

Fast Mining Approach for Automatic Recognition of Rare Raga Swaram Notes

T.V.Mahendra
Associate Professor
Dept.of CSE
NBKRIST,Vidyanagar

G.Venugopal
Associate Professor
Dept.of CSE
VITS,Kavali

Abstract - Data Mining is a powerful technology to discover and analyze large data sets. It has its applications in various fields of Arts, Science and Engineering. One such field is Music. Music is a form of arts which comes under fine arts category. Carnatic music is a form of south Indian classical music which comprises of swarams (7 notes) to evolve music. This research work deals with automatic identification of rare Carnatic raga Swaram notes through Data Mining algorithms. The training sets considered for the work are Avarohanam notes of 72 melakartha raga and 212 Janya raga. C4.5 decision tree algorithm, Random Tree and Rule Induction algorithm were utilized to classify the Melakartha raga and the Janya raga. However the Janya raga swaram notes were also investigated through the use of appropriate feature relevance algorithms namely Feature ranking, Correlation based Feature Selection (CFS) filtering, and Fast Correlation based Filter (FCBF) filtering. The Melakartha raga data set was accurately classified with 100% accuracy by all the aforementioned classification techniques while predictor attributes selected through feature ranking algorithm produced nearly 90% accurate classification with Rule Induction algorithm on the Janya raga data.

Keywords - Carnatic raga swarams, Classification, Feature Selection, Rule Induction.

I. INTRODUCTION

Data Mining [1], an analysis part of knowledge discovery with immense potential helps to classify and access hidden details from a database. Data mining [2] helps in finding predictive information that experts may miss because it lies outside their expectations. Data Mining [1-2] detects patterns based on user queries with Classes, Clusters, Association, Sequential Patterns and Feature Selection. Feature selection [3] in supervised learning attempts to find a feature subset that produces higher classification accuracy. Classification [3] is the task of generalizing known structure, to apply to new data. Classification, a data mining technique is used to predict group membership for data instances. The above said data mining techniques can be applied to various fields' viz. medical, business, arts etc.

One such field in arts is Music. Music can be classified as Western Music and Indian Classical Music [4]. The Indian classical music can be further viewed as Hindustani Music and Carnatic Music. Carnatic music is a South Indian Classical music which consists of Raga (melody), seven Swaras (note with pitch), and Thal (rhythmic beat) to produce a musical concert note [5-6]. Each Raga consists of template Arohana and Avarohana from which notes (Swaras) are produced. It consists of 72 main Melakartha raga from which many combinations of Janya raga can be formed.

In this paper we apply Feature Selection and Classification methods to raga swara notes to automatically recognize raga and we perform a comparative analysis on the performance of classification and feature selection algorithms.

The rest of the paper is organized as follows. Section 2 discusses the Literature survey on classification of music and raga identification. Section 3 gives the description on Carnatic Music. Section 4 narrates the Data Mining methods utilized to recognize Carnatic Music Swaras, while Section 5 presents the Experimental Results. Section 6 concludes the paper with a view on possible extensions to this research.

II. LITERATURE SURVEY

The following paragraphs briefly narrate the research work carried out thus far in the field of music and raga identification. Shao et.al, [7] proposed an unsupervised clustering method based on a given measure of similarity that could be generated by Hidden Markov Models. The authors also made an attempt to better characterize music content by suggesting a novel segmentation scheme based on intrinsic rhythmic structure analysis of music and features that were extracted from those segments. The performance of the feature segmentation was superior to the traditional fixed-length method according to their stated experimental results. Their preliminary results also revealed the fact that the suggested method was comparable to a supervised method of classification. The proposed method was executed on four different kinds of music viz, Pop, Country, Jazz and Classical. Moreover the HMM results that were obtained, varied from 88% to 100%. Later Basili et.al, [8] investigated the impact of machine learning algorithms in the development of automatic music classification models aiming to capture genres distinctions. The authors believed that the study of genres as bodies of musical items aggregated according to subjective and local criteria. The authors attempted to model the process as an example- driven learning task. The authors investigated the impact of different musical features on the inductive accuracy by first generating a moderate-sized collection of examples for widely recognized genres and later evaluated the performances of different learning algorithms.

