Identifying Ragas in Indian Music

Vijay Kumar*, Harit Pandya*, C.V. Jawahar
International Institute of Information Technology, Hyderabad, India
{vijaykumar.r@research., harit.pandya@research., jawahar@}.iiit.ac.in

Abstract—**In this work, we propose a method to identify the ragas of an Indian Carnatic music signal. This has several interesting applications in digital music indexing, recommendation and retrieval. However, this problem is hard due to (i) the absence of a fixed frequency for a note (ii) relative scale of notes (iii) oscillations around a note, and (iv) improvisations. In this work, we attempt the raga classification problem in a non-linear SVM framework using a combination of two kernels that represent the similarities of a music signal using two different features-pitch-class profile and n-gram distribution of notes. This differs from the previous pitch-class profile based approaches where the temporal information of notes is ignored. We evaluated the proposed approach on our own raga dataset and CompMusic dataset and show an improvement of 10.19% by combining the information from two features relevant to Indian Carnatic music.**

I. INTRODUCTION

Raga - the melodic framework or formalization of melodies found in Indian Classical music (Carnatic and Hindustani) composes of a sequence of swaras depicting the mood and sentiments. Indian music has seven basic swaras (notes) namely (*Sa*, *Ri*, *Ga*, *Ma*, *Pa*, *Dha*, *Ni*). There are hundreds of ragas in Indian Carnatic music derived from 72 parent or Janaka ragas [20], formed by the combination of 12 *swarasthanas*. Identifying ragas is a central problem for appreciating, comparing and learning Indian music. Due to the overwhelming number, complicated structure and minor variations of ragas, even humans find it difficult to identify them, without years of practice. In this work, we report our initial attempt to automatically identify the *Ragas* in Indian Carnatic music.

The identification of ragas is very cognitive, and comes only after adequate amount of exposure. For automatic identification, some of the characteristics of ragas have to be converted into appropriate features. This becomes particularly challenging for Indian music due to the following reasons which needs to be addressed while converting a music piece into swara strings. (i) A music piece may be composed from multiple instruments during a performance. (ii) Unlike Western music, the notes in Indian music are not on a absolute scale but on a relative scale (iii) There is no fixed starting swara in a raga. (iv) Notes in Indian music do not have a fixed frequency but rather band of frequencies (oscillations) around a note. (v) The sequence of swaras in the ragas are not fixed and various improvisations are allowed [21] while citing a raga as long as the characteristics of raga are intact. These factors pose a serious challenge for automatic detection of ragas. Also, these factors make the problem distinct from applications such as genre recognition, comparison of vibrato music [22], emotions recognition etc., which are usually solved by extracting well defined features such as MFCC [25], melodic histogram [23]

from the main melody or bass line [23] of the music and classifying using *K-NN*, *naive Bayes* [24] or *SVM* [25] .

Inspite of the above mentioned challenges, there is an underlying structure found in a raga that can be captured. For example, one can identify a raga by finding its most prominent swara by counting the number of occurrences or the duration of each swara [9]. This may give insights into the set of notes and their frequencies in a raga, thereby helping in identification. *Gamakas*, the variations of pitch around a note can be used to identify a raga as only certain type of variations are allowed in each raga. Characteristic-motifs, similar to *Pakads* in Hindustani music, are the repetitive characteristics phrases of a raga that provide vital information in identifying a raga as these characteristic phrases vary from one raga to another.

In this work, we attempt the raga classification problem using a non-linear SVM and a combination of two different kernels. *We introduce the kernels to suit the Indian Carnatic music that represents the similarities of a raga based on pitch-class profile and n-gram note distribution.* This differs from the previous pitch-class profile based approaches where the temporal information of notes is ignored. This approach allows us to learn a decision boundary in the combined space of Pitch-class profile and n-gram note distribution, where different ragas are linearly separable. Given a music piece, we initially extract the predominant pitch values at every instant of time, convert to cents scale, map them to single octave and identify the stable note regions similar to [2]. These notes are then used to construct Pitch-class profile and n-gram distribution. While Pitch-class profile represent the distribution of pitch values, n-gram distribution provide the information about the occurrence of short sequence of notes. Thus our approach incorporates the information from both of these features unlike the previous approaches [2], [4] where either of these features were used but not both. We evaluate our approach on an extensive CompMusic dataset [2] consisting of 170 tunes corresponding to 10 ragas and achieve an improvement of 10.19% in accuracy.

