Egyptian Informatics Journal (2012) **13**, 59–63

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

Exploring different approaches for music genre classification

Antonio Jose Homsi Goulart, Rodrigo Capobianco Guido *, Carlos Dias Maciel

Received 31 December 2011; revised 6 March 2012; accepted 12 March 2012
Available online 18 April 2012

KEYWORDS

Music genre classification;
Entropy;
Fractals;
Wavelets;
SVMs

Abstract In this letter, we present different approaches for music genre classification. The proposed techniques, which are composed of a feature extraction stage followed by a classification procedure, explore both the variations of parameters used as input and the classifier architecture. Tests were carried out with three styles of music, namely blues, classical, and lounge, which are considered informally by some musicians as being "big dividers" among music genres, showing the efficacy of the proposed algorithms and establishing a relationship between the relevance of each set of parameters for each music style and each classifier. In contrast to other works, entropies and fractal dimensions are the features adopted for the classifications.

1. Introduction

Lots of facts make automatic music genre classification (AMGC) intelligent systems vital nowadays. The ease of downloading and storing music files on computers, the huge availability of albums on the Internet, with free or paid download, peer-to-peer servers and the fact that nowadays artists deliberately distribute their songs on their websites, make music database management a must. Another recent tendency is to consume music via streaming, raising the popularity of on-line radio stations that play similar songs based on a genre preference. In addition, browsing and searching by genre on the web and smart playlists generation choosing specific tunes among gigabytes of songs on personal portable audio players are important tasks that facilitate music mining.

On the other hand, music genre classification is, as described ahead, an ambiguous and subjective task. Also, it is an area of research that is being con-tested, either for low classification accuracy or because some say that one is not able to classify genres that does not even have clear definitions [1–3].

End users are nonetheless already accustomed to browse both physical and on-line music collections by genre, and this approach is proven to be at least reasonably effective. Particularly, a recent survey [1], for example, found that end users are more likely to browse and search by genre than by recommendation, artist similarity or music similarity, although these

* Corresponding author.
E-mail address: guido@ieee.org (R.C. Guido).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.
http://dx.doi.org/10.1016/j.eij.2012.03.001

Production and hosting by Elsevier

alternatives were each popular as well. Another study [2] shows that genre is so important to listeners that the style of a piece can influence their liking for it more than the piece itself. Finally [3], shows that categorization in general plays an essential role in music appreciation and cognition.

Examining the works described in Section 2, plus people's impression in general, we observed that there is no claim against the fact that the degree of irregularity noted in a certain song may be an indication of its genre. Further-more, the same holds true when considering the distribution of information in it, i.e., a classical song, for instance, seems to have more ''information'', or content, than a child melody in the same interval of time. Therefore, the use of fractal dimension and entropy, which represent those properties of a certain signal, are valid hypotheses. Based on this point-of-view, we investigated their performance for AMGC.

The remainder of this work is organized as follows. Section 2 presents a re-view on literature about music genre classification techniques, covering the state-of-the-art in the field. The proposed approach is described with details in Section 3. Section 4 lists the tests carried out with different classification schemes, input parameters, and music styles that we adopted. Lastly, useful comments and conclusions are included in Section 5, demonstrating that prominent results were achieved, strongly stimulating further research in this area.

2. Literature review

Mckay and Fujinaga [4] elaborated a paper on why should researchers continue efforts to enhance the area of AMGC. The issues they point out are related to ambiguity and subjectivity in the classifications and the dynamism of music styles. It takes a lot of expertise and time to manually classify recordings, and also there is limited agreement among human annotators when classifying music by genre. Very few genres have clear definitions and there is often significant overlap among them. Also, classifications tend to be by artist or album rather than by individual recordings, and metadata found in mp3 tags tend to have unreliable annotations. Finally, new genres are introduced regularly, and the understanding of existing genres changes with time.

The ground-breaking work of Dannenberg et al. [5], based on naive bayesian and neural network approaches, identifies one out of four styles of a musician improvisation. They were testing a performer's ability to consistently produce intentional and different styles. A database was elaborated to train the classifiers, and an accuracy of 98% was achieved when classifying among four styles. When using eight classifiers, trained to return ''yes'' or ''no'' for eight different styles, they got an overall accuracy of 77–90%.

