
A statistical generalization of music graph spectral mining

Alberto Pinto

PINTO@DICO.UNIMI.IT

Dip. di Informatica e Comunicazione, Università degli Studi di Milano, Via Comelico 39, I-20135 Milano (Italy)

Keywords: music graphs, spectral mining, complex representation, structural similarity

Abstract

Graph spectral representation of pitch class sequences provides an effective tool for mining structural music information. Here we show how it is possible to extend this approach by embedding rhythmical information into graph representation through a complex valued function defined on the graph edge set allowing for both pitch-class and rhythmical modelling through coefficients of variation. Experiments run on MIREX 05 dataset show that spectral distances calculated from this representation outperform existing methods.

1. Introduction

Content-based searching for music in large collections is getting more and more attention from information retrieval, data mining and learning fields for its practical purposes as well as the scientific challenges involved with capturing the notion of similarity on such a subjective and high dimensional domain. The first challenge is finding a good representation for music that encodes specific musical features necessary for the application, while maintaining a highly discriminative power. Graphs are an appropriate data structure to reduce a melody's complexity and degrees of freedom to a level where efficient and effective matching can be applied.

The first graph model oriented to mining similarly structured musical themes has been proposed in (Pinto & Haus, 2007); the underlying idea is that melodic sequences can be clustered by partitioning them into equivalence classes characterized by having isomorphic representative graphs. Graphs are labeled and the similarity function involves all possible isometries and

graph powers up to a fixed degree. This makes the model not suitable for online applications. An improvement of the model relies on a characterization of melodic graphs with less than 12 nodes by their second immanantal polynomials (Pinto, 2008). A strong limitation of this approach is that it provides a model for the melodic surface of a melodic sequence disregarding any rhythmical information.

In this paper we present a graph spectral modelling of melodic sequences through a statistical representation of the rhythmical surface of melodies via variation coefficients. The rest of this paper is organized as follows. More details on the construction of the graph representation and the spectra of their matrices is given in Section 2. The new representation and mining procedures are described in Section 3. Experiments on the MIREX'05 collection is given in Section 4. We conclude the paper with some final remarks and future work in Section 5.

2. Graph model

The graphs used in the approach presented in this paper reflect the musical structure on a somewhat abstract level. Specifically, the vertices of the graph represent the different pitch classes (so each graph always has 12 vertices). The sequence of notes in a melody is encoded by adding an edge between two vertices if there exists a transition between the two corresponding pitches somewhere in the melody. Let M be a melodic sequence of length $m = |M|$ and consider the sequences of pitches $\{p_s\}_{s \in I}$ and durations $\{d_s\}_{s \in I}$, $I = \{1, \dots, m\}$. Then let \mathbb{Z}_{12} be the space of pitches, or pitch classes, in the 12-tone system. We define the graph G with vertex set $V_G = \mathbb{Z}_{12}$ and edge set whose elements are the edges e_s such that

$$e_s : \begin{cases} p_s \rightarrow p_{s+1} & \text{for every couple } (p_s, p_{s+1}) \subseteq M \\ p_m \rightarrow p_1 & \text{for the couple } (p_m, p_1) \end{cases}$$

where $s = 1, \dots, m - 1$. Figure 2 shows a graph representation of the melodic sequence in Figure 1.

Appearing in *Proceedings of the 7th International Workshop on Mining and Learning with Graphs*, Leuven, Belgium, 2009.



Figure 1. An example of melodic sequence

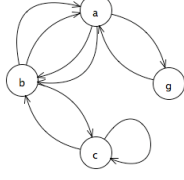


Figure 2. Graph representing the sequence in 1

To compute similarity between melodies, algebraic structures like the adjacency matrix A or the laplacian matrix $L = D - A$, where D is the degree matrix, are associated with each graph. The concept of music similarity is independent from node permutation, so the graph distance measure has to be invariant under permutations of vertexes and this supports our dealing with spectra of those matrices (Cvetkovic et al., 1995), (Sarkar & Boyer, 1996) (Umeyama, 1988). The magnitude-sorted set of eigenvalues of the laplacian matrix is very informative about the original graph structure and topology and therefore a good representation of the melody (Demirci et al., 2007). Since the laplacian spectrum can be seen as a high dimensional vector, the Euclidean distance between two spectra is computed to perform the graph matching.