Related Works: There are some attempts made in identifying the raga in a music. One method for raga classification is through the transcription of raga directly into swaras at every intervals of time and classifying using a classifier such as K-NN or SVM. In [6], relative frequencies are used instead of using an absolute frequencies as the notes have fixed ratio of frequencies. Though this approach addresses the issue with scale, it cannot handle multiple instruments. In [7], authors try to rectify this issue by identifying and extracting the fundamental frequency of the singer. All the other frequencies in the scale are then marked down based on their respective ratio with the identified fundamental frequency of the singer. Simple string matching techniques are then

employed to compare ragas. Though this approach performs reasonably well, improvisations and initial transcription errors may severely deteriorate the performance.

Few works have focused on a particular characteristic of a raga and designing the features accordingly to capture them. For example, authors of [8] compared *arohan* and *avrohan*. In [9], *Vadi* swara is included in their method. Authors in [4] treat the raga classification similar to word recognition problem. They assume the raga compositions to be composed of words formed from the alphabet of notes used in Indian classical music. They used an Hidden Markov Model (HMM) to learn the transition of notes in different ragas based on two observations. First, the sequences of notes for different ragas are well defined in Indian Music and a model based on discrete states with transitions between them possibly captures these sequences. Second, the notes are small in number requiring a simpler HMM setup. Note that, this method has a large dependency on the initial transcription of the notes. Even though this approach incorporates the temporal information of notes, it ignores the pitch distribution, which also provides vital information in identifying a raga.

Authors of [1] used a Gaussian mixture model (GMM) based HMM using three features, Chroma, MFCC and timbre features. They combined all the three features resulting in a 62 dimensional feature vector. They trained on four ragas namely *Darbari*, *Khamaj*, *Malhar* and *Sohini* and achieved reasonable performance. In [2], [10], pitch-class profile distribution is used to classify a raga. They created 3 variants of the pitch-class profile based on type of bin weighting and note stabilization. Simple K-NN classifier with KL-divergence distance is used to classify the ragas. They conducted experiments on extensive dataset of 10 ragas consisting of 170 tunes with at least 10 tunes in each raga. Note that, we conduct our experiments on this dataset and achieve superior performance. Also, note that in the approaches based on pitch-class profiles, temporal information of notes is ignored. However, in our approach, we partially capture the temporal information of notes of a raga by computing the n-gram histogram. By combining the n-gram histogram with pitch-class profiles, performance is further improved.

Kernels are used earlier to improve the performance of many audio and music related tasks such as classification and segmentation [16]–[18] by designing application specific kernels for SVM.

II. KERNELS FOR INDIAN CARNATIC MUSIC

In this section, we describe some of the characteristic features of Indian music and describe how the information provided by them can be used to identify a raga.

A. Features of Raga

Authors in [1], [2] characterize the raga by following characteristics:

1) *arohana* and *avarohana*: A raga has a fixed ascent (arohana) and descent (avarohana) swaras, without any strictness of their sequence in recitation. There are certain rules that are essential while citing a raga, though are not strictly followed. Also, many ragas
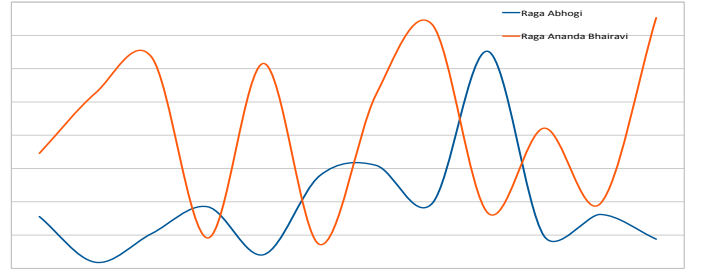

Fig. 1: Comparison of pitch class profiles for ragas *Abhogi* and *Ananda Bhairavi*

have same arohana and avarohana swaras making it difficult to uniquely identify a raga based on this characteristics alone.

