Automatic Music Genre Classification Using Bass Lines

Umut Şimşekli
Department of Computer Engineering,
Boğaziçi University 34342 Bebek, İstanbul, Turkey
umut.simsekli@boun.edu.tr

Abstract

A bass line is an instrumental melody that encapsulates both rhythmic, melodic, and harmonic features and arguably contains sufficient information for accurate genre classification. In this paper a bass line based automatic music genre classification system is described. "Melodic Interval Histograms" are used as features and k-nearest neighbor classifiers are utilized and compared with SVMs on a small size standard MIDI database. Apart from standard distance metrics for k-nearest neighbor (Euclidean, symmetric Kullback-Leibler, earth mover's, normalized compression distances) we propose a novel distance metric, perceptually weighted Euclidean distance (PWED). The maximum classification accuracy (84%) is obtained with k-nearest neighbor classifiers and the added utility of the novel metric is illustrated in our experiments.

1. Introduction

Along with the rapid growth of the music databases, music information retrieval (MIR) became a popular topic in computer science. Automatic music genre recognition is one of the important tasks in MIR. Correct classification of music pieces with respect to their genres would be very useful for internet music search engines, musicologist or listeners. It would also be quite useful for interactive computer music systems since music genre encapsulates semantic information about the given music piece.

Previously, McKay and Fujinaga reported very high classification accuracy on their 3-root 9-leaf MIDI data set. Their accuracy was 98% on root genres and 90% for leaf genres [7]. Using the same data set, Çataltepe et al. obtained 93% accuracy for the root genres [1]. However classification of polyphonic MIDI data is not practical. In order to make use of a music genre classifier which only works on polyphonic MIDI, one would need all the MIDI recording of the corresponding audio files. This requires an accurate polyphonic transcription system and arguably this is a much harder problem than genre classification. In order to circumvent the polyphonic transcription problem, we focus on a bass line based music genre classification system.

A bass line is an instrumental melody, which is played by a low-pitched instrument such as electric bass or double bass. In most musical styles the bass lines form a bridge between the melody and the rhythm section. Hence they encapsulate both rhythmic, melodic, and harmonic information in a simple, monophonic melody. Our hypothesis is that the bass lines are sufficiently rich source of information for accurate genre identification and in this paper we will test this assumption on MIDI data.

We believe that this approach is useful as there exists already quite accurate systems which extract the bass lines from polyphonic audio [9]. Hence, a natural next step of our work is to use bass line transcription as a preprocessing step in genre classification.

The rest of the paper is organized as follows: in Section 2, the MIDI database is described. In Section 3 and Section 4, the methodology is described. In Section 5, the results are presented. Section 6 concludes this paper.

2. Database Description

In this study we used McKay and Fujinaga's MIDI data set which consists of 3 root and 9 leaf classes. As opposed to [7] and [1] we used a different genre taxonomy. Here Jazz, Rhythm & Blues, and Rock are the root genres where each root genre is divided into three leaf genres as shown in Table 1. The bass lines were manually extracted from the MIDI files. On the other hand each leaf genre in the data set contains 25 MIDI recordings. 80% of these 225 files were used for training and the remaining 20% was used for testing.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.1006

Table 1. The genre taxonomy.

Jazz	Rhythm & Blues	Rock
Bebop	Blues Rock	Hard Rock
Swing	Funk	Metal
Bossa Nova	Rock'n Roll	Alter. Rock

3. Feature Extraction

There are several feature sets which were used in various studies. Dannenberg et al. used low-level features on MIDI data such as mean and standard deviations of key number, duration, duty factor, pitch, and volume [2]. Tzanetakis et al. used pitch histograms where each histogram represented the pitch distribution of the given recording [11]. In [6], McKay presented several high level features where these features almost cover all the features those were presented in [2] and [11].

In view of the fact that we are dealing with monophonic data, pitch histograms could be useful since they capture some information about the harmonic structure of the piece. However, these histograms are key (tonality) dependent, which means two identical songs that are played in different keys would have completely different pitch histograms. In order to have key independent features, "melodic interval histograms" were used in this study. Apart from pitch histograms, melodic interval histograms reflect the information about the order in which notes are played. Each bin of a melodic interval histogram is indexed by a number which determines the number of semitones between two adjacent notes in the bass line [6]. In this study we normalized the histograms where magnitude of a bin represents the fraction of the melodic interval of the corresponding bin.

When the melodic interval histograms of a bass line is considered, one can expect a peak at the bin with label zero due to the nature of the bass lines. Two melodic interval histograms which belong to different genres are presented in Figure 1. The main assumption here is different genres would have different melodic interval histograms of bass lines. In the next section several distance metrics will be explored in order to measure how "different" two melodic interval histograms are.

4. Classification

As mentioned in Section 2, the data set has a systematic hierarchy. In order to make use of this property, the classification was performed hierarchically. In our case, the root genre of a recording is determined firstly. Then the leaf genres which belong to the previously determined root genre is determined. Hence we will need four different classifiers: one handling the root genres, the remaining three handling the leaf genres. Moreover, hierarchical classification also increased the genre classification accuracy in [7].

