International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-11, September 2020

Machine Learning Algorithms for Indian Music Classification Based on Raga Framework

Kalyani C. Waghmare, Balwant A. Sonkamble

Abstract: ***The supervised and unsupervised learning methods in Machine Learning are successfully applied to solve various real time problems in different domains. The Indian Music has a base of Raga structure. The Raga is melodious framework for composition and improvisation. The identification and indexing of Raga for Indian Music data will improve efficiency and accuracy of retrieval being expected by e-learners, composers and classical music listeners. The identification of Raga in Indian Music is very difficult task for naïve user. The application of machine learning algorithms will definitely be best key idea. The paper demonstrates K-means and Agglomerative clustering methods from unsupervised learning nonetheless K Nearest Neighbor, Decision Tree and Support Vector Machine and Naïve Bayes classifiers are implemented from supervised learning. The partition of 70:30 is done for training data and testing data. Pitch Class Distribution features are extracted by identifying Pitch for every frame in an audio signal using Autocorrelation method. The comparison of above algorithms is done and observed supervised learning methods outperformed.***

Keywords: Classification, Clustering, Indian Classical Raga, Performance measures, Pitch Features.

I. INTRODUCTION

The Indian Classical Music (ICM) has very rich background. The ICM is divided in two types North Indian and South Indian music. The difference in both music is mainly because of their geographical locations, languages and pronunciation. Though they have differences but the core part of both the type is a Raga. The Raga is collection of Melodic phrases of different Swaras. The ICM has seven Swaras which are sung in 3 different ways as Sharp, flat, and soft. The combination of various types of Swaras exhibit emotions in Raga which are suitable for specific time and mood. The Ragas in ICM are distinguished based on the type of Raga, time at which Raga to be sung, a mood of Raga. The Raga could be sung in three octaves, if the first or lower octave starts at 110Hz then middle octave starts at 220Hz and upper octave starts at 440Hz.

Revised Manuscript Received on September 30, 2020.
* Correspondence Author
Kalyani C. Waghmare*, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India. E-mail: kcwaghmare@pict.edu.
Balwant A. Sonkamble, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India. E-mail: basonkamble@pict.edu.

In ICM the freedom is given to artist for improvising the performance by systematically following the rules of Raga. Due to this the performances of same Raga by two different signer or even same singer may vary.

The Raga can be identified based on different parameters such as number of notes and which notes used in ascending and descending sequence of notes, catchy phrases sung and movement of notes. Pitch based features are used for exploring patterns to identify Raga. Based on these Pitch features Raga characteristics such as Vadi, Samvadi Swar, Aaroh Avaroh patterns and pakad of Raga are identified, and are provided to different classifier. The experts try to identify repeating and catchy phrases in musical excerpt. These phrases reveals the parameters of Raga which are useful for Raga Recognition. The correct Note identification is main task in this whole process, which is quite difficult. So researchers tried with other low level features such as Zero Crossing Rate, Cepstral features, brightness, Mel Frequency Cepstral Features (MFCC). The correct Raga identification task from musical excerpt is very complicated task for untrained person. So the application of data mining techniques for Raga identification is the best key idea. Data mining techniques made it easier to build model from extracted features and identify hidden patterns in various domains like medical science for predicting diseases, weather forecasting, and agriculture for crop yield prediction or crop disease identification and so on. The paper is organized as follows: Section II describes the relevant literature for Raga identification using Data Mining techniques. Section III briefs about the work done proposed feature extraction algorithm. Section IV gives details of experimental results and its analysis. Section V Conclusion.

II. RELATED WORK

The Raga identification is mainly performed using its fundamental parameters such as Vadi –Samvadi Swar, Aaroh-Avaroh, Time, chalan and Gamak. To extract these parameters Pitch features need to be calculated. In [1] researchers extracted Chroma features using Music Information Retrieval (MIR) Toolbox in Matlab[2]. The dataset of four Raga from North Indian Music is taken into consideration for experimentation. The Raga parameters are identified based on Chroma features. The machine learning classifiers such as Random forest, C4.5, Bayesian network and K-star are used from WEKA tool [3]. The average accuracy observed is greater than 85 % for all classifiers.