2) *Vadi* and *Samwadi*: Vadi is the predominant swara in the raga and Samwadi is the next predominant swara. When arohana and avrohana in a raga have same swaras, Vadi and Samwadi could be used to distinguish them.

3) *Gamakas*: Gamakas refers to ornamentation used in the performance of Indian music. Unlike Western music, Indian music do not have a fixed frequency for a swara (note) and can have various variations (movements) around a note. These variations are called as *Gamakas*. The variations can occur in multiple forms. For example, it could be a rapid oscillation around a note or a slow transition from one note to another. For every raga, only a certain type of Gamakas (variations) are allowed around a swara giving an important clue for identification.

4) *Characteristic-motifs* - are the characteristic-phrases of a raga in Carnatic music which help in identifying a raga [5]. These are similar to Pakads in Hindustani music.

B. Kernels for Pitch-Class profiles:

Pitch-class profile distribution is used as a feature to classify ragas in [2], [10]. They provide a discriminative information based on distribution of notes in a raga. Even in different ragas with a same set of notes, the phrases often differ enough that one can see a recognizable difference in their pitch profiles. We thus define a kernel for pitch-class profiles that define the similarity of ragas based on pitch-class distribution.

We use the procedure employed in [2] to obtain the pitch-class profile of a music signal. Pitch values are detected at regular intervals of 10 ms from a given polyphonic audio music using a predominant melody extraction algorithm such as [13]. The frequency at each interval of the signal is determined by applying a Discrete Fourier Transform to different blocks of the audio and considering only the most *energetic* frequencies present in the audio signal. Pitch values are then extracted from these frequencies and are tracked based on how continuous a pitch values are in time and frequency domain. The pitch values are then converted to the cents scale with a tuning scale of 220 Hz. All the pitch values are then mapped to a single octave and stable note regions are identified. Finally, pitch values in these regions are quantized to nearest available
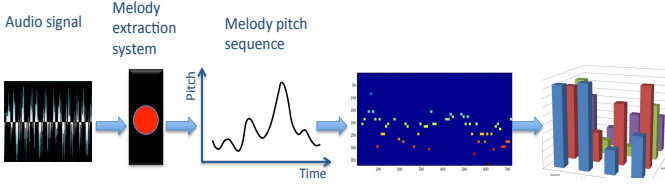
Fig. 2: Procedure to compute the n-gram. Predominant melody is extracted from a given polyphonic audio signal and a pitch class profile is constructed. From the pitch class profile, stable notes are identified and n-grams are computed.

Fig. 3: Two tunes belonging to same raga may have different pitch-class profiles (left) but their n-gram histogram of notes (right) show high similarity. We have used $n = 2$.

note value in 220 Hz equi-tempered scale. In order to measure the similarity of pitch-class profiles belonging to two different tunes, distribution intervals must be aligned in terms of the locations of corresponding scale degrees. In [2], this is done by a cyclic rotation of one of the distributions to achieve alignment of its tonic note interval with that of the other distribution. In the absence of tonic note of each tune, all possible alignments between two pitch class profiles are considered and the one that minimizes a certain distance measure is selected. Fig 1 shows the pitch-class profiles of ragas *Abhogi* and *Ananda Bhairavi*. Note how the pitch-class profiles for two different ragas differ providing vital information about a raga.

The histogram bins can be weighted in multiple ways. In [2], two types of binning are considered. One is based on the number of instances of a note, and another is the total duration of a note over all instances in the music piece. We consider the duration weighting of a bin which basically defines the average time spent around a raga. We define a kernel for pitch-class profile that gives a measure of similarity between two music pieces. It is a well known practice to use Kullback-Leibler (KL) divergence for comparing histograms. However, as KL-divergence is not symmetric, we symmetrize it as

$$\hat{D}_{KL}(\psi(x_i), \psi(x_j)) = \hat{d}_{KL}(\psi(x_i)|\psi(x_j)) + \hat{d}_{KL}(\psi(x_j)|\psi(x_i))$$