Another classic work in the area is the one of Tzanetakis and Cook [6]. They proposed three different feature sets to represent timbral texture, rhythmic and pitch content. Short-time Fourier Transform (STFT), Mel-frequency Cepstral Coefficients (MFCCs), Wavelet Transform (WT) [7], and some additional parameters were used to obtain feature vectors. With these vectors, they could train statistical pattern recognition classifiers such as simple Gaussian, Gaussian Mixture Model, and k-Nearest Neighbor [7], by using real world audio collections. They achieved correct classifications of 61% for 10 musical genres.

Li et al. [8] worked on a comparative study between timbral textural, rhythmic content features and pitch content features versus features based on Daubechies Wavelet Coefficient Histograms (DWCHs). For the classifications, they used Support Vector Machines (SVMs), Linear Discriminant Analysis (LDA) and some other learning methods. They also tested the use of One-Against-All (OAA) and Round-Robin (RR) approaches. They used both first seconds of and middle parts of musics to carry out tests. The best overall accuracy (74.2%) was achieved when using DWCH features and an SVM classifier based on the OAA approach, being this test carried out with middle parts of songs (seconds 31–60).

Ezzaidi and Rouat [9] proposed two methods. They divided the musical pieces into frames and then got MFCCs from averaged spectral energies. Finally, for comparison purposes, they used Gaussian Mixture Models (GMMs) [10], obtaining a maximum of 99% recognition.

Silla et al. [11] adopted multiple feature vectors that were selected from different time segments from the beginning, middle and final parts of the music, and pattern recognition ensemble approach, according to a space–time decomposition dimension. Naive-Bayes, decision trees, k Nearest-Neighbors, SVMs and Multilayer Perceptron Neural Networks were employed. The best accuracy obtained was 65.06% when using Round-Robin on Space–time ensemble.

Panagakis and Kotropoulos [12] proposed a music genre classification frame-work that considers the properties of the auditory human perception system, i.e., 2D auditory temporal modulations representing music and genre classification based on sparse representation. The accuracies they obtained outperformed any rate ever reported for the GTZAN and ISMIR2004 datasets, i.e., 91% and 93.56%, respectively.

Paradzinets et al. [13] explored acoustic information, beat-related and timbre characteristics. To obtain acoustic information they used Piecewise Gaussian Modeling (PGM) features enhanced by modeling of human auditory filter. To do so, they obtained the PGM features, then applied critical bands filter, equal loudness and specific loudness sensation. To extract the beat-related characteristics, they used wavelet transforms, getting the 2D-beat histograms. For the timbre characteristics, they collected all detected notes with relative amplitude of their harmonics and then computed their histograms. Among others issues, their results show: (i) an improvement when using perceptually motivated PGM instead of basic PGM, i.e. accuracy of 43% versus 40.6%; (ii) training different NNs for each genre is better than training only one NN with all the genres being considered, which corresponds to an average accuracy of 49.3%.

What is shown is that a lot of work is being done in the area, but most of the approaches explore the timbre texture, the rhythmic content, the pitch content, or their combinations. As illustrated above, our work explores the use of entropies and fractal dimensions, thus, eliminating the use of musical information such as harmony, melody, beat and tempo. Information theory concepts are the basis of our approach.

3. The proposed approach

Our approach consists of a feature extraction stage followed by a classification step. For the first stage of tests, we adopted feature vectors of five components each one. The features are

extracted directly from the digital music files. Particularly, each song was divided into frames of 1024 samples with 50% overlap between consecutive frames. Then, for each frame, we calculated the entropy (E) via the energy approach [14], i.e.,

$$-\sum_{i=0}^{1023} p_i \log_2(p_i) \quad (1)$$

being p_i the proportion of the total signal energy, i.e., the energy of the frame divided by the energy of the entire signal. This criterion was adopted because it turned out to be more stable than the amplitude and frequency approaches.

