TALA CLASSIFICATION IN CARNATIC MUSIC USING AUDIO THUMBNAILING

Amulya Sri Pulijala and Suryakanth V Gangashetty

International Institute of Information Technology, Hyderabad, India

ABSTRACT

The concept of Raga and Tala is integral part of Indian Classical music. Raga is the melodic component while Tala is the rhythmic component in the music. Hence, Tala classification and identification is a paramount problem in the area of Music Information Retrieval (MIR) systems. Although there are seven basic Talas in Carnatic Music, a further subdivision of them gives a total of 175 ragas. Statistical and machine learning approaches are proposed in Literature Survey to classify Talas. However, they use complete musical recording for training and testing. As part of this paper, a novel approach is proposed for the first time in Carnatic music to classify Talas using repetitive structure called Thumbnails.

KEYWORDS

Tala Classification, Carnatic Music, Audio Thumbnails, SVM, CNN

1. INTRODUCTION

Music has become integral part of our lives today. With the digital revolution and growth of computational power, browsing and storage has become accessible and effective. It has paved a new way of generation and analysis of music in the area of music signal processing. The Natyasastra of Bharata and the Sangitaratnakara of Sarangadeva are the oldest existing sources of information on Indian Classical Music. Carnatic and Hindustani are the two broad variations of classical music in India based on its geographical association. Carnatic belongs to the southern part, while Hindustani is from northern part of the sub-continent. Swara, Raga and Tala can be described as important elements in Indian classical music. Raga is the melodic part and Talam is the rhythmic component.

There are 12 notes or swara which forms the primary aspect of Carnatic music as well as Hindustani classical music, along with raga and tala. It is described as “Sruthi Mata, Laya Pitha”, which means Shruthi or Tonic or base frequency is considered as Mother whereas Tala is like father [1]. Tala has no reference in the earliest system of music, popularly referred as "Samagana". However, it existed during Gandarva music. There are various classification schemes of Tala in Carnatic Music. The ancient 108 anga Talas, the 72 Melakartha Tala system and the Suladi Sapta Tala system are some of the classification schemes. Among these Sapta Tala system is prominent one which was popularized by Purandaradasa. Seven Talas in Carnatic music are as follows namely, Adi, Rupaka, Eka, Jhampe, Dhruva, Matya, Ata, Triputa etc.

These seven talas are further subdivided, based on the change in Tala due to change in five Jaathis (Jaathis of Tala means that the amount of beats that a laghu can take). The five Jaathis are as follows: Tisra, Chatusra, Khanda, Misra and Sankeerna. Thus we get total of 35 Talas after dividing on the basis of Jathis. These 35 Talas allow further subdivision based on five

David C. Wyld et al. (Eds): CCSEA, BIoT, DKMP, CLOUD, NLCAI, SIPRO - 2020
pp. 195-201, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101015

Gathis/Nadais(Gathis means speed). The five Gathis are same as above Jaathis. Finally after Jathi and Gathi subdivision of Seven Talas, we get a total of 175 talas in Carnatic Music. Three elements namely, Jaathi name, Tala name and Gati name are required to describe a Tala [2].

Tala has ten important features called dasapranas. The following is the brief description of these Dasapranas [3]:

- Anga : Part or Limb
- Jati : type or kind. It describes variations in Anga(Laghu)
- Kriya: Action
- Kaala: Duration or measurement of time
- Graha: Describes where song commences, may not be at the beginning of tala
- Marga: Path. Describes duration of kriya/action. In other words, how tala is performed in various different songs
- Kala: Denotes number of matras in which kriya is subdivided
- Laya: Time gap between two consecutive kriyas. It sets the tempo
- Yati: rhythmic pattern in composition with reference to anga
- Prasthara: detailed elaboration of rhythmic pattern

Alex and his team [4] has worked on Tala Classification with three different types of Talas. [5] aimed at estimating the tala or akshara period using self similarity matrix. [16] compared beat detection, sound energy algorithm and frequency selected sound algorithm in order to classify talas. Deep Neural Network with Group delay was used by [17] for onset detection of mrudangam strokes. [18] used various data driven approaches to generate rhythm/ tala. Gaussian Models were used by [19] to classify Talas and Ragas. However, all these methodologies are using complete musical recording for training purposes. Hence, the feature set and time required for training and testing is significantly more.

2. PROPOSED METHODOLOGY

We propose a method of classification of Talas using Audio thumbnailing. The algorithm proposed is presented in Algorithm 1. We use self similarity matrices to generate thumbnails which are representative part of musical recording. Self similarity matrix is an important feature in any time series as it captures repetitions in the form of path-like structures. These repetitions are captured and the most repetitive path is found and named as thumbnail. Before generating thumbnails, the musical piece is normalized with respect to the tonic frequency of the singer. Tonic identification is performed in two stages as in [6]. We performed multi-pitch analysis of given audio signal in order to identify pitch class. Advantage of multi-pitch analysis is, it gives the drone sound which constantly runs in the background. In the next stage, estimate the octave in which tonic of singer lies and then analyzes the predominant melody. The audio signal is then normalized with the help of the tonic frequency before computing chroma features.