In all classifiers K-star classifier is outperformed. The melodic framework of Raga can be obtained by using Pitch values. These Pitch values used by the researcher's different ways such as Pitch Class Distributions, constructing statistical models to identify the Raga of a given input. The lots of work done in Raga recognition using Pitch values is based on Pitch Distribution. In [4, 5, 6] authors created variants of pitch-class profile distribution based on bin weighting and note stabilization. In [4] author used tree based classifier and K-Nearest Neighbor classifier witch PCD and Pitch Class Dyad Distribution (PCDD). The Cross Validation (CV) and Holdout methods are used to analyze the results of PCD and PCDD. In CV method PCDD have better than PCD but in Holdout method the reversed results are observed. Author decided to develop the more robust methods for analyzing melodies with gliding tones. In this approach the loss of temporal properties in melodic phrases is observed, which are important for Raga Recognition. In [5] authors suggested to increase the number of bins in pitch distribution to smooth out the minor fluctuations and exploit gamak effectively. The variations of KNN classifier with Kullback-Leibler (KL) distance measure is used for classification. The overall accuracy with Pitch distribution achieved is 76.5% but increase in number of bins in the pitch distribution not proved to be much useful.

As extension to [4] authors developed new Raga Recognition system in [6]. In this work Pitch Feature extraction is done by Sawtooth Waveform Inspired Pitch Estimator (SWIPE) algorithm and found better result as compared to HPS algorithm [7] and algorithm in YIN [8]. After Pitch tracking the distribution of Pitch is done by three ways, such as a twelve-dimensional PCD, 120 or 240 bin Fine grain Pitch Distribution (FPD), and a kernel-density pitch distribution (KPD) based on Kernel Density Estimation (KDE) function. KNN classifier is applied on three distribution with different distance measures namely Euclidean, Bhattacharya, L3, City Block. It is observed the combination of KPD with Bhattacharya distance KNN worked better results. To overcome the limitations of Pitch Distribution by modeling Pitch sequence information, the use of n-gram model or Markov Model is proposed by researchers. In [9] Pitch contour is extracted, for manually marked motifs. Hidden Markov Model (HMM) is defined for each motif. The authors observed that the motifs are identified correctly, if the phrases are long enough. The authors observed that the significant understanding of Rhythm and acoustics is necessary for extracting melodic motifs.

In [10] the results are improved by combining the concept of n-gram model and Pitch Distribution. KL distance measure is used for comparing two Pitch distribution. 4-gram histogram of notes is created. A multiclass SVM classifier is used by combining two linear kernels. The overall accuracy is improved than previous approaches. In [11, 12] phrase detection based Raga identification is done. In this approach the main moto was to identify phrases and then to apply Pitch distribution or similarity measure. In [11] Authors performed automatic identification of melodic phrases based on the Time Delayed Melody Surface (TDMS) which is considering the concept of delay coordinates from Time series analysis. After identifying the TDMS, the KNN classifier is applied on it with different distance measures. The concept of Topic Modeling in text classification using vector space model is utilized in [12]. In this work all possible 2sec. segments are considered as candidate patterns in every sample. First, unique Intra-recording patterns are identified and then from whole dataset unique patterns are identified. The vocabulary of pattern is constructed. Each pattern is similar to term in the text information retrieval. Term Frequency Inverse Document Frequency (TFIDF) is applied. The features obtained after TFIDF are used to train classifier. The KNN, Naïve Bayes, SVM, Random Forest classifiers are implemented with 12 fold cross- validation method. The Naïve Bayes classifier performed well and author observed phrased based Raga recognition outperforms. The unsupervised learning approach is used in [13, 14] with MFCC features for Raga Recognition. The combination of MFCC and Chroma features is done and achieved better results. But the limitation of this work is, the algorithm is not implemented on any standard datasets.