Their study investigated the factors affecting automated genre classification. The work mainly concerned western genres and traditional Malay music. Eight genres were introduced namely, *Dikir Barat, Etnik Sabah, Inang, Joget, Keroncong, Tumbuk Kalang, Wayang Kulit, and Zapin*. A total of 417 tracks from various Audio Compact Discs were collected and used as the dataset. Results showed that various factors such as the musical features extracted, classifiers employed, the size of the dataset, excerpt length, excerpt location and test set parameters enhanced classification results. Rao et.al, [10] presented an overview of techniques for Nearest Neighbor classification that placed focus on mechanisms for finding distance between neighbors using Cosine Distance, Earth Movers Distance and formulae were used to identify nearest neighbors and algorithms for classification in training and testing in order to identify Melakarta raga in Carnatic music. From the derived results it was concluded that Earth Movers Distance produced better results than Cosine Distance measure. Focused work on raga identification was initiated by Rajeswari Sridhar and T.V.Geetha, [11] who proposed a method to identify the raga of a Carnatic music signal. The input polyphonic music signal was analyzed and made to pass through a signal separation algorithm to separate the instrument and the vocal signal [12]. After extracting the vocal signal the extracted signal was segmented using the proposed segmentation algorithm. The authors presented the results on raga identification based on frequency of pitch in arohana [13]. The frequency components of the signal were determined and mapped into the swara sequence to detect the Raga of the particular song. Koduri et.al, [14] investigated the properties of a raga and the process by which people identified the raga. The authors presented a survey on the past raga recognition techniques and correlated them with human techniques, in both north Indian (Hindustani) and south Indian (Carnatic) music systems. They identified the major drawbacks and proposed minor yet multiple enhancements to the state-of-the-art raga recognition technique. Surendra Shetty et al [15] identified raga based on Note Transcription with signal processing and pitch variation. The authors depicted a system that took an audio file as input and transformed the same into a sequence of notes, and recognized the raga by extracting its arohana avarohana pattern. The system was analyzed for 50 different raga and tested for 20 different raga with 3-5 songs of each raga rising to 90 songs in entirety. The system performed with an accuracy of 95%. Shreyas Bayes et al [16] also made use of the same pitch class analysis for application in the notes of Hindustani music, whereas Sudha R [17] et al identified the raga using Musical Instrument Digital Interface (MIDI) file. Krishna A S [18] et al too reported on raga identification relating to Carnatic Music using Hidden Markov Chains.

Previous and existing research investigations have focused on either the pitch frequency or determination of notes or on western music classification. There have been misclassification rates in the previously declared experimental results. However this research work concentrates on classification of Swara notes to identify the raga using data mining techniques. We have made use of feature selection and classification algorithms to accurately recognize the raga for Carnatic music.

III. CARNATIC MUSIC ELEMENTS

The important elements[4] of carnatic music are Sruthi which is musical pitch, swara which specifies note with lower or higher position, raga which prescribes the rules for melody, and thala which refers to beat set (measure of time) of particular composite the other elements are the arohana that defines ascending order of swaras in the raga, and avarohana that defines the descending order. The arohana and avarohana are orthogonal in that the ascending order and descending order can be different for a raga. The swaras are seven notes [6] of scale in music as s shadja, rishabh, gandhar, madhyam, pancham, dhaiivat and nishad, and are shortened to Sa, Ri, Ga,

Ma, Pa, Dha, and Ni and written as S, R, G, M, P, D, N. These seven swaras have fifteen forms namely S, R1, R2, R3, G1, G2, G3, M1, M2, P, D1, D2, N1, D3, N2.

The raga is classified based on linearity (all Swaras) and non-linearity (missing swaras) of arohana and avarohana. Samporna raga are called Melakartha raga with all 7 notes with variation in rishabha, gandhar, madhyam, pancham, daivatham and nishagam. It totally constitutes 72 raga. Some sample raga is given in Table 1. The second classification includes Janya raga which have been derived from the Melakartha raga. Sample Melakartha raga and Janya raga is given in Table 1 and Table 2.