3. Rhythmical extension

Our goal is to extend this idea in order to embed rhythmical information into the model. Recent approaches to graph matching (Wilson et al., 2005) suggest that a complex number can be used to represent 2-component measurement vector on a graph edge. There are, of course, many ways in which this can be done, and the key point is how this mapping has to be defined and which features are to be taken into account. At the moment, the more effective mappings for music graphs make use of Lewin's theory (Pinto & Tagliolato, 2008) which allows to model rhythmical features through different generalized interval systems (GIS). In section 4 we will compare our approach with the best complex mapping available in literature.

Here we adopt a completely different approach which aims at modelling distributions of interval-related durations with a single coefficient of variations. Let us consider the above mentioned duration sequence $\{d_s\}_{s \in I}$, $I = \{1, \dots, m\}$. By dividing each duration

value by the preceding one

$$g_s = \frac{d_s}{d_{s-1}} \quad s \in [2, \dots, m] \quad (1)$$

$$g_s = \frac{d_s}{d_m} \quad s = 1 \quad (2)$$

we get a time-stretching independent sequence $\{g_s\}_{s \in I}$, $I = \{1, \dots, m\}$ that can be used to represent rhythmical values more effectively than rough durations. Then we consider the $12 \times 12 \times m$ tensor $B = \{b_{i,j,k}\}$ depicted in Fig. 3 and defined as follows:

$$b_{i,j,k} = g_k \quad \text{iff} \quad p_k = i \wedge p_{k+1} = j \quad (3)$$

Thus we have that fibers $f_{i,j} = [b_{i,j,1}, \dots, b_{i,j,m}]$ are m -dimensional vectors containing rhythmical values corresponding to each interval (i, j) . Fibers $f_{i,j}$ can be regarded also as temporal distributions of intervals (i, j) , thus we can describe each fiber through its mean $\mu_{i,j}$ and variance $\sigma_{i,j}$ or, better, combine them into a single coefficient generating the map:

$$(i, j) \mapsto c_{i,j} = \frac{1 + \sigma_{i,j}}{1 + \mu_{i,j}}$$

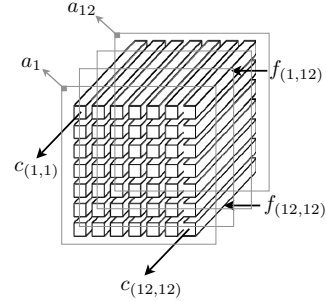


Figure 3. Tensor representation for durations. In each vertical section a_k , $k \in \{1, \dots, m\}$ element $b_{i,j,k}$ represents the duration of interval (i, j) at time k . Fibers $f_{i,j}$ contain all rhythmical values corresponding to a single interval (i, j) . Each fiber $f_{i,j}$ is then represented by its modified coefficient of variation $c_{i,j}$.

By combining interval and rhythmical representations we get the following map:

$$\mathbb{Z}_{12} \times \mathbb{Z}_{12} \times \mathbb{R} \rightarrow \mathbb{C}$$

$$(i, j, c_{i,j}) \mapsto t_{i,j} = c_{i,j} \cdot e^{\frac{i\pi}{6}|i-j|}$$

generating the representative matrix $T = \{t_{i,j}\}_{i,j \in \{1, \dots, 12\}}$ that can be regarded as an adjacency matrix of a simple graph $G(M)$ whose edges (i, j) have been weighted by coefficients $t_{i,j}$. Thus we can replace the adjacency matrix A with T and calculate the signature of the laplacian matrix of $L = T - D(T)$.