III. WORK DONE

The untrained people usually compare two tunes based on note variations. This is very difficult for untrained people even trained person cannot recognize it 100% correctly. Data Mining is multidisciplinary field. It construct the model from extracted features and provides encouraging results. In this paper we performed experimentation by considering 8 Ragas namely Asavari, Bageshree, Bhairavi, Bhoopali, Darbari Kanada, Malkauns, Vrindavani Sarang, and Yaman. The recording of 1400 samples each of around 5-6min length is done in sound proof room with Harmonium and Tabla. The samples are stored in .wav format with sampling frequency 44100Hz, and 16bps, in Monochannel. Every sample is divided into frame of 20ms with 25% overlapping. This dataset is self-generated dataset for research work. The Pitch features are extracted using autocorrelation. The singer sings the same song with different tonic. It is necessary to normalize the values with respect to the same Tonic. The Tonic pitch is the Pitch value of the full signal which is considered as Middle octave ‘Sa’. The remaining values are calculated as per the ratios in [15]. All the pitch values normalized with Tonic Pitch 220 Hz. The Pitch values distribution is done with 12 bins per octave. First octave is from 110Hz to 220 Hz, Second 220Hz to 440 Hz, and 440Hz to 880 Hz. The occurrence of Pitch values of each sample is stored in 36 bins PCD. While considering Pitch features, the PCD of these samples are taken as feature vector. To design the reference model of each Raga 70% samples are considered in training set and 30% in testing set. K-means clustering, Agglomerative clustering, K-nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM) and Naïve Bayes algorithms are implemented by considering PCD feature vectors.

A. K-means Clustering

K-means clustering algorithm is partition based algorithm. K-means is very simple to understand and easily adaptable to new examples.

The limitation of K-means is choosing 'K' value and centroid values manually. K-means clustering algorithm is sensitive to initial value of 'K'. The complexity of K-means is O(nkl), where n is number of data points, k is number of clusters, l number of iterations.

Consider S is the training dataset,

$S = (x_i)$ where i=1,2,…., N training samples.

x_i is d-dimensional feature vector

x_i= PCD = {FC1,FC2,…..FCd}

Cen={c1,c2,…,c8}

Cen is centroid values for eight clusters

The steps followed in K-means algorithm are as follows

Step1: Calculate Euclidean distance of each data point with every cluster.

$$\forall_{i=1..N} \ \forall_{j=1..8} \ d[i,j] = dist(xi,xj)$$

Step2: Allocate datapoint to cluster with minimum distance

$$\forall_{i=1..N} \ clust[i] = index(min(d(xi,:)))$$

Step 3: Modify centroid value of all clusters by taking average of all datapoints.

Step 4: Repeat step 1, 2, 3 till changes occur in cluster.

B. Agglomerative Hierarchical Clustering (AHC)

Agglomerative clustering is hierarchical clustering algorithm which uses bottom-up approach. Initially each data point is considered as single cluster. At each iteration algorithm merges the most similar clusters, until one cluster or 'K' cluster are formed. The complexity of Agglomerative Hierarchical Clustering algorithm is $O(n^3)$ where n is number of data points. In AHC algorithm complexity may be reduced by storing distance in Min-Heap data structure. Consider 'S' training dataset in K-means and follow the steps given below to find 'k' clusters

Step1: Calculate Euclidean distance of each data point with other.

$$\forall_{i=1..N} \ \forall_{j=1..N} \ D[i,j] = dist(xi,xj)$$

Step 2: Find the minimal value Di,j ,and merge clusters i,j.

Step 3: Transform d, by substituting row and column of i, j with iUj.

$$D = D - \{D_{i},:, D_{j},:, D:,I, D:,j\} \cup \{D_{iUj},:\}$$

Step 4: If the number of clusters or entries in D obtained larger than 'k', go to step 1, else end.