Table 1 Sample Melakartha Raga

Ragam	S1	S2	S3	S4	S5	S6	S7	S8
kanakAn'gī	S	R1	G1	M1	P	D1	N1	S
rathnAn'gī	S	R1	G1	M1	P	D1	N2	S
gAnamUrthī	S	R1	G1	M1	P	D1	N3	S
vanaspathī	S	R1	G1	M1	P	D2	N2	S
mAnavathī	S	R1	G1	M1	P	D2	N3	S
thAnarUpī	S	R1	G1	M1	P	D3	N3	S
sEnAvathī	S	R1	G2	M1	P	D1	N1	S

Table 2 Sample Janya Raga

Janya Ragam	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
8_dhanyAsī	S	G2	M1	P	N2	S	*	*	*	*	*	*	*
15_bowLi	S	R1	G3	P	D1	S	*	*	*	*	*	*	*
15_gowLa	S	R1	M1	P	N3	S	*	*	*	*	*	*	*
15_gowLipanthu	S	R1	M1	P	N3	S	*	*	*	*	*	*	*
15_gowri	S	R1	M1	P	N3	S	*	*	*	*	*	*	*
15_gundhakriya	S	R1	M1	P	N3	S	*	*	*	*	*	*	*
15_jaganmOhini	S	G3	M1	P	N3	S	*	*	*	*	*	*	*
15_malaHari	S	R1	M1	P	D1	S	*	*	*	*	*	*	*

IV. CLASSIFICATION FRAMEWORK OF RAGAS

The data mining framework comprises of a feature relevance analysis phase, classification and test phase. The system design is presented in Figure 1.

The individual data sets were first preprocessed and Feature relevance algorithm was applied for the pre-processed Janya raga.

The Melakartha raga and Janya raga were later classified using classification algorithms. The results were verified with a test data for each class of raga.