Computation of Self similarity Matrix: Chroma vector representation of the song is used to compute the self-similarity matrix. Given a sequence $X = (x_1, x_2, ...x_N)$ the self similarity matrix is used to compare all the elements with each other in the sequence. This gives us N-square self similarity matrix $S \in \mathbb{R}^{N \times N}$ defined by

$$\mathbf{S}(n,m) = s(x_n, x_m)$$

Where $n, m \in [1:N]$.

Chroma features are an array of 12-dimensional vectors from the short-term Fourier Transform of the musical recording which shows the pitch distribution over time. Concept of Self similarity matrix is fundamental in computing structural properties of any music recording [7]. Notable property of SSM is that, repetitions give rise to path-like structures.

Enhance Self Similarity Matrix: In order to enhance the self-similarity matrix, the paths parallel to the diagonal are smoothed out using convolutional filters. The irrelevant noisy structures in the matrix are suppressed using thresholding and scaling.

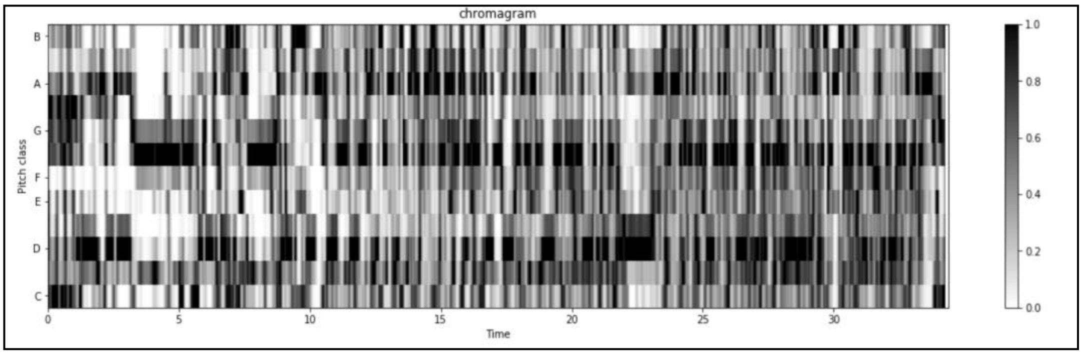

Fig.1. Chromogram of Song Inta Chala in Adi Talam

- Computation of Fitness Score: Algorithm proposed by Muller [7] is used to compute the fitness score. Briefly, the following are the steps involved:
- Computing the Score Matrix D for every section of SSM . This score matrix is used to compute the path score, which represents the extent of repetitive nature.

$$\mathbf{D}(n,m)=\mathbf{S}^{\alpha}(n,m)+\max\{\mathbf{D}(i, j)|(i, j)\in\Phi(n,m)\}$$

Where $\mathbf{S}^{\alpha}$ is the submatrix of SSM S from n to m and $\Phi(n,m)$ is the set which consists of predecessors that precede (n,m) in SSM

- Computing the Coverage Score. This represents the total length of musical recording does optimal segment cover.
- Computing the fitness score using Score Matrix and Coverage score. In other words, fitness score is the harmonic mean of path score and coverage score of given audio.

Algorithm 1: Audio Thumbnailing Algorithm

Input: Chroma vectors of music recording
Output: audio segment of maximal fitness score

```
function GenerateThumbnail(chroma, threshold)
    ssm ← Computation of Self Similarity Matrix
    for all I,j,s,t 0 ≤ i ≤ j ≤ audio do
        φ (i,j) ← Computation of Fitness Score
    α ← argmax(i,j, φ (i,j))
    return α
```

- The last step is to build the classification model. Machine learning has paved way for new area of research in the field of audio classification. Support Vector Machines [8] and CNN-RNN [9] are proved methodologies in the field of audio classification with high accuracies. We propose to use the audio thumbnail to extract the feature vector for classification, which consists of the chroma features, spectral contrast and Mel spectrogram features. The baseline model used for this classification task is Support Vector Machine [8].

- Parameters for the classifier were tuned to optimal value. Three fold cross validation were used for training and testing. Apart from this, a CNN-RNN Model [9] is also used to train the classifier with convolution layers and recurrent ones.