Once we have the entropy value of each frame, we could form the feature vector, composed by:

- *Feature 1*: average entropy of the entropies of each music frame.
- *Feature 2*: standard deviation of the entropies of each music frame.
- *Feature 3*: maximum entropy among all the entropies of each music frame.
- *Feature 4*: minimum entropy among all the entropies of each music frame.
- *Feature 5*: maximum entropy difference among consecutive frames of the music signal.

After all the tests were carried out, we adopted a sixth element for the feature vector, namely, the fractal dimension of each frame, obtained on time-domain, via the box counting method [15]. Then, new experiments were performed using the best combination of the previous parameters, including this new one, and the best classifier architecture.

For all the tests, we used 90 examples of tunes equally divided on three distinct genres, namely blues, classical and lounge music. All the songs were ripped from CDs at 44.1 kHz sampling rate, 16-bit resolution, wave format. The first stage of feature extraction was based on time analysis. The entropy values were extracted directly from the wave files. In the next step, we switched the songs samples to the frequency domain via Discrete Wavelet Transform to extract entropy values. For the final test, the fractal dimension of each frame was again obtained, on time-domain via the box counting method.

The classification stage was based on the use of combined SVMs [7]. The first type of classifier was based on the training of three SVMs. Each one was trained to return the value 1 in case of blues, classical or lounge, respectively, and −1 for the other cases. The second was trained to return 1 in case of classical music, and the third one in case of lounge music. The second type of classifier also used three SVMs, but training each one to return 1 in case of recognition of its genre, never returning −1. Time and frequency feature vectors were used in each kind of test. As we got better results with frequency values and with the second kind of SVM architecture, the fractal dimension was adopted as a sixth element in the feature vector and a fifth test was made to check if this extra information would improve the classification.

4. Tests and results

The songs were divided into training and testing stages. Using 10% of each style for training lefts 90% for the tests; training with 20% allows testing with 80%, and so on until 90% for training and 10% for testing. A song used for training was never used for testing. In each round the tunes were randomly designated for each step. Five configurations were used for the tests. The first one (results shown in Table 1) consists of time-domain features extraction and the first type of classifier. The second test (Table 2) is also performed by using time-domain, but with the second type of classifier. The third experiment (Table 3) was carried out using frequency-domain features extraction and classifier of the first type. For the fourth test (Table 4) we used frequency-domain features extraction followed by the second type of classification.

Results show us that features extracted in frequency-domain had higher accuracy in the classification. We could also notice that training each SVM with a specific genre without mentioning the others was better than teaching the classifier what is a genre and what is not, as we can observe in the tables, which show better results in bold-face. So we ran a fifth test (Table 5) by using the best combination we obtained (frequency-domain features extraction and second type of classifier), and including the fractal dimension as a sixth element in the feature vector.

Overall, we perceived that frequency-based parameters have shown better results than time-based ones. Particularly, fractal dimensions have not contributed to the classifications, worsening the results in terms of accuracy, therefore, it was not considered as being a good parameter to distinguish among music genres. Another interesting point is the fact that the proposed architecture adopted for classification, which is based on M independent SVMs, being M the number of music styles, improved the traditional classification schemes which are based on one, or a few, classifier(s).

Table 1 Classification obtained with time features and first architecture of classifier, as de-scribed in the text. In each column, we have shown between parentheses the number of songs that belong to each one of the styles. The values that appear in bold-face correspond to the higher accuracies we observed.

Training	Tests	Accuracy blues	Accuracy classical	Accuracy lounge
10% (3)	90% (27)	21/27 = 77.8%	25/27 = 92.6%	14/27 = 51.8%
20% (6)	80% (24)	22/24 = 91.8%	20/24 = 83.3%	8/24 = 33.3%
30% (9)	70% (21)	17/21 = 80.9%	17/21 = 80.9%	10/21 = 47.6%
40% (12)	60% (18)	13/18 = 72.2%	13/18 = 72.2%	9/18 = 50%
50% (15)	50% (15)	12/15 = 80%	11/15 = 73.3%	13/15 = 86.8%
60% (18)	40% (12)	7/12 = 58.3%	8/12 = 66.8%	6/12 = 50%
70% (21)	30% (9)	8/9 = 88.9%	8/9 = 88.9%	6/9 = 66.8%
80% (24)	20% (6)	3/6 = 50%	2/6 = 33.3%	1/6 = 16.8%
90% (27)	10% (3)	2/3 = 66.8%	2/3 = 66.8%	0/3 = 0%

Table 2 Classification obtained with time features and second type of classifier.