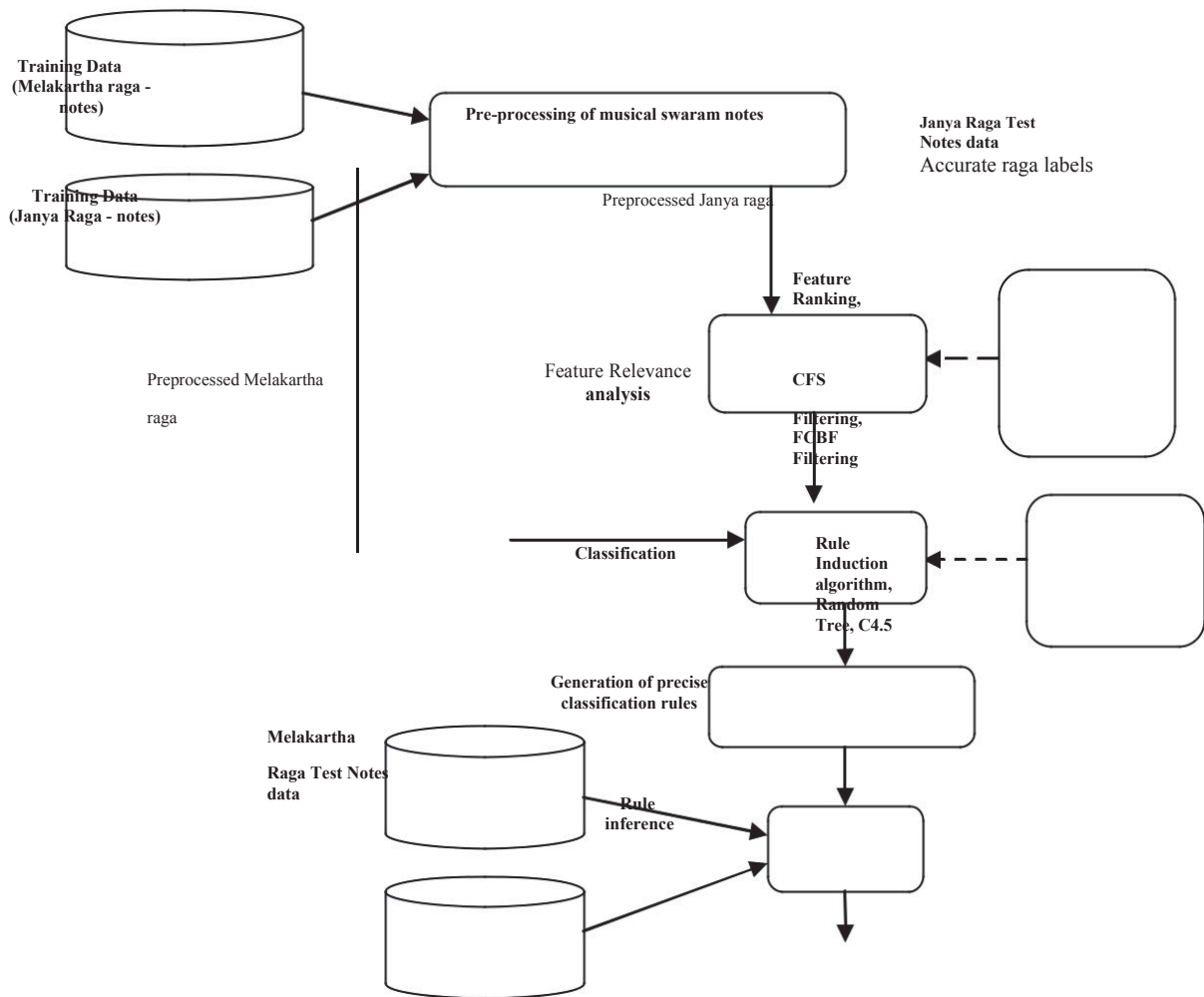


Figure 1. Proposed Data Mining Framework for Melakartha Raga and Janya Raga Classification

4.1 Data Training

The training data utilized in this analysis includes a set of 72 Melakartha raga avarohanam notes and a set of 212 Janya raga avarohanam notes. The swaram notes are unique to each raga. Melakartha raga has 8 predictor attributes named S1 to S8 while Janya raga has 13 predictor attributes referred as S1 to S13.

4.2 Pre-processing of Data

Data preprocessing is the task of formatting the raw training data to a form, suitable for execution of the data mining techniques. The raga was partitioned into two data sets. The first subset comprised of Melakartha raga and the other constituted the Janya raga derived from the Melakartha raga. Missing values noted in the Janya raga dataset were appropriately replaced with suitable values.

4.3 Raga Analysis

The Melakartha and Janya raga training records were explored individually. Each of the phases involved in the analysis is detailed below.

4.3.1 Melakartha Raga

The pre-processed Melakartha raga comprised of 72 instances defined by 8 predictors and one target class. This training data was fed to the classification phase to analyze and investigate the suitable classification algorithm that recognized the precise class of each raga. The three classification algorithms explored were C4.5 Decision tree algorithm, Random Tree algorithm and Rule Induction algorithm. All the three classification techniques produced 100 % classification accuracy on all the training records. However the Rule Induction classification algorithm showed promising results in both the training datasets. The pseudo-code of this algorithm [20] [21] is narrated in the following sub-section.

1) Rule Induction Classification Algorithm

Classification [4] [20] is a prominent phase in data mining that attempts to accurately place a given data under a certain category by evaluating the stored rules in the classifier knowledge base. Rule induction is an area of machine learning in which formal rules are extracted from a set of observations.

For each class C

Initialize to the set of all examples in E While E contains examples in class C

Create a rule R with an empty left-hand side predicate class

C

Until R is 100% accurate (or no remaining predictors) do For each attribute A not in R and each value V

Consider adding the condition (attribute-value) pair $A \rightarrow V$ on the left hand side of R

Select a and V to maximize the accuracy and covering of the attribute

Add $A \rightarrow V$ to R

Remove the examples covered by R from E

The rules extracted may represent a full scientific model of the data, or merely local patterns in the data. Decision tree algorithms are based on a divide-and-conquer approach to the classification problem. They work in a top-down manner, seeking at each stage an attribute to split on, that separates the best classes, and then recursively processes the resulting partitions. Sample rules generated by the Rule Induction Algorithm on the Melakartha raga dataset is given below in Table 3.