3. IMPLEMENTATION:

The algorithm for tonic identification and audio thumbnailing is implemented in Python. Tonic is identified and the musical recording is normalized with tonic frequency [6]. We then extract chroma features using short term Fourier Transform as explained above. Then Self Similarity matrix is found and smoothened using Forward backward smoothening. Forward Backward smoothing uses a convolutional filter averaging in diagonal direction both in forward and backward direction. Smoothing enhances path like structures. Figure 1 demonstrates the chromagram for a portion of performance of Tala Adi, Raga Begada called Inta Chala by Ariyakudi Ramanuja Iyer. Figure 2 shows Self similarity matrix before and after enhancement of same musical piece.

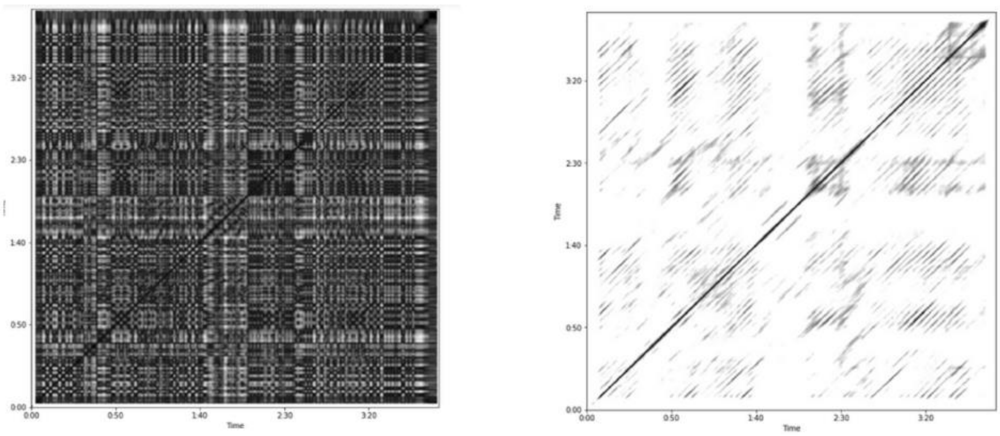

Fig.2. Self Similarity Matrix before and After Enhancing and Smoothing for Song Inta Chala in Adi Tala

These functions are implemented using librosa library [14] in Python. A thumbnail is then selected after computation of fitness score. As mentioned above SVM and CNN RNN is trained for the classification. SVM parameters are selected and tuned for extensive grid search algorithm [15]. A six dimensional feature set consists of features such as, short term fourier transform, mel frequency cepstrum, chroma, melspectogram, spectral contrast, tonnetz. Support Vector Machine classifier is optimized by three fold cross validation using GridSearch CV in sklearn library [13]. The performance of Hyper parameters [11][12] are evaluated and tuned for effective accuracy. Best parameters found on dataset with Kernel radial basis is, C: 10 and gamma as e -8.

Our CNN-RNN structure is inspired by [9]. As part of this architecture we train the classifier with characteristics learned from both convolutional as well as recurrent layers. The intuition behind taking this parallel approach is that, convolutional properties captures spatial relationship among features well, while recurrent is effective for capturing temporal characteristics. The CNN has 4 convolutional layers interleaved with 2 pooling operations, then followed by a dense layer. RNN is implemented with LSTM layer and batch size of 128. CNN RNN is implemented in Python using Keras library [20].

Dataset: The dataset provided by CompMusic [10] is used to evaluate the model. Total 140 songs, of which 100 songs were used as training set and 40 songs were used as test data set. Five Talas were used which are as follows: Adi talam (cycle of 8 beats), Rupaka talam(cycle of 6 beats), Chaturasra jati Eka talam(cycle of 4 beats), Trisra jati Eka(cycle of 3 beats), Khanda Jati Eka(cycle of 5 beats). Selection of these Talas is based on unique number of beats. Table 1 illustrates the division of dataset for training and testing.

The CNN-RNN classification model significantly showed higher accuracy levels compared to standard SVM Classifier. CNN classifier produces a test accuracy of 84% where the SVM classifier gave an accuracy of 65%.

TABLE I. Table Describing Dataset For Tala Classification

Name of the Tala	Number of Beats	Training, Testing Set	Average Duration of songs in min
Adi	8	(20,8)	07.20
Rupaka	6	(20,8)	08.10
Chaturasra Jati Eka	4	(20,8)	06.40
Tisra Jati Eka	3	(20,8)	09.05
Khanda Jati Eka	5	(20,8)	08.20

4. CONCLUSION AND FUTURE WORK

Using the thumbnailing instead of complete musical recording to extract features for training and testing reduces the feature vector sizes to significant extent. However, extracting the thumbnail is computationally not an easy task. It takes 90sec to compute thumbnail for an average recording of 6 min.

Tala classification is not explored by many in the area of Music Information Retrieval and the problem is unique to genres of Indian subcontinent. Existing researchers have used several statistical methodologies, using complete song. This paper proposes a novel technique to use repetitive structure of musical recording, namely audio thumbnail to perform the classification. Audio thumbnailing in Carnatic music is not an explored area of research. We would like to extend our work on thumbnails towards mrdangam strokes identification and classification of all 175 talas in Carnatic music.