$$\hat{d}_{KL}(\psi(x_i)|\psi(x_j)) = \sum_k \psi(x_i(k)) \log \frac{\psi(x_i(k))}{\psi(x_j(k))} \quad (1)$$

where $\psi(x_i(k))$ is the k-th bin of the pitch-class profile $\psi(x_i)$ of the music sample x_i. Finally, we create a kernel for Pitch-class profile as follows,

$$K_1(i,j) = \exp(-D_{KL}(\psi(x_i), \psi(x_j))) \quad (2)$$

C. Kernel for n-gram distribution:

Pitch-class profiles provide information about the distribution of notes, however, they miss the temporal information of notes. Ragas usually contain repetitive *Characteristic-phrases or motifs* which provide a complementary information in identifying a raga. However, extracting these characteristic-motifs from a music itself is a challenging problem. Even humans find it difficult to identify them without years of practice. This is basically due to complex structure of these characteristic-motifs and their occurrence. These are usually spread throughout the raga without any specific time in their occurrence. They may also contain insertions of other swaras (notes) in between making the problem difficult to identify
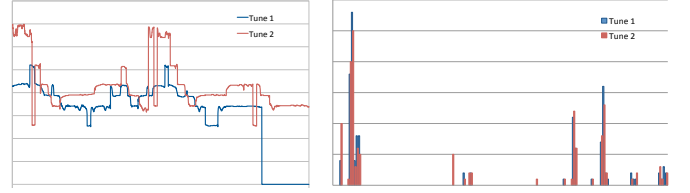
them automatically. One possible way to capture them to certain extent is by computing the n-gram distribution of the swaras in the ragas, similar to the way Pakads are captured in Hindustani music [4].

We employ the following procedure to obtain the n-gram histogram of a raga (Fig 2). Initially, we extract the dominant melody, pitch values, pitch contours, and stable note regions as explained in the previous section. Once the notes are identified, we construct the n-gram histogram feature as follows. Given a music sample x_i and n, we find all k-gram histogram ($k = 1, 2, \ldots, n$) and concatenate them to produce final feature vector.

$$\phi(x_i) = [H_1^T H_2^T ... H_n^T]^T = [H_k^T]^T \text{ where } k = 1, 2, \ldots n$$

H_k is a k-gram histogram of notes.

The use of all the k-grams ($k = 1, 2 \ldots, n$) is motivated by the fact that occurrence of the characteristic-phrases in a music is usually noisy and may contain insertion of notes in between them. We limited to 4-gram histogram of notes as it becomes computationally expensive to go beyond 4-grams. In Fig 3, we show an example that demonstrates how the n-gram histogram of two tunes are highly similar even though the pitch class profiles show some variation.

We define a kernel to represent the similarity of ragas based on n-gram histogram of notes. We found the radial basis function (RBF) kernel to be effective for capturing the this similarity defined as,

$$K_2(i,j) = \exp(-\frac{||\phi(x_i) - \phi(x_j)||_2^2}{2\sigma^2}) \quad (3)$$

In the absence of tonic note, we align the n-gram distributions of two tunes in terms of locations of corresponding scale degrees. Given a n-gram distribution of two tunes, we initially align the 1-gram distribution of two tunes through cyclic rotation as explained in the previous section. Once the correspondence between scale degrees of two 1-grams is obtained, we use this correspondence to align the n-grams.

Note that, both of the above defined kernels based on Pitch-class profiles and n-gram histogram are valid as the term $exp(-a)$ is always positive and greater than 0, for $a \geq 0$.