Table 3 Sample classification rules for Melakartha Raga

Antecedent	Consequent
IF S3 in [G1] -- S7 in [N1] - - S4 in [M1]	Ragam in [kanakAn'gi]
IF S3 in [G1] -- S7 in [N2] - - S4 in [M1] -- S6 in [D1]	Ragam in [rathnAn'gi]
IF S3 in [G1] -- S4 in [M1] -- S6 in [D1] -- S7 in [N3]	Ragam in [gAnamUrthi]
IF S3 in [G1] -- S6	Ragam in [vanaspathi]

in [D2] - - S4 in [M1] -- S7 in [N2]	
IF S3 in [G1] -- S6 in [D2] - - S4 in [M1] -- S7 in [N3]	Ragam in [mAnavathi]
IF S3 in [G1] -- S6 in [D3] - - S4 in [M1]	Ragam in [thAnarUpi]

The classification accuracy is verified by presenting a new Melakartha raga test data whose class is unknown, to the Rule base and comparing the obtained class value with the expected class value. The correct class label was obtained.

4.3.1 Janya Raga

The pre-processed Janya raga contained 212 instances predicted by 13 attributes with a single target class. The classification accuracy of the aforementioned techniques did not show promising results when all the predictor features were considered. Hence the feature selection techniques were employed to identify the most contributing features to classification. Feature Relevance [3] is a process that aims at filtering a subset of attributes from the initial training dataset based on the relevance, significance and contribution of the attribute in providing accurate classification results. Three feature selection algorithms were employed viz, Feature Ranking, Correlation based Feature Selection (CFS) and Fast Correlation based Filter (FCBF). Their performance was assessed based on the classification accuracy after feature selection. Investigation in this manner revealed that the predictor attributes selected by the Feature Ranking algorithm improved the classification accuracy of Rule Induction algorithm on the Janya raga data.

1) Feature Ranking Method

The Feature Ranking algorithm is also known as Univariate supervised feature ranking algorithm using CHI-2 criterion [19]. This algorithm ranks the input attributes according to their relevance. A cutting rule enables selection of a subset of these attributes. This algorithm however does not take into account the redundancy of the attributes. The significance p-value threshold was set to 0.001. The number of raga predictor attributes selected was 8. The reduced feature sub set result was used as the set of predictor attributes for the Classification phase to evaluate the classification accuracy after feature selection.

The pseudo-code of the feature selection algorithm that filtered the best set of features for the Janya raga is given below.

Input: Set of all features-F

Output: Reduced Feature subset –FF Algorithm:

Step 1: Compute X^2 (Chi-Square) statistic for each class-feature combination.

Step 2: Determine features that maximize X^2

Step 3: Sort the feature in the order of decreasing X^2 .

Step 4: Add the features one-by-one to the reduced feature subset (FF) till the cardinal number is attained.

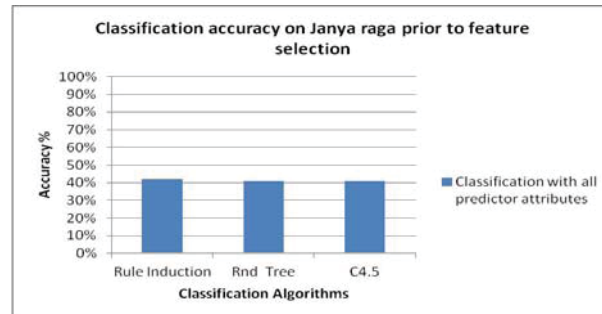
The sample rules obtained on classification with the Rule Induction algorithm on the Janya raga data is given below in Table 4.

Table 4 Sample classification rules for Janya Raga

Antecedent	Consequent
IF S3 in [S] -- S4 in [G3]	Ragam in [8_AHiri]
IF S5 in [D1] -- S3 in [G2]	Ragam in [8_bhUpALam]
IF S2 in [S]	Ragam in [8_punnAgavarALi]
IF S5 in [G2] -- S2 in [R2]	Ragam in [10_Sindhubhairavi]
IF S6 in [N1]	Ragam in [13_kalagada]

IF S5 in [D1] -- S4 in [M1]	Ragam in [14 vasantabhairavi]
IF S6 in [D1] -- S7 in [S]	Ragam in [15 ardhradhesi]
IF S6 in [D1] -- S2 in [R1] - - S8 in [S]	Ragam in [15 gujjari]
IF S7 in [D1] -- S4 in [P]	Ragam in [15 GummakAmbhoji]

after feature selection is depicted in Figure 2 and Figure 3 respectively.



