MELODIC PATTERN EXTRACTION IN LARGE COLLECTIONS OF MUSIC RECORDINGS USING TIME SERIES MINING TECHNIQUES

Sankalp Gulati[*] sankalp.gulati@upf.edu
Joan Serrà[†] jserra@iiia.csic.es
Vignesh Ishwar[*] vigneshishwar@gmail.com
Xavier Serra[*] xavier.serra@upf.edu

[*]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[†]Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain

ABSTRACT

We demonstrate a data-driven unsupervised approach for the discovery of melodic patterns in large collections of Indian art music recordings. The approach first works on single recordings and subsequently searches in the entire music collection. Melodic similarity is based on dynamic time warping. The task being computationally intensive, lower bounding and early abandoning techniques are applied during distance computation. Our dataset comprises 365 hours of music, containing 1,764 audio recordings representing the melodic diversity of Carnatic music. A preliminary evaluation based on expert feedback on a subset of the music collection shows encouraging results. In particular, several musically interesting relationships are discovered, yielding further scope for establishing novel similarity measures based on melodic patterns.

1. INTRODUCTION

The identification of repeating structures in a musical piece is fundamental to its analysis, understanding and interpretation [4]. In music information research (MIR), repeating structures at different levels are addressed, including long duration repetitions such as themes, choruses and sections [7], and short duration repetitions such as motifs and riffs [3]. However, relatively less efforts are dedicated to melodic motif discovery using audio music recordings [5], and hence, it still remains a challenging task [2]. In this paper we address this task for Carnatic music, where the melodic patterns are highly emphasized for the rāga-based characterization of the melodies [9, 12].

We demonstrate preliminary results of our data-driven unsupervised approach for melodic pattern discovery in large audio music collections. Results indicate that our approach is robust to different tonic pitches of the lead artist, non-linear timing variations, global tempo changes and added melodic ornaments. All the discovered melodic patterns (approximately 15 million) are available online [1].

[1] http://dunya.compmusic.upf.edu/motifdiscovery/

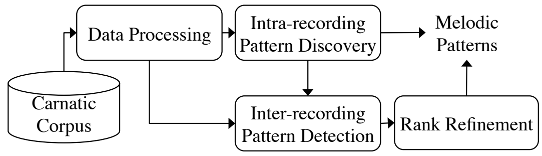

Figure 1. Block diagram of the proposed approach.

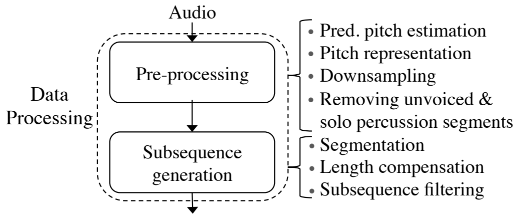

Figure 2. Block diagram of the data processing module.

2. METHOD

In Fig. 1 we show the block diagram of our proposed approach. A brief explanation of each of the block is given below.

2.1 Data Processing

All the steps involved in the data pre-processing block are listed in Fig. 2. For predominant pitch extraction we use the algorithm proposed by Salamon and Gómez [11]. We use the implementation of this algorithm available in Essentia 2.0 [1]. Pitch values are converted from Hertz to Cents using the tonic pitch of the lead artist as the reference frequency [6]. It enables us to compare melodies across different artists sung at different tonic pitches. To reduce the computational load we downsample the pitch sequence to a sampling frequency of 45 Hz. A concert of Carnatic music typically contains a solo percussion section, we therefore dedicate an extra effort to discard such segments stage using a classification-based approach.

Pitch subsequences (or melodic pattern candidates) are generated in a brute force manner by using a sliding window of length 2 s with a hop size of one sample (22 ms). Since the length of a melodic pattern can vary across repetitions, to handle it we consider 5 uniformly time scaled versions of each subsequence during the distance computation. Further, we filter out musically uninteresting melodic patterns (patterns containing a single musical note) by applying a heuristic-based criterion on the local variance of the pitch subsequences.

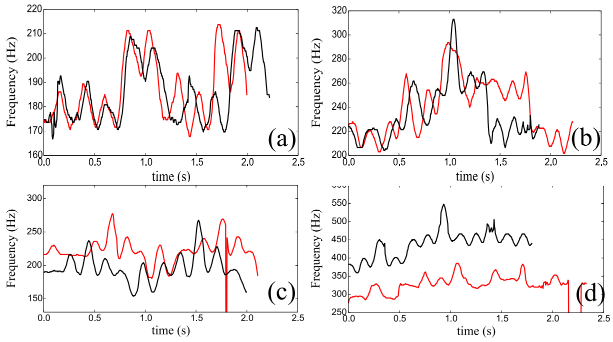

Figure 3. Examples of the discovered melodic patterns.