III. CLASSIFICATION

We identify a raga by combining the information from two different and relevant features, Pitch-class profiles and n-gram distribution. We incorporate this systematically into an SVM
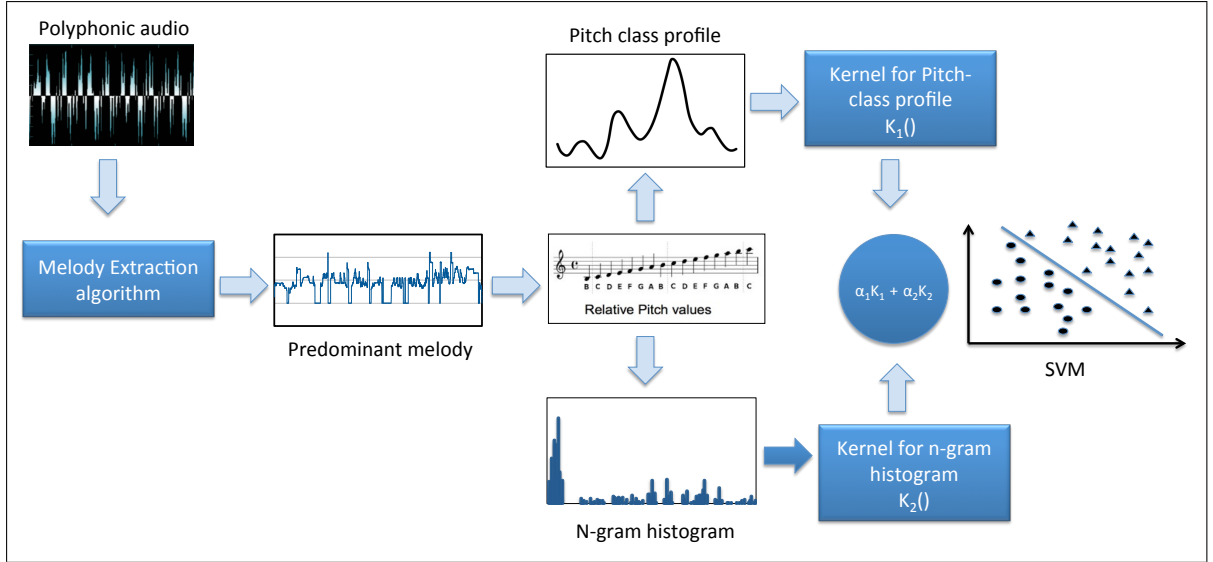
Fig. 4: Overview of our proposed approach. Predominant melody is extracted from a given polyphonic audio signal and pitch values are identified. An SVM model is learnt using two different non-linear kernels that define the similarities of ragas based on Pitch-class profiles and n-gram histogram of notes.

framework by defining a combined kernel over them. This is in contrast to previous pitch-class profile based approaches where the temporal information of notes is ignored.

Given a set of training pairs $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$, traditional SVM tries to find the maximum-margin hyperplane defined by the parameter w that separates points with $y_i = 1$ from those having $y_i = -1$. This is achieved by solving the following optimization problem:

$$\arg \min_{w,b,\xi_i} \frac{1}{2}||w||_2 + C \sum_{i=1}^{n} \xi_i$$
$$\text{s.t.} \quad y_i(w^t x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \quad (4)$$

where ξ_i's are the slack variables denoting the violations made by the training points. During inference, a test sample x is predicted by finding the sign of $(w^t x + b)$. Alternatively, solution could be achieved by maximization of dual formulation:

$$J(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^t x_j$$
$$\text{s.t.} \quad \sum \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C. \quad (5)$$

For the above formulation, a test sample x is predicted by finding the sign of $\sum_{i=1}^{m} \alpha_i y_i x_i^t x + b$ where m denote the number of support vectors. The dual formulation allows to apply the Kernel Trick to compute the dot product in the feature space $< \phi(x_1), \phi(x_2) >= K(x_1, x_2)$, $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ without explicitly computing the features $\phi(x_i)$. With the use of a kernel, above formulation becomes

$$J(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$\text{s.t.} \quad \sum \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C. \quad (6)$$

Any given test sample can be labeled using the sign of $\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b$.

For multi-class problems, one can build multiple binary classifiers and adopt strategies like *one-vs-rest* or *one-vs-one* to infer a decision or extend the binary SVM to handle multiple classes through a single optimization.