These classification rules were used to determine the class of a new Janya raga record whose class label was previously unknown. Accuracy is validated by verifying the obtained classification result with the expected class value.

The experimental results for both the Melakartha raga and Janya raga are given in Section 5.

V. EXPERIMENTAL RESULTS

The performance of the classification algorithms was evaluated based on the accuracy obtained on both the Melakartha raga and the Janya raga. Accuracy [9] is defined as the number of samples correctly classified relative to the total number of samples in the data set. However in the case of the Janya raga, the results of the feature selection algorithm play a pivotal role in enhancing the classification accuracy.

5.1 Melakartha Raga

The C4.5, Random Tree and Rule Induction algorithms produced 100% classification accuracy with all the predictor attributes on the Melakartha raga data. The impact of the feature selection algorithms on the classification of the Melakartha raga data was minimal. Hence they were not utilized in further investigations. The accuracy is validated by presenting a Melakartha raga record whose class value is unknown and verifying the predicted class obtained on utilizing classification rules from the Rule Induction Algorithm.

5.2 Janya Raga

The Janya raga was analyzed based on accuracy after applying feature selection algorithms viz, Feature ranking, CFS and FCFB. Table 5 clearly depicts the improvement in classification accuracy after feature selection.

Table 5 Comparison of classification accuracy on Janya Raga

Algorithms	Classification Accuracy (%)		
	Rule Induction	Random Tree	C4.5
Without Feature Selection	42	41	41
Feature Ranking	90	43	20
CFS	20	41	20
FCFB	20	41	20

Fig.2 Classification accuracy on Janya raga before feature selection

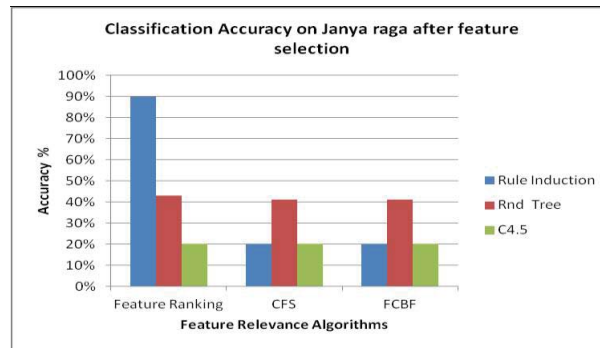


Figure 3 Comparison of classification accuracy on Janya raga with filtered predictor features

The cumulative results of our research findings is presented as a graph in Figure 4. The classification accuracy of the Rule Induction algorithm on the Janya raga is increased by 50% after selection of predictor features through Feature selection algorithm.

The training set with all swarasthanas are given as input parameters. Raga name was considered as target output for classification. The feature relevance algorithm is applied to remove redundancy of swarams in Janya raga.

In the case of the Feature relevance algorithm the Misclassification significance level was set to 100% and minimum support was set to 1. The Minimum split was set to 1 and Minimum support was 100% for the classification algorithm. By following these parameters Rule induction algorithm produced a maximum of 120 accurate classifications. Iteratively the rules were framed for the remaining subsets. Finally the maximum result with minimum misclassification rate was given by Feature Ranking Algorithm and Rule Induction Classification Algorithm for Janya raga. The classification accuracy on Janya raga before feature section and

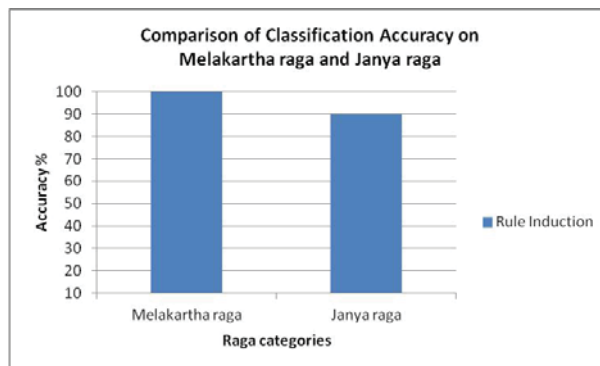


Figure 4. Comparison of classification accuracy on Melakartha raga and Janya raga

The results of this research suggest that the Rule induction algorithm shows higher classification accuracy in recognizing the class of a musical raga note. Moreover the use of feature relevance analysis in raga classification is also brought out in the case of the Janya raga notes.