4. Experimentation

MIREX'05 is a subset of RISM A/II consisting of about 500 monophonic incipits of about 10-40 notes that has become a standard dataset for the evaluation of melodic similarity methodologies. For the MIREX 2005 competition 11 incipits were chosen as queries and a ground truth was established by 35 music experts (Typke et al.,), who estimated which incipits of the subset were melodically similar to each of the proposed queries.

As explained in Section 2 the methodology consists in the evaluation of the spectral signature ($\in \mathbb{C}^{12}$) of each representative graph followed by a nearest neighbor search. To each melody M (and to each query Q) in the database have been associated spectral vectors $z(M) \in \mathbb{C}^{12}$, one for each graph model. For each query we evaluated the euclidean distances between $z(M)$ and $z(Q)$ of the patterns from Q .

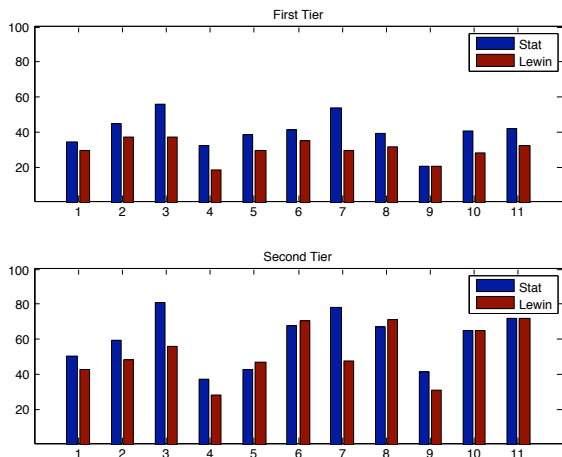


Figure 4. First and second tiers for the 11 queries.

The results obtained by the retrieval technique for the different graph models are presented in Figure 4. They represent first and second tiers, i.e. the number of relevant patterns within the first n and $2n$ retrieved patterns, over the total number of relevant (ground truth) patterns. Results have been compared with the $w_{Meandur} + e^{i\frac{\pi}{6}} w_{pc}$ representation (Pinto & Tagliolato, 2008) based on Lewin's GIS theory, which proved to outperform analogous rhythm-and-pitch representations.

5. Conclusions

We presented a novel graph spectral representation through the introduction of rhythmical information into graph edge labels. We provided a complex map

which associates to each interval a coefficient reflecting the variation of the duration distribution for that specific interval throughout the melody. The evaluation has been performed on the MIREX'05 dataset showing that spectral distances calculated from this representation outperform the best GIS-based models, thus better fitting the perceptual distance between melodies. Future work deals with the study of new tensor maps and first results in this direction are very promising. We expect this approach to be a starting point of new statistical and tensor-based learning methods for mining music graphs.

References

- Cvetkovic, D., Doob, M., & Sachs, H. (1995). *Spectra of graphs: Theory and application*. Heidelberg and Leipzig: JA Barth Verlag.
- Demirci, M. F., van Leuken, R. H., & Veltkamp, R. C. (2007). Indexing through laplacian spectra. *Computer Vision and Image Understanding, Accepted for publication*.
- Pinto, A. (2008). Mining music graphs through immanantal polynomials. *In Proceedings of the 6th International Workshop on Mining and Learning with Graphs*.
- Pinto, A., & Haus, G. (2007). A novel xml music information retrieval method using graph invariants. *ACM Transactions on Information Systems, 25*, 19.
- Pinto, A., & Tagliolato, P. (2008). A generalized graph-spectral approach to melodic modeling and retrieval. *Multimedia Information Retrieval* (pp. 89–96).
- Sarkar, S., & Boyer, K. L. (1996). Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *CVPR '96: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)* (p. 458). Washington, DC, USA: IEEE Computer Society.
- Typke, R., den Hoed, M., de Nooijer, J., Wiering, F., & Veltkamp, R. C. A ground truth for half a million musical incipits. *Journal of Digital Information Management, 3*, 34–39.
- Umeyama, S. (1988). An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell., 10*, 695–703.
- Wilson, R., Hancock, E., & Luo, B. (2005). Pattern vectors from algebraic graph theory. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27*, 1112–1124.