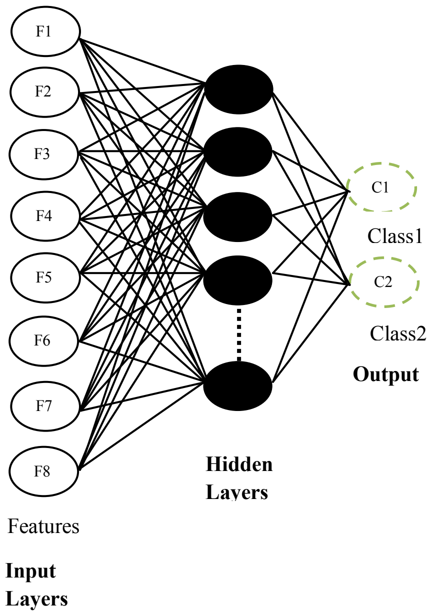

Fig.2: Architecture of Neural Network

The 150 samples with eight features(elements) each is given as inputs to the training network with two classes of targets Carnatic as 1 and western as 0 (Targets representing 150 samples of 2 elements). Each sample was associated in the columns. Out of 150 samples, 90(60%) samples were used for training and 30(20%) samples were used for validation and remaining 30(20%) samples were used for testing. Then the hidden neurons 20, is selected and then train the network.

Table 1: Sample Database with different feature values for carnatic and western songs

Features/ Songs	RMS Energy	Roll Off	Pitch	In-Harmonicity	Flatness	Entropy	Low Energy	Brightness
Carn-S1	0.60401	19186.0840	1338.8755	0.48762	0.6350800	0.93661	0.65637	0.91106
Carn-S2	0.65677	19067.6514	1382.7606	0.33812	0.6763900	0.95286	0.66667	0.91106
Carn-S3	0.089404	2896.9308	412.0690	0.48568	0.1807600	0.96527	0.53634	0.53373
West-S1	0.094135	3693.1126	741.2748	0.44849	0.0457300	0.84575	0.53264	0.32203
West-S2	0.09577	5991.6138	185.6500	0.48951	0.0979460	0.87434	0.5042	0.37937
West-S3	0.081682	6400.4082	164.3176	0.47109	0.0885710	0.88939	0.54056	0.515

Table 2: Confusion Matrix for Neural Network

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	74	1
	Western	2	73

4.2 SVM Classifier

The next evaluation of our experiment is using the SVM (Support Vector Machine Classifier). A point of a data is visualized as a *p*-dimensional hyperplane vector. The finest hyperplane is that symbolizes the major parting, or margin, among two classes. Such hyperplane, it is identified as the maximum-margin and the linear classifier defines it as maximum margin classifier; and optimal stability for the perceptron. The input data is specified on an eight column and 150 elements of each eight features in each columns. Then classify the inputs in to two groups namely Carnatic(1) and Western(0). Randomly selects the tests and train data using crossvalind function. Then train the selected data using SVMtrain function. After the training classify the datasets using SVMclassify function and then evaluate the performance and accuracy of the classifier.

Table 3: Confusion Matrix for SVM

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	74	1
	Western	3	72

The Table 3 shows the confusion matrix for SVM classifier. Matrix clearly shows that out of 75 carnatic songs 74 are identified as carnatic and 1 is misclassified as western and similarly for western songs, out of 75 western songs 73 are correctly classified as western and remaining 2 were misclassified as carnatic. Hence the overall accuracy of the SVM classifier is about 97.3%.

4.3 Logistic Regression Classifier

Logistic Regression is one of the types of statistical classification probabilistic model employed for envisage the conclusion of a variable dependent on categorical information based on one or more features. It assess the experimental values of the features in a qualitative retort model. In our case the logistic regression classifier has been used to classifiy the western and carnatic musics and finally obtained the accuracy rate of 95.33% from the confusion matrix as shown in Table 4.

Table 4: Confusion Matrix for Logistic Regression

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	71	4
	Western	3	72

4.4 K-Star Classifier

K-Star classifier is an instance based classifier that is the similarity between the test instance class and the training instance class as computed by some similarity functions. It is different from other instance based classifiers in that it utilizes the function entropy-based distance. In our experiment the overall accuracy rate obtained using K-Star Classifier is 97% from the confusion matrix as shown in Table 5.

Table 5: Confusion Matrix for K-Star

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	74	1
	Western	3	72

4.5 J48 Classifier

It generates the c4.5 decision tree in pruned or unpruned manner. These decision tree were used for classification and the c4.5 is also named as statistical classifier. The decision tree were build from the set of training data. We have used this classifier and the obtained the overall classification accuracy of 99.33% from the confusion matrix as shown in Table 6.

4.6 Adaboost Classifier

Adaboost machine learning algorithm can be used in combination with other learning algorithms in order to improving the performance.

Adaboost machine learning algorithm is high sensitive towards the outliers and noisy data.

Table 6: Confusion Matrix for J48

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	75	0
	Western	1	74

We have used the adaboost classifier for the classification of western and carnatic music classification and obtained the overall accuracy of 98% from the confusion matrix as shown in Table7.