C. Decision Tree classifier

Decision tree algorithm is recursive algorithm. Starting from root node partitioning is done based on one feature recursively. The selection of attribute for partitioning is performed on the basis of different measures such as Entropy, Gini index, and Information Gain. Tree algorithm uses greedy approach to train the tree model. It is easy to understand and designing the rules to identify patterns. Nonetheless takes relatively more time to train the model and calculations becomes more complex for more class labels and number of features. If the feature dimension is huge then pre-pruning or post-pruning of tree needs to be performed. The complexity of Decision Tree algorithm is O (mn*log(n)), where n is number of tuple and m is number of number of features.

Consider S training dataset from KNN. Apply following steps to train the model.

A = {A0, A1,…..Ad} is set of attributes for splitting.

SA= Φ, set of attributes selected for splitting.

Step 1: Select attribute FCi for splitting the dataset 'S' at level 'l' by calculating Gini index

$$Gini(S) = 1 - \sum_{j=1}^{m} P_j^2$$

Where m =8, no. of class labels, Pj is probability of tuples belong to 'j' class.

$$P_j = \frac{|\{t \in S : t|y| = y_j\}|}{|S|}$$

Where |S| is the total no. of records in S.

Let ai ∈ A and ∄ (SA) be the attribute, to partition S into S1 and S2

$$Gini_{ai}(S) = \frac{|S_1|}{|S|} \times Gini(S_1) - \frac{|S_2|}{|S|} \times Gini(S_2) \qquad \forall_{ai \in A}$$

$$\Delta Gini(ai) = Gini(S) - Gini_{ai}(S) \qquad \forall_{ai \in A}$$

$$Splitattribute = ai \qquad \forall_{A\, ai = max(\Delta Gini(ai))}$$

$$SA = SA \cup \{ai\}$$

$$S = S - \{t \in S : \Delta Gini(ai) = 0\}$$

Step2: Split S into S1 and S2 on attribute ai

$$S_1 = \{t \in S : t(ai) \leq x\} \ \& \ S_2 = \{t \in S : t(ai) > x\}$$

Step 3: Repeat step 2, 3 till |SA| = |A|

D. Support Vector Machine (SVM) classifier

SVM constructs a separating hyperplane, which maximizes the margin between the data points belong to two classes. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are pushed up against the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the lower the generalization error of the classifier.

The equation of hyperplane for binary classifier is

$f(x) = x_i\beta + b$

Where β ∈ Rd, b is real no., $x_i \in R^d$

The margin will be

max(2/‖ β‖) subject to $x_i\beta + b \geq 1$ if $y_i = 1$

max(2/‖ β‖) subject to $x_i\beta + b \leq -1$ if $y_i = -1$

In SVM Multiclass classifier using one-vs-one k*(k-1)/2 classifiers are trained on data from two classes. If xt is testing tuple then yt will be

$y_t = \forall_{j=1..k*(k-1)/2} \max(vote[x_{tj}])$

$Vote\ [x_{tj}] = \forall_{j=1..k*(k-1)/2} Count(y_j = y_t)$

E. K-Nearest Neighbor (KNN) classifier

KNN is easy to understand and implement. It doesn't required to evaluate any parameter from given data so it is defined as non-parametric method. The main problem in KNN is selection of 'K' value. Usually, researchers run the KNN algorithm several times with different values of K and choose the K that reduces the number of errors.

Let S be the training dataset,

$S = (x_i, y_i)$ where i=1,2,…., N training samples.

x_i is d-dimensional feature vector

x_i= PCD = {FC_1,FC_2,…..FC_d}

y_i ∈ {1, 2, 3, 4, 5, 6, 7, 8}

The steps followed in KNN algorithm are as follows

Step1: Calculate Euclidean distance of each data point with other.

$$\forall_{i=1..N} \forall_{j=1..N} \; d[i,j] = dist(x_i, x_j) \quad \text{where } i \neq j$$

Step2: $D = \langle\langle x_{11}, x_{12}, \ldots, x_{1N}\rangle, \langle x_{21}, x_{22}, \ldots, x_{2N}\rangle, \ldots \langle x_{N1}, x_{N2}, \ldots, x_{NN}\rangle\rangle$

Step3: Sort distance of every data-point in ascending order

$Ds = \langle\langle x_{ij}, x_{ij+1}, \ldots, x_{iN}\rangle\rangle$

$\forall_{i=1..N} \forall_{j=1..N} \; x_{ij} \leq x_{ij+1}$

Step4: Select labels of first 'K' entries from sorted distance D.