The type of kernel and parameters to select depend largely on the application at hand. Several kernel functions have been proposed in the literature starting from generic RBF kernel to an application specific kernels [15]. Any clue for similarity could be captured in the form of kernel, as long as it is closed under positive semi-definiteness, $K \geq 0$. One can also define a kernel as a linear combination of the individual kernels each representing different kinds of similarity, under a condition that each individual kernel is valid.

$$K = \sum_i \alpha_i K_i \quad (7)$$

The weights α_i can be selected heuristically based on the cross-validation errors, or learned in a multiple kernel learning framework [14]. For raga classification problem, we define our kernel as linear combination $\alpha_1 K_1$ and $\alpha_2 K_2$ of two different kernels representing similarities of an audio based on Pitch-class profiles and n-gram histogram of notes. This provides a mechanism to systematically combine the similarities using two heterogeneous features into a single max-margin framework. The weights α_i are selected here based on the cross-validation error. Our entire approach is summarized in Fig 4.

IV. RESULTS AND DISCUSSIONS

A. Datasets:

We evaluate the performance of our proposed approach on our dataset and CompMusic dataset [2]. These two datasets are summarized in Table I. While our data set is small consisting of only 4 ragas with limited instruments, CompMusic dataset is extensive consisting of 10 ragas and variety of musical instruments.

TABLE I: Summary of Our and CompMusic dataset.

Dataset	Composition of tunes
Our dataset	60 tunes, 5 artists, 4 ragas, 2 instruments
CompMusic dataset	170 tunes, 31 artists, 10 ragas, 27 instruments

(i) Our Dataset: To evaluate our method, we created a dataset comprising of 4 ragas namely *Kalyanavasantham, Nattakurinji, Ranjani, and Bilhari*. All audio files are of type instrumental of type flute and are of approximately 20 minute duration from CD recordings. We divided these full length recordings into 1 minute audio clips to create our dataset. Each clip is 44.1 KHz sampled, stereo-channel and m4a encoded.

(ii) CompMusic Dataset: We test our method on another dataset from the authors of [2]. CompMusic dataset is an extensive dataset that includes compositions from several artists spanning several decades, male and female, and all the popular instruments. The clips were extracted from the live performances and CD recordings of 31 artists, both vocal (male and female) and instrumental (Veena, Violin, Mandolin and Saxophone) music. The dataset consisted of 170 tunes from across 10 ragas with at least 10 tunes in each raga (except Ananda Bhairavi with 9 tunes). The duration of each tune averages 1 minute. The tunes are converted to mono-channel, 22.05 kHz sampling rate, 16 bit PCM. The composition of dataset is shown in Table II.

TABLE II: Composition of CompMusic dataset.

Raga	Total tunes	Average duration (sec)	Composition of Tunes
Abheri	11	61.3	6 vocal, 5 instrumental
Abhogi	10	62.0	5 vocal, 5 instrumental
Ananda Bhairavi	09	64.7	4 vocal, 5 instrumental
Arabhi	10	64.9	8 vocal, 2 instrumental
Atana	21	56.7	12 vocal, 9 instrumental
Begada	17	61.1	9 vocal, 8 instrumental
Behag	14	59.7	12 vocal, 2 instrumental
Bilhari	13	61.3	10 vocal, 3 instrumental
Hamsadwani	41	57.0	14 vocal, 27 instrumental
Hindolam	24	60.0	15 vocal, 9 instrumental

B. Results

We initially conducted experiments on our dataset. We implemented the feature extraction procedure for ragas as described in [2]. Polyphonic audio signals are converted to predominant melody using melody extraction software [11]. Pitch-class profiles and n-grams are extracted as explained in the Section 2. We randomly divide the dataset into training and testing set so that half is used for training and other half is used for testing. We conducted 10 trials with random training and testing sets and report the mean accuracy. We compare our approach with the approach proposed by [2]. Authors in [2] calculated the pitch-class profile in multiple ways namely, P1, P2 and P3 based on whether only stable regions are considered or not and weighting of the bins. In P1 and P2, only stable

regions of the pitch class are considered while in P3, all the regions are considered. The difference in P1 and P2 lies in type of weighting the bins. In P1, a note bin is weighted by the number of instances of the note, and in P2 by the total duration over all instances of the note in the music piece. In [2], k-NN classifier with KL-divergence distance is used as a classifier. For our approach, we report results using $n = 2, 3$ and 4 grams while calculating the n-gram kernel K_2. We selected values of α_1 and α_2 as 5 and 3 respectively based on cross-validation errors. We set the value of RBF parameter $\sigma = 0.01$ and $C = 5$ through a grid-search and using cross-validation errors. Results are shown in Table IV and the best results for both methods are shown in Table III. It is clear that, our approach which combines Pitch-class profiles and n-gram histogram of notes achieves superior performance compared to [2] where only pitch-class profile is used.