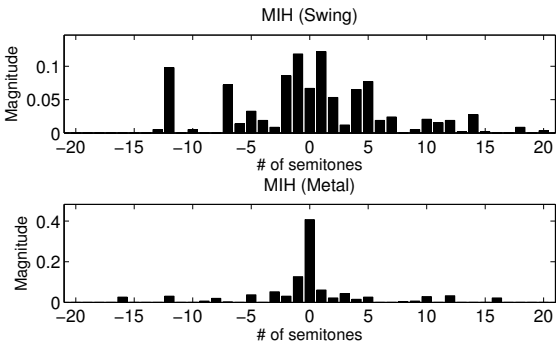

Figure 1. Typical MIHs of two genres.

In this study the k-nearest neighbor classifiers were investigated. In the k-nearest neighbor algorithm -one of the simplest machine learning algorithms- the label of a test instance is determined with respect to its k-nearest neighbors. In this context the "nearness" is determined by a distance metric and five different distance metrics were explored in order to compute the distance between two melodic interval histograms.

4.1. The Euclidean Distance

The Euclidean distance is a very well-known and popular distance metric. In d dimensions, the Euclidean Distance between two points P and Q is computed as follows:

$$\mathrm{EUC}(P,Q) = \sqrt{\sum_{i=1}^{d} (p_i - q_i)^2}.$$

4.2. The Earth Mover's Distance

The Earth Mover's Distance (EMD) is a metric for measuring the dissimilarity between two distributions [8]. The EMD measures the minimum amount of "work" needed to transform one distribution into the other. Let $P = \{p_1, p_2, ..., p_d\}$ and $Q = \{q_1, q_2, ..., q_d\}$ be two different melodic interval histograms and c_{ij} be the distance between p_i and q_j. In this study c_{ij} is computed by using the Euclidean distance. The aim is to find f_{ij} which minimize the total cost:

$$\sum_{i=1}^{d}\sum_{j=1}^{d} c_{ij} f_{ij},$$

subject to some constraints where f_{ij} is the flow between p_i and q_j. This problem can be solved via linear programming. Given all f_{ij} the EMD is defined as:

$$\text{EMD}(P,Q) = \frac{\sum_{i=1}^{d}\sum_{j=1}^{d} c_{ij} f_{ij}}{\sum_{i=1}^{d}\sum_{j=1}^{d} f_{ij}}.$$

Although the EMD was first designed for image processing applications, it has been also used for measuring the music similarity [5], [10].

4.3. The Kullback-Leibler Divergence

In the context of Information Theory, the Kullback-Leibler divergence is a measure of the difference between two probability distributions. For two distributions P and Q, KL divergence is defined as

$$\text{KL}(P||Q) = \sum_{i=1}^{d} p_i \log \frac{p_i}{q_i}.$$

However the KL divergence is not a distance metric since it is not symmetric. In order to overcome this problem KL_2 is defined as follows:

$$\text{KL}_2(P||Q) = \text{KL}(P||Q) + \text{KL}(Q||P).$$

4.4. The Normalized Compression Distance

The Normalized Compression Distance (NCD) is a distance metric which is based on non-computable Kolmogorov complexity. From the NCD perspective two bass lines are similar if we can compress one of the melodic interval histogram in the other. Despite the Kolmogorov complexity is not computable, it is possible to approximate it by using compression methods such as the Ziv-Lempel algorithm [1]. The NCD between two melodic interval histograms is defined as follows:

$$\text{NCD}(P,Q) = \frac{\max\{\text{K}(P|Q), \text{K}(Q|P)\}}{\max\{\text{K}(P), \text{K}(Q)\}}.$$

where $\text{K}(P)$ is the Kolmogorov complexity of P and $\text{K}(P|Q)$ is the conditional Kolmogorov complexity. As in proposed in [3] the conditional Kolmogorov complexity is estimated by the following formula:

$$\text{K}(P|Q) \approx \text{K}(PQ) - \text{K}(P).$$

The Kolmogorov complexity is estimated by the compressed length of P, and $\text{K}(PQ)$ is the approximate Kolmogorov complexity of concatenation of P and$Q$. Besides various types of applications, NCD showed good performance at measuring musical similarity [1], [4].

4.5. Perceptually Weighted Euclidean Distance

Consonance is a psychoacoustics term which is basically the harmony between two notes that are played simultaneously. When played together, two notes that sound "good" are called consonant and they are called dissonant if they sound unpleasant. Figure 2 shows the different levels of consonance.

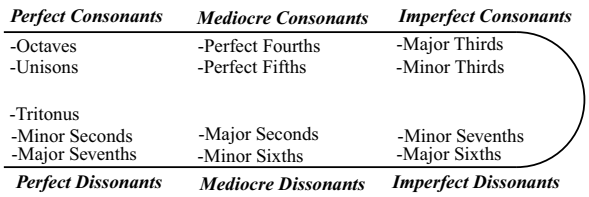

Figure 2. Levels of consonance.