Training	Tests	Accuracy blues	Accuracy classical	Accuracy lounge
10% (3)	90% (27)	10/27 = 37%	25/27 = 92.6%	23/27 = 85.2%
20% (6)	80% (24)	9/24 = 37.5%	19/24 = 79.2%	16/24 = 66.7%
30% (9)	70% (21)	13/21 = 61.9%	15/21 = 71.4%	20/21 = 95.2%
40% (12)	60% (18)	11/18 = 61.1%	15/18 = 83.3%	17/18 = 94.5%
50% (15)	50% (15)	7/15 = 46.8%	14/15 = 93.3%	9/15 = 60%
60% (18)	40% (12)	4/12 = 33.3%	10/12 = 83.3%	5/12 = 41.8%
70% (21)	30% (9)	3/9 = 33.3%	9/9 = 100%	4/9 = 44.5%
80% (24)	20% (6)	3/6 = 50%	6/6 = 100%	3/6 = 50%
90% (27)	10% (3)	1/3 = 33.38%	2/3 = 66.8%	3/3 = 100%

Table 3 Classification obtained with frequency-domain features and first type of classifier.

Training	Tests	Accuracy blues	Accuracy classical	Accuracy lounge
10% (3)	90% (27)	14/27 = 51.8%	23/27 = 85.2%	25/27 = 92.6%
20% (6)	80% (24)	16/24 = 66.8%	16/24 = 66.8%	19/24 = 79.2%
30% (9)	70% (21)	18/21 = 85.8%	10/21 = 47.6%	13/21 = 61.9%
40% (12)	60% (18)	12/18 = 66.8%	15/18 = 83.3%	11/18 = 61.1%
50% (15)	50% (15)	10/15 = 66.8%	11/15 = 73.3%	8/15 = 53.3%
60% (18)	40% *(12)*	7/12 = 58.3%	8/12 = 66.8%	8/12 = 66.8%
70% (21)	30% (9)	9/9 = 100%	6/9 = 66.8%	7/9 = 77.8%
80% (24)	20% (6)	6/6 = 100%	4/6 = 66.8%	3/6 = 50%
90% (27)	10% (3)	3/3 = 100%	2/3 = 66.8%	3/3 = 100%

Table 4 Classification obtained with frequency-domain features and second type of.

Training	Tests	Accuracy blues	Accuracy classical	Accuracy lounge
10% (3)	90% (27)	23/27 = 85.1%	24/27 = 88.8%	24/27 = 88.8%
20% (6)	80% (24)	19/24 = 79.2%	16/24 = 66.8%	19/24 = 79.2%
30% (9)	70% (21)	17/21 = 80.9%	18/21 = 85.8%	18/21 = 85.8%
40% (12)	60% (18)	13/18 = 72.2%	16/18 = 88.8%	15/18 = 83.3%
50% (15)	50% (15)	12/15 = 80%	13/15 = 86.6%	12/15 = 80%
60% (18)	40% *(12)*	10/12 = 83.3%	11/12 = 91.8%	10/12 = 83.3%
70% (21)	30% (9)	7/9 = 77.8%	8/9 = 88.8%	7/9 = 77.8%
80% (24)	20% (6)	6/6 = 100%	4/6 = 66.8%	3/6 = 50%
90% (27)	10% (3)	3/3 = 100%	2/3 = 66.8%	3/3 = 100%

Table 5 Classification obtained with frequency-domain features fractal dimension as a new feature and second type of classifier.