2.2 Intra-recording Pattern Discovery

We perform an exact pattern discovery by computing the similarity between every possible subsequence pair obtained within an audio recording. We obtain the top $N = 25$ closest subsequence pairs as seed patterns. We compute melodic similarity between two subsequences using a DTW-based distance measure [10]. We use a step condition of $\{(1, 0), (1, 1), (0, 1)\}$ and the squared Euclidean distance as the cost function. In addition, we apply the Sakoe-Chiba global constraint with the band width set to 10% of the pattern length. To make DTW distance computations tractable for a large number of subsequences we apply cascaded lower bounds as described in [8]. In total, we do nearly 1.4 trillion distance computations and discover 79,000 seed patterns.

2.3 Inter-recording Pattern Detection

We consider every seed pattern as a query and perform an exhaustive search over all the subsequences obtained from the entire audio music collection. For every seed pattern we store top $M = 200$ closest matches. We use the same similarity measure and lower bounding techniques as mentioned above. In total, we do nearly 12.4 trillion distance computations and obtain over 15 million melodic patterns.

2.4 Rank Refinement

The lower bounds we use for speeding up distance computations are not valid for any variant of DTW. However, once the top matches are found, nothing prevents us from reordering the ranked list using any variant of DTW. A DTW variant with step condition of $\{(1, 2), (1, 1), (2, 1)\}$ is used for rank refinement. Furthermore, we investigate different distance measures used in the computation of the DTW cost matrix.

3. DISCUSSION

The data used in study comprises 365 hours of music, containing 1,764 audio recordings covering diverse Carnatic music recordings. Some of the discovered patterns are shown in Fig.3. We see that our approach robustly extracts patterns in different scenarios such as large local time warpings (b), uniform scaling (a) and across different tonic pitches (c and d). An evaluation based on expert feedback showed encouraging results, where the musician found striking similarity between phrases of two different ragas, between phrases in sung melodies and the melodies played on instruments (Violin or Veena), and phrases sung by different artists. All these examples are available online [2].

It is interesting to analyze the role of lower bounds used in the distance computation. For intra-recording pattern discovery task we saved over 75% of the DTW distance computations by using cascaded lower bounds, and 99% of the DTW distance computations in the inter-recording pattern detection. Time series mining techniques can enable several computationally challenging tasks in MIR that use DTW-based similarity measures, including the discovery of melodic motifs.

4. REFERENCES

[1] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. Essentia: an audio analysis library for music information retrieval. In *Proc. of Int. Society for Music Information Retrieval Conf.*, pages 493–498, 2013.

[2] T. Collins, S. Böck, F. Krebs, and G. Widmer. Bridging the audio-symbolic gap: The discovery of repeated note content directly from polyphonic music audio. In *Audio Eng. Society Conf.: 53rd Int. Conf.: Semantic Audio*, 2014.

[3] D. Conklin. Discovery of distinctive patterns in music. *Intelligent Data Analysis*, 14:547–554, 2010.

[4] N. Cook. *A guide to musical analysis*. J.M. Dent and Sons, London, UK, 1987.

[5] R. B. Dannenberg and N. Hu. Pattern discovery techniques for music audio. *Journal of New Music Research*, 32(2):153–163, 2003.

[6] S. Gulati, A. Bellur, J. Salamon, H. G. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra. Automatic tonic identification in Indian art music: approaches and evaluation. *Journal of New Music Research*, 43(1):55–73, 2014.

[7] J. Paulus, M. Müller, and A. Klapuri. State of the art report: Audio-based music structure analysis. In *Proc. of Int. Society for Music Information Retrieval Conf.*, pages 625–636, 2010.

[8] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Addressing big data time series: mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data*, 7(3):10:1–10:31, September 2013.

[9] P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy. Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research*, 43(1):115–131, January 2014.

[10] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech, and Language Processing*, 26(1):43–50, 1978.

[11] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.

[12] T. Viswanathan and M. H. Allen. *Music in South India*. Oxford University Press, 2004.

[2] http://dunya.compmusic.upf.edu/motifdiscovery/results/349ebbb3-3cb6-4634-b485-3975ee1134d4/1953369