The idea behind the proposed distance is to weight the Euclidean distance by making use of consonance and dissonance. Despite the bass lines are monophonic we can assume that the song has the consonant or dissonant intervals if two consecutive bass notes have that interval too. Hence before computing the Euclidean distance, we can weight the bins of the melodic interval histograms with respect to the corresponding bin's consonance level. Here one should have six different weights for the consonance and dissonance classes. Hence the PWED between two melodic interval histograms, P and Q is computed as follows:

$$\text{PWED}(P,Q) = \sqrt{\sum_{i=1}^{d} c_i (p_i - q_i)^2}.$$

where all c_i are non-negative. In order to decrease training complexity, the weights were grouped in three classes: C_1 for perfect consonants and dissonants, C_2 for mediocre consonants and dissonants, C_3 for imperfect consonants and dissonants. Note that all $c_i \in \{C_1, C_2, C_3\}$. On the other hand, since we cannot exactly compute the c_i values that maximizes the overall accuracy, a small subset of $\mathbb{R}^3$ was explored.

5. Results

Before searching for the distance metric and parameters that give the best accuracy, we explored the overall performance of all k-NN configurations. In order to do that, firstly the root accuracies were obtained by using all configurations. Then the leaf accuracies were obtained as if the root labels were correctly classified. The maximum root and leaf accuracies achieved are given in Table 2.

Table 2. Maximum k-NN accuracies.

	Root Acc. (%)	**Leaf Acc.** (%)
ED	97.78	84.44
EMD	93.33	80.00
KL_2D	93.33	80.00
NCD	75.56	66.67
PWED	100.00	86.67

In order to maximize the overall accuracy all the distance metrics and all reasonable k values were explored. The maximum accuracy was 100.00% for the root labels and 84.44% for the leaf labels. The k-NN configurations which yielded the best results are given as follows.

Table 3. The best k-NN configurations.

Type	**D. Metric**	**Parameters**
Root	PWED	$k_{root} = 19$ $C^{1:3}_{root} = (0.5, 2.0, 0.5)$
Leaf (Jazz)	EMD	$k_{jazz} = 1$
Leaf (R&B)	PWED	$k_{rnb} = 14$ $C^{1:3}_{rnb} = (2.0, 0.5, 0.5)$
Leaf (Rock)	NCD	$k_{rock} = 11$

In order to compare the success of the k-NN classifiers, Support Vector Machines (SVMs) were also tested on the same data set. Apart from the kernels that are commonly used in SVMs (linear, quadratic, polynomial, Gaussian), a PWED based kernel was also used by utilizing the three "kernelization" methods that was proposed in [12]. Although the maximum accuracy was achieved with the SVM with the linear kernel, the accuracy was still lower than the accuracy that was achieved with the k-NN classifier. The overall results are shown in Figure 3.

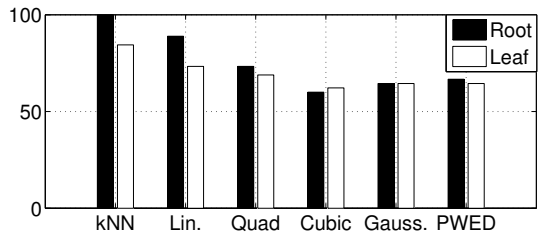

Figure 3. k-NN and SVM accuracies.

6. Discussions and Conclusion

In this study an automatic music genre classification system is described. Apart from the previous works, we simplified the genre recognition problem and discarded all the instrument parts except the bass lines. The main idea in this study is that the bass lines are sufficiently rich source of information for accurate genre identification. We tested this assumption and obtained encouraging classification accuracy (84%) with the k-NN classifiers and the added utility of the novel metric (PWED).

7. Acknowledgements

We would like to thank Cory McKay and Carlo Tomasi for sharing their data set and software. We would also want to thank A. Taylan Cemgil for suggestions. This work was supported by the MS scholarship from TÜBİTAK.

References

[1] Z. Cataltepe, Y. Yaslan, and A. Sonmez. Music genre classification using midi and audio features. *JASP*, 2007:150–150, 2007.

[2] R. B. Dannenberg, B. Thom, and D. Watson. A machine learning approach to musical style recognition. In *ICMC*, 1997.

[3] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitanyi. The similarity metric. *IEEE TIT*, 50:3250–3264, 2004.

[4] M. Li and R. Sleep. Melody classification using a similarity metric based on kolmogorov complexity. In *SMC*, 2004.

[5] B. Logan and A. Salomon. A music similarity function based on signal analysis. In *ICME*, 2001.

[6] C. McKay. Automatic genre classification of midi recordings. Master's thesis, McGill University, 2004.

[7] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *ISMIR*, 2004.

[8] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *ICCV*, 1998.

[9] M. Ryynänen and A. Klapuri. Automatic bass line transcription from streaming polyphonic audio. In *ICASSP*, 2007.

[10] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. V. Oostrum. Using transportation distances for measuring melodic similarity. In *ISMIR*, 2003.

[11] G. Tzanetakis, A. Ermolinskiy, and P. R. Cook. Pitch histograms in audio and symbolic music information retrieval. In *ISMIR*, 2002.

[12] A. Zamolotskikh and P. Cunningham. An assessment of alternative strategies for constructing emd-based kernel functions for use in an svm for image classification. In *CBMI*, 2007.