Step 5: Return mode of 'K' labels

Step 6: Repeat step 5, 6 for K=1..sqrt(N) .

IV. EXPERIMENTAL RESULT

The Data considered for experimentation is generated for 08 different Raga. 25 samples sung by each singer of every Raga are recorded in a vacant room with Tanpura as drone and Tabla for rhythm. The experimentation is done on samples recorded by six singers. The file is stored in .wav format with sampling frequency 44100Hz, and 16bps. The frame size is considered as 20ms with 25% overlapping.

Initially, Autocorrelation algorithms used for getting Pitch features. The distribution of these values is done in PCD of 36 bins. The Partitioned based K-means clustering algorithm and Hierarchical Agglomerative clustering algorithm is implemented and the results are documented in Table I. Decision Tree, SVM, KNN and Naïve Bayes classifier results for Pitch features are shown in Table II.

TABLE I Raga Identification using clustering algorithms for Self-generated data

Name of Algorithm/ Performance Measure	K-means	Agglomerative
Accuracy in % - Autocorrelation	79.04%	83.85%
F1-score in % - Autocorrelation	16.18%	10.54%

TABLE II Raga Identification using different classification algorithms for Self-generated data

Name of Classifier/ Performance Measure	Decision Tree	Naïve Bayes	KNN	SVM
Accuracy in % - Autocorrelation	95.19%	82.19%	89.46%	84.74%
F1-score in % - Autocorrelation	80.76%	28.77%	57.87%	38.99%

In Pitch features, it is necessary to accurately identify Tonic and the other Notes. The slight error in Tonic may affect the Note values and ultimately PCD, which is feature vector for the algorithms. The results include accuracy and F1-score, calculated from the confusion matrix. In Accuracy True Positive as well as True Negatives are considered, so to analyze the performance of multiclass classifiers we need to observe other performance measure F1-score as well.F1-Score is combination of both Precision and Recall. It makes balance between Precision and Recall and that's why it is very useful to choose classifier for any dataset.

The accuracy of clustering algorithm is average but its F1-score is very less which shows True Positive rate of both the algorithm is need to be improved.

In this experimentation it is observed that all the classifiers worked well but to increase true positive rate, consideration of melodic phrases will be beneficial. As we have converted continuous in discrete form of 36 dimensions, the Decision Tree classifier outperformed among all the classifiers. The performance of SVM is better than Naïve Bayes classifier but not that of KNN. SVM is suitable for high dimensional data. The performance of SVM may improve in phrase based features as feature vector will be of high dimensional. To improve the performance of KNN we can analyze its parameters such as K values, weights for classes, or distance function and try for modified KNN.

V. CONCLUSION

Raga plays a major role in Indian Classical Music. It is a mixture of melody and Rhythm. The Indian Classical Music is very complex. Though the various rules are defined in it, still while performing, the artists make improvisation in it. As per our expectation Machine learning algorithms are successfully contributed in Indian Music classification using Raga framework. In this paper we implemented the supervised and unsupervised learning methods with PCD features. The results shows that the supervised learning methods are more suitable for identifying Raga correctly than unsupervised methods. In future we are planning to automatically find repeating phrases in Music for more accurate results, as the Raga identification by trained people is mainly performed based on repeating patterns of Notes and their variations.