TABLE III: Comparison of best results of approach [2] with our approach on various datasets (%).

Method	Our dataset	CompMusic dataset
Koduri et al. [2]	96.3	73.2
Our approach	97.3	83.39

TABLE IV: Comparison of performance of approach [2] for various pitch-class profiles with our approach on our dataset (%).

Method	1-NN	3-NN	5-NN	7-NN
P1 [2]	96.1	95.0	93.67	90.17
P2 [2]	96.3	96.0	95.5	94.0
P3 (12 bins) [2]	94.2	86.2	83.2	79.5
P3 (24 bins) [2]	94.0	85.0	79.6	71.5
P3 (36 bins) [2]	94.8	92.3	88.5	83.5
P3 (72 bins) [2]	94.8	90.3	81.8	75.8
P3 (240 bins) [2]	92.5	83.0	84.8	70.2
Our approach (2-gram)	96.0			
Our approach (3-gram)	97.7			
Our approach (4-gram)	97.3			

In another experiment, we tested our approach on Comp-Music dataset. We used the same experimental procedure as described above. Results are shown in Table V with best results shown in Table III. Results in the table clearly demonstrates the superiority of our approach. The best accuracy obtained by our approach is 83.39% which is higher than their best reported accuracy 73.2%. This clearly agrees to our intuition that including temporal information of a raga with pitch-class profiles improves the performance.

C. Effectiveness of n-grams and Pitch-class profiles:

In order to understand the effectiveness of n-gram and pitch-class profiles in identifying a raga, we performed an experiment by considering either of these features and together. Table VI shows the result of the experiment. It is clear that, both these features provide a vital clue about a raga and combining them improves the performance.

TABLE V: Comparison of performance of approach [2] for various pitch-class profiles with our approach on CompMusic dataset (%).

Method	1-NN	3-NN	5-NN	7-NN
P1 [2]	51.3	48.6	52.6	51.2
P2 [2]	73.2	72.1	71.4	67.9
P3 (12 bins) [2]	72.0	68.1	69.2	68.9
P3 (24 bins) [2]	71.3	67.7	69.2	66.0
P3 (36 bins) [2]	68.8	65.91	66.5	64.3
P3 (72 bins) [2]	67.2	63.8	64.8	61.0
P3 (240 bins) [2]	63.6	57.9	58.5	57.2
Our approach (2-gram)	79.43			
Our approach (3-gram)	81.32			
Our approach (4-gram)	83.39			

TABLE VI: Improvement in classification performance due to various kernels (%).

Feature	SVM with kernel K_1	SVM with kernel K_2	SVM with kernel $K = \alpha_1 K_1 + \alpha_2 K_2$
2-gram	70.51	57.34	79.43
3-gram	70.51	60.25	81.32
4-gram	70.51	63.41	83.39

Our initial attempt has demonstrated the utility of combining two relevant features of music by defining kernels popular in machine learning. There is much more to achieve before we obtain a reliable raga recognition system.

V. Conclusion

In this paper, we looked into the problem of raga identification in Indian Carnatic music. Based on the observation that, existing methods are either based on pitch-class profiles or n-gram histogram of notes but not both, we tried to incorporate both of them in a multi-class SVM framework by linearly combining the two kernels. Each of these kernels capture the similarities of a raga based on Pitch-class profiles and n-gram histogram of notes. This is in contrast to previous pitch-class profile based approaches where the temporal information of notes is ignored. We evaluated our proposed approach on CompMusic dataset and our own dataset and show that combining the clues from pitch-class profiles and n-gram histogram indeed improves the performance.

Acknowledgment

References

[1] Pranay Dighe, Parul Agrawal, Harish Karnick, Siddartha Thota and Bhiksha Raj, *Scale independent raga identification using chromagram patterns and swara based features*, IEEE International Conference on Multimedia and Expo Workshops, 2013.