Table 7: Confusion Matrix for Adaboost

		Predicted Class	
		Carnatic	Western
Actual Class	Carnatic	73	2
	Western	1	74

Table 8: Method-1 and Method-2 classification accuracy

Classifier	Method-1 Accuracy (in %)	Method-2 (proposed) Accuracy (in %)
SVM	92.42	97.3
Neural Network	91	98
Logistic Regression	87	95.33
K-Star	92	97
J48	94	99.33
Adaboost	91.33	98

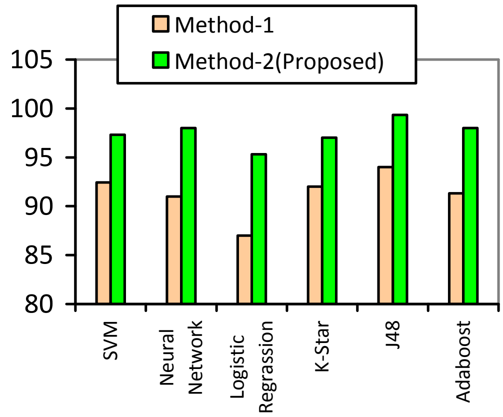

Fig.3: Comparison graph for Method-1 & Method-2 of classification accuracy

The comparison table shown in Table 8 clearly outfits our experimental work by obtaining the higher accuracy rate than the initial method(method 1). Method-1 conatins the accuracy rate of classification using initially selected fourteen features. Method-2 contains the accuracy rate of classification using selected robust features as discussed earlier. As for our categorization of music signals as western or carnatic the selected robust feature perform well in classifciation as compared to other features. From the results obtained above it is concluded that method-2 shows the better results.

5. CONCLUSION AND FUTURE WORKS

Our experimental results shows that the samples undergone for pattern recognition using Method-1(fourteen features set) & Method-2 (eight selected(robust) features from Method-1) and Confusion matrix(accuracy) is plotted for both the methods and the performance analysis shows that the Method-2(proposed method) has overall higher performance than Method-1. As a result Method-2 performs well than Method-1. Thus the extracted features such as RMS, Low Energy, Rolloff, Brightness, Pitch, Inharmonicity, Flatness and Entropy forms the thereshold in the classification of music signals as western or carnatic and the future work can be extended on automatic ragga identification, automatic artist voice identification and other computerized identification can be enhanced by considering the different features of music signals.

6. REFERENCES

[1]. G. Agostine, M. Longari, and E. Pollastri,“ Musical Instrument Timbres Classification with Spectral Features”, EURASIP Journal on Applied Signal Processing, pp.5-14, May 2003.

[2]. Dalibor Mitrovic, Matthias Zeppelzauer, and Christian Breiteneder, “Features for Content-Based Audio Retrieval”, Advances in computers,Vol 78, pp.71-150, 2010.

[3]. Geoffroy Peeters, “Spectral and Temporal Periodicity Representations of Rhythm for the Automatic Classification of Music Audio Signal “, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 5, July 2011.

[4]. Hunt, M., Lennig, M., and Mermelstein. P, “Experiments in syllable-based recognition of continuous speech”, In Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp.880-883, 1996.

[5]. Jia-Min Ren, Jyh-Shing, Roger Jang, “Discovering Time-Constrained Sequential Patterns for Music Genre Classification”, IEEE Transactions on Audio, speech, and language processing, Vol. 20, No. 4, May 2012.

[6]. Jin S.Seo , SeungjaeLee,”Higher-order moments for musical genre classification”, Elsevier, Signal Processing 91, pp.2154–2157, March 2011.

[7]. A. Klapuriand M. Davy, “Signal processing methods for music transcription”, Springer, New York, 2006.

[8]. T. Li and M. Ogihara, “Music Generic Classification with Taxonomy”, In proceedings of the IEEE International Conference on Acoustics, Speech, And Signal Processing, Volume-5, pp.197-200, March 2005.

[9]. M. Liu and C Wan, “A Study on Content based Classification and Retrieval of Audio Database”, In proceedings of International Symposium on Database Engineering and Applications, pp.339-345 IEEE Computer Society, July 2001.

[10]. Mehdi Banitalebi-Dehkordi & Amin Banitalebi-Dehkordi, “Music Genre Classification Using Spectral Analysis and Sparse Representation of the Signals”, J Sign Process System DOI 10.1007/s11265-013-0797-4, July 2013.

[11]. H. Misra , S. Ikbal, S. Sivadas and H. Bourlard, “ Multi-Resolution Spectral Entropy Feature for Robust asr”, In proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Vol-1, pp.253-256, , March 2005.

[12]. Noor Azilah Draman, Sabrina Ahmad, and Azah Kamilah Muda “Recognizing Patterns of Music Signals to Songs Classification Using Modified AIS-Based Classifier”, J.M. Zain et al. (Eds.): ICSECS 2011, Part II, CCIS 180, pp. 724–737.

[13]. Pasi Saari, Tuomas Eerola, and Olivier Lartillot, “Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music”, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 6, August 2011.

[14]. G. Pfeiffer. “The importance of perceptive adaptation sound features for audio content processing”, In proceedings SPIE Conference, Electronic Imaging, Storage and Retrieval for Image and Video Databases, VII, pp.328-337, San Jos SE, California 1999.

[15]. A.Ramalingam and S. Krishanan, “Gaussian Mixture Modelling Using Short Time Fourier Transform Features for Audio Finger Printing”, In proceedings of IEEE International Conference on Multimedia and Expo, pp.1146-1149, July 2005.

[16]. Shin-Cheol Lim, Jong-Seol Lee, Sei-Jin Jang, Soek-Pil Lee, and Moo Young Kim, ”Music-Genre Classification System based on Spectro-Temporal Features and Feature Selection”, IEEE Transactions on Consumer Electronics, Vol. 58, No. 4, November 2012.

[17]. Thair Nu Phyu, “Survey of Classification Techniques in Data Mining,”In the Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol-1 IMECS, Hong Kong, March 2009.

[18]. A. Wang, A. Divakaran, A. Vetro, S.F. Chang, and H. Sun, “Survey of compressed-domain featuresused in audio-visual indexing and analysis. Journal of Visual Communication and Image Representation”, 14920; pp.150-183, June 2003.