VI. CONCLUSION AND FUTURE WORK

Data mining in the field of music is a relatively young field of research. Comprehending the structure of raga is at present feasible for the trained musicians alone. In view of this, to utilize computational techniques to aid in musical raga prediction is certainly a promising sphere of research. Moreover raga classification enables even common man with little knowledge of music to identify the complex structure of raga. The system proposed in this research work makes the above mentioned task easier as it identifies the raga to which each concert belongs using arohana and avarohana pattern. This work is found to be successful for both the source of raga namely Melakartha raga and Janya raga. Possible extensions to this work involve application of Genetic Algorithm methods in classification of raga to improve the result.

REFERENCES

- [1] Data Mining Concepts, <http://www.theartofscience.com/text/dmwhite/dmwhite.htm>
- [2] Principles in Data mining, http://en.wikipedia.org/wiki/exploring_the_possibility
- [3] Data mining Technologies, <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/>
- [4] Ries, R. E. (1969). The Cultural Setting of South Indian Music. *Asian Music* (University of Texas Press) 1 (2): 22– doi: 10.2307/833909
- [5] Gaurav Pandey, et al., Tansen: A system for automatic Raga identification, *IJCAI*, 2003
- [6] Priya K and Dr. Geetha Ramani R, 2012, GA method of evolving Carnatic Music Concert Swaras in the proceeding of *ICCA 2012*, ACM.
- [7] Xi Shao, C. Xu, and M.S. Kankanhalli, Unsupervised classification of music genre using hidden markov model, in *Proc. of the IEEE International Conf. of Multimedia Expo*, 2004.
- [8] Roberto Basili, Alfredo Serafini, Armando Stellato, Classification of Musical Genre: A Machine Learning Approach, @c 2004 Universitat Pompeu Fabra.
- [9] Noris Mohd Norowi, Shyamala Doraisamy, Rahmita Wirza, Factors affecting Genre Classification: An investigation incorporating non-western musical forms, 2005 Queen Mary, University of London, pp. 13-20
- [10] B. Tarakeswara Rao¹, Sivakoteswararao Chinnam², P *International Journal of Advanced Computer Science and Applications*, Vol. 1, No. 4, 2011, pp.43-48
- [11] Rajeswari Sridhar¹, and T.V. Geetha, Raga Identification of Carnatic music for Music information retrieval, *International journal on recent trends in Engineering*, vol. 1,2009
- [12] Rajeswari Sridhar and T.V. Geetha, Comparison of signal separation algorithms as applied to Carnatic music, *FRSM 2007*
- [13] Rajeswari Sridhar and T.V. Geetha, Swara identification of Carnatic music, IEEE Computer Society press, *proceeding of ICIT 2006*.
- [14] Gopala Krishna Koduri, Sankalp Gulati, Preeti Rao, A Survey of Raaga Recognition Techniques and Improvements to the State-of-the-Art, *Proceedings of Sound and Music Computing*, 2011.Italy.
- [15] Surendra Shetty and K.K. Achary, Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern, *International Journal of Recent Trends in Engineering* Vol. 1, No. 1, May 2009,pp.362-366
- [16] Shreyas Belle and Preethi Rao, 2009, Raga Identification by using Swara Intotation, *Journal of ITC sangeet research academy*, vol 23, 2009.
- [17] Sudha.R and Karthiravel A,2009,A system tool for identifying ragas using MIDI files in proceedings of *ICCEE2009*, IEEE Explore, Dubai, 644-647,DOI: 10.1109/ICCEE.2009.49
- [18] Krishna A S and Raj Kumar PV, 2011, Identification of carnatic ragas using hidden markov models in *proceedings of IEEE Explore*, Smolience, 107-110, DOI: 10.1109/SAMI.2011.5738857
- [19] Tanagra Tutorials and Software, www.tanagraltd.com
- [20] Mrs.Shomona Gracia Jacob and Dr. R.Geetha Ramani. Article: Discovery of Knowledge Patterns in Clinical Data through Data Mining Algorithms: Multi-class Categorization of Breast Tissue Data. *International Journal of Computer Applications* 32(7):46-53, October 2011. Published by Foundation of Computer Science, New York, USA.
- [21] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*.