[2] Koduri Gopala-Krishna, Sankalp Gulati and Preeti Rao, *A Survey of Raaga Recognition Techniques and Improvements to the State-of-the-Art*, Sound and Music Computing, 2011.

[3] Honglak Lee, Peter Pham, Yan Largman and Andrew Y Ng, *Unsupervised feature learning for audio classification using convolutional deep belief networks*, Neural Information Processing Systems, 2009.

[4] Gurav Pandey, Gaurav P, Chaitanya Mishra and Paul Ipe, *Tansen : A System For Automatic Raga Identification*, Indian International Conference on Artificial Intelligence, 2003.

[5] Vignesh Ishwar, Shrey Dutta, Ashwin Bellur, Hema A. Murthy, *Motif Spotting in an Alapana in Carnatic Music*, International Society on Music Information Retrieval, 2013

[6] Preeti Rao and Anand Raju, *Building a melody retrieval system*, National Conference on Communications, 2002.

[7] Rajeswari Sridhar, and T.V. Geetha, *Raga Identification of Carnatic music for Music Information Retrieval*, International Journal of Recent Trends in Engineering, 2009.

[8] S. Shetty and K. Achary, *Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern*, International Journal of Recent Trends in Engineering, 2009.

[9] S. Shetty and K. Achary, *Raga Identification of Carnatic music for Music Information Retrieval*, International Journal of Recent Trends in Engineering, 2009.

[10] P. Chordia and A. Rae, *Raag recognition using pitch- class and pitch-class dyad distributions*, International Society on Music Information Retrieval, 2007.

[11] *http://essentia.upf.edu.*

[12] Swift, Gordon N, *Ornamentation in South Indian Music and the Violin*, Journal of the Society for Asian Music, 1990.

[13] J. Salamon, E. Gomez, *Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics*, IEEE Transactions on Audio, Speech and Language Processing, 2012.

[14] Mehmet Gonen, Ethem Alpaydn, *Multiple Kernel Learning Algorithms*, Journal of Machine Learning Research, 2011.

[15] Subhransu Maji, Alexander C. Berg, Jitendra Malik, *Efficient Classification for Additive Kernel SVMs*, Pattern Analysis and Machine Intelligence, 2012.

[16] Lie Lu, Hong Jiang Zhang, Stan Z. Li, *Content-based audio classification and segmentation by using support vector machines*, MultiMedia Systems, 2003.

[17] Na Yang, Rajani Muraleedharan, JoHannah Kohl, Ilker Demirkol, Wendi Heinzelman, and Melissa Sturge-Apple, *Speech-based Emotion Classification Using Multiclass SVM with Hybrid Kernel and Thresholding Fusion*, IEEE Workshop on Spoken Language Technology, 2012

[18] Joder, C., Essid, S., and Richard, G., *Alignment Kernels for Audio Classification with application to Music Instrument Recognition*, European Signal Processing Conference, 2008

[19] Pranay Dighe, Harish Karnick, Bhiksha Raj, *Swara Histogram Based Structural Analysis And Identification Of Indian Classical Ragas*, International Society for Music Information Retrieval, 2013.

[20] Mandayam Bharati Vedavalli, *Sangita sastra sangraha: A Guide to theory of Indian music,* Page 25.

[21] Bruno Nettl, Melinda Russell, *In the Course of Performance: Studies in the World of Musical Improvisation,* Chapter 10, Page 219.

[22] Felix Weninger, Noam Amir, Ofer Amir, Irit Ronen, Florian Eyben, Bjrn Schuller, *Robust feature extraction for automatic recognition of vibrato singing in recorded polyphonic music.*, International Conference on Acoustics, Speech and Signal Processing, 2012.

[23] Umut Simsekli, *Automatic Music Genre Classification Using Bass Lines*, International Conference on Pattern Recognition, 2010.

[24] Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang, *Learning Naive Bayes Classifiers for Music Classification and Retrieval*, International Conference on Pattern Recognition, 2010.

[25] Kamelia Aryafar, Sina Jafarpour, Ali Shokoufandeh, *Automatic musical genre classification using sparsity-eager support vector machines*, International Conference on Pattern Recognition, 2012.