Training	Tests	Accuracy blues	Accuracy classical	Accuracy lounge
10% (3)	90% (27)	18/27 = 66.8%	18/27 = 66.8%	26/27 = 96.3%
20% (6)	80% (24)	18/24 = 75%	16/24 = 66.8%	20/24 = 83.3%
30% (9)	70% (21)	16/21 = 76.2%	16/21 = 76.2%	19/21 = 90.5%
40% (12)	60% (18)	13/18 = 72.2%	11/18 = 61.1%	15/18 = 83.3%
50% (15)	50% (15)	13/15 = 86.6%	13/15 = 86.6%	15/15 = 100%
60% (18)	40% (12)	8/12 = 66.8%	11/12 = 91.8%	10/12 = 83.3%
70% (21)	30% (9)	6/9 = 66.8%	9/9 = 100%	7/9 = 77.8%
80% (24)	20% (6)	5/6 = 83.3%	6/6 = 100%	5/6 = 83.3%
90% (27)	10% (3)	3/3 = 100%	3/3 = 100%	3/3 = 100%

5. Conclusions

In this article, we described a combined algorithm for music genre classification based on some specific parameters and on a set of SVMs. Our classifier presented a maximum of 100% of accuracy, but requiring 80% of the entire database, which corresponds to 72 songs, to train it. On the other hand, when only 10% of the database was used to train it, correct recognition rates varied from 51.8% to 92.6%. Thus, although full accuracy was reached by using a considerable part of the database for training, a modest training dataset was sufficient to produce strong classification rates, i.e., the proposed ap-

proach demonstrated prominent results with a considerable ability to generalize. In terms of computational costs, the proposed frequency-based approach required an extra effort to run, due to the DWT computations, however, it presented better results, as discussed above. Anyway, both frequency-based and time-based implementations are quite fast, allowing real-time use based on Digital Signal Processors (DSPs) or Field Programmable Gate Arrays (FPGAs).

References

[1] Lee JH, Downie JS. Survey of music information needs, uses, and seeking behaviours: preliminary findings. In: Proceedings of the international conference on music, information retrieval; 2004.

[2] North AC, Hargreaves DJ. Liking for musical styles. Music Scientae 1997;1(1):109–28.

[3] Tekman HG, Hortacsu N. Aspects of stylistic knowledge: what are different styles like and why do we listen to them? Psychol. Music 2002;30(1):28–47.

[4] McKay C, Fujinaga I. Musical genre classification: is it worth pursuing and how can it be improved? In: 7th Int conf on music, information retrieval (ISMIR-06); 2006.

[5] Dannenberg RB, Thom B, Watson D. A machine learning approach to musical style recognition. In: Proceedings of the international computer music conference; 1997. p. 344–7.

[6] Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process 2002;10(5):293–302.

[7] Duda RO, Hart Peter E, Stork David G. Pattern classification. 2nd ed. John Wiley & Sons; 2001.

[8] Li T, Ogihara M, Li Q. A Comparative study on content-based music genre classification. In: Proceedings of the 26th annual international ACM SI-GIR conference on research and development in information retrieval. Toronto: ACM Press; 2003. p. 282–9.

[9] Ezzaidi H, Rouat J. Automatic musical genre classification using divergence and average information measures. Research report of the world academy of science, engineering and technology; 2006.

[10] Deng L, O'shaugnessy D, Deng Deng. Speech processing: a dynamic and optimization-oriented approach. Marcel Dekker; 2003.

[11] Silla Jr CN, Koerich AL, Kaestner CAA. A machine learning approach to automatic music genre classification. J Braz Comput Soc 2008;14(3).

[12] Panagakis Y, Kotropoulos C, Arce GR. Music genre classification via sparse representations of auditory temporal modulations. In: 17th European signal processing conference (EUSIPCO); 2009.

[13] Paradzinets A, Harb H, Chen L. Multiexpert system for automatic music genre classification. <http://liris.cnrs.fr/Documents/Liris-4224.pdf>. Research report; 2009.

[14] Cover T, Thomas J. Elements of information theory. 2nd ed. John Wiley & Sons; 2006.

[15] Al-Akaidi M. Fractal speech processing. Cambridge University Press; 2004.