REFERENCES

1. Hiteshwari Sharma, Rasmeet S.Bali, "Comparison of ML classifiers for Raga recognition," International Journal of Scientific and Research Publications, Vol. 5, No. 10, Oct. 2015, pp.1-5.
2. O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio", International Conference on Digital Audio Effects, Bordeaux, Finland 2007 pp. 1-8.
3. Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald David Scuse, "Weka Manual for version 3-7-8", University of Waikato, New Zealand, Jan. 2013, pp. 1-327.
4. Parag Chordia and Alex Rae, "Raga recognition using Pitch Class and Pitch Class Dyad Distribution," in 8th International Society of Music Information Retrieval Conference, Vienna, Austria, 2007, pp. 431-436.
5. G. Koduri, S. Gulati, P. Rao, " A survey of Raaga Recognition techniques and improvements to the state-of-the-art," in Conference of sound and Music, Padova, Italy, 6th -9th July 2011, pp.1-4.
6. Parag Chordia and Senturk S., "Joint recognition of Raga and tonic in North Indian music, " in Computer Music Journal, Vol. 37, No.- 3, Sept. 2013, pp. 82-98.
7. Patricio de la Cuadra, Aron Master, Craig Sapp, " Efficient Pitch Detection Techniques for Interactive Music," in Proceedings of International Computer Music Conference, Havana, Cuba, 17^{th} -22^{nd} Sept. 2001, pp. 1-4.
8. A. D. Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," Journal of the Acoustical Society of America, 2002, pp.1917–1930.
9. P. V. Rajkumar, K. P. Saishankar, and M. John, "Identification of Carnatic Raagas using Hidden Markov Models," in IEEE 9th International Symposium on Applied Machine Intelligence and Informatics, 27^{th} -29^{th} Jan. 2011, pp. 107–110.
10. V. Kumar, H. Pandya, C. V. Jawahar, " Identifying Ragas in Indian Music," in IEEE 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24^{th} -28^{th} Aug.2014, pp. 767-772.

11. Sankalp Gulati, J. Serra, K. Ganguli, S. Senturk, Xavier Serra, "Time-Delayed Melody Surfaces for Raga Recognition," in Proceedings of the 17th International Society for Music Information Retrieval Conference, New York, USA, 7th -11th Aug.2016, pp.751-757.
12. Sankalp Gulati, J. Serra, V. Ishwar, S. Senturk, Xavier Serra, " Phrased based Raga Recognition using vector space modeling," in IEEE International Conference on Acoustics, Speech, and Signal Processing, Shanghai, China 20th -25th Mar. 2016 pp.66-70.
13. Loretta Maria Paul Sheba, A. Revathy, "Raga Identification of Carnatic music using iterative clustering Algorithm," IEEE International Conference on Communication and Signal Processing, Melmaruvathur, Indian, 3rd -5th April 2014, pp. 1932-1936.
14. K. V. Deshmukh, P. J. Deore, "Raga identification using MFCC and Chroma Features," International Journal of Advanced Research in Computer Science, Vol. 8, No. 5, May-June 2017, pp.725-729.
15. Dr. Vidyadhar Oke. 2008 3rd Feb. "What is the meaning of Shrutis?" online available: http://www.22shruti.com.

AUTHORS PROFILE

Mrs. Kalyani C. Waghmare is currently pursuing Ph.D. in Computer Engineering at Pune Institute of Computer Technology, affiliated to Savitribai Phule Pune University, Pune, and Maharashtra, India. She did her M. E. in Computer Science Engineering and Information Technology from Vishwakarma Institute of Technology, Pune in 2006. She is working as Assistant Professor in Computer Engineering Department at Pune Institute of Computer Technology, Pune Since 1998. Her areas of interest are Data Mining, Design and Analysis of Algorithms, Data Structures, Information Retrieval, and Music, Speech processing.

Dr. Balwant A. Sonkamble has completed Ph.D. in Computer Science and Engineering from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra, India. He did his M.E. in Computer Engineering from Pune Institute of Computer Technology, Pune in 2004. He is Professor in Computer Engineering Department at Pune Institute of Computer Technology, Pune, India. His areas of interest are Artificial Intelligence and Speech Processing.