ACKNOWLEDGEMENTS

We express our heartfelt gratitude to Comp Music Team for providing Carnatic music dataset. I would like to thank my Mother who supported selflessly while writing this paper. We would like to thank everyone who supported us in this endeavour.

REFERENCES

[1] Rowell, Lewis. "Paradigms for a Comparative Mythology of Music." Journal of the Indian Musicological Society 18.2 (1987): 14.

[2] Music, Robert Morris. "Sets, Scales and Rhythmic Cycles A Classification of Talas in Indian Music."

[3] Venkatarathnam, R. "Musicometric Dynamism of the Jatis." Journal of the Indian Musicological Society 13.1 (1982): 20.

[4] Kan, Alex, et al. "A Comparison of Machine Learning Approaches to Classify Tala COMP 562." (2017).

[5] A. Srinivasamurthy and X. Serra, "A supervised approach to hierarchical metrical cycle tracking from audio music recordings," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, 2014, pp. 5217-5221, doi: 10.1109/ICASSP.2014.6854598.

[6] Gulati, Sankalp, et al. "Automatic tonic identification in Indian art music: approaches and evaluation." Journal of New Music Research 43.1 (2014): 53-71.

[7] M. Muller, N. Jiang, and P. Grosche, “A robust fitness measure for capturing repetitions in music recordings with applications to audio thumbnailing,” IEEE Transactions on audio, speech, and language processing, vol. 21, no. 3, pp. 531–543, 2012.

[8] M. A. Hearst, “Support vector machines,” IEEE In-410 telligent Systems, vol. 13, no. 4, p. 18–28, 1998. 411 [Online]. Available: https://doi.org/10. 412 1109/5254.708428

[9] L. Feng, S. Liu, and J. Yao, “Music genre 414 classification with paralleling recurrent convolution-415 al neural network,” arXiv preprint ar-416 Xiv:1712.08370, 2017

[10] Porter, Alastair, Mohamed Sordo, and Xavier Serra. "Dunya: A system for browsing audio music collections exploiting cultural context." Britto A, Gouyon F, Dixon S. 14th International Society for Music Information Retrieval Conference (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 101-6.. International Society for Music Information Retrieval (ISMIR), 2013.

[11] Friedrichs, Frauke, and Christian Igel. "Evolutionary tuning of multiple SVM parameters." Neurocomputing 64 (2005): 107-117.

[12] Lameski, Petre, et al. "SVM parameter tuning with grid search and its impact on reduction of model over-fitting." Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Springer, Cham, 2015. 464-474.

[13] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.

[14] McFee, Brian, et al. "librosa: Audio and music signal analysis in python." Proceedings of the 14th python in science conference. Vol. 8. 2015.

[15] Jiménez, Álvaro Barbero, Jorge López Lázaro, and José R. Dorronsoro. "Finding optimal model parameters by discrete grid search." Innovations in Hybrid Intelligent Systems. Springer, Berlin, Heidelberg, 2007. 120-127.

[16] Nitha, K. P., and E. S. Suraj. "An Algorithm for detection of Tala in Carnatic Music for Music Therapy Applications." International Journal of Research in Engineering, IT and Social Sciences, ISSN 2250-0588, Volume 09 Issue 5, May 2019, Page 416-42

[17] Sebastian, Jilt, and Hema A. Murthy. "Onset Detection in Composition Items of Carnatic Music." *ISMIR*. 2017.

[18] Guedes, Carlos, Konstantinos Trochidis, and Akshay Anantapadmanabhan. "Modeling carnatic rhythm generation: A data driven approach based on rhythmic analysis." Proceedings of the 15th Sound & Music Computing Conference. 2018.

[19] Heshi, Rushiraj, et al. "Rhythm and timbre analysis for carnatic music processing." Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics. Springer, New Delhi, 2016.

[20] Chollet, Francois. Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG, 2018.

AUTHORS

Amulya Sri Pulijala, works as Scientist in Indian Space Research Organisation. She is doing Masters in Computer Science and Engineering from International Institue of Information Technology, Hyderabad. She has done her graduation from Jawaharlal Technological University in 2013. She was awarded for the project titled “Work stealing algorithm for Multi Core NUMA Architecture”. Her interests include Music Signal Processing, Multicore architecture.

Suryakanth V Gangashetty works as Asst. Professor in International Institute of Information Technology, Hyderabad. He received his Doctorate from Indian Institute of Technology, Madras. His research interests include Speech processing, neural networks, signal processing, pattern recognition, soft computing, machine learning, image processing, natural language processing, artificial intelligence. His research has been chronicled in more than 110 journals and conference papers.