International Journal of Semantic Computing
Vol. 3, No. 2 (2009) 183–208
© World Scientific Publishing Company

A FEATURE SELECTION APPROACH FOR AUTOMATIC MUSIC GENRE CLASSIFICATION

CARLOS N. SILLA JR.

Computing Laboratory, University of Kent
Canterbury, CT2 7NF, Kent, UK
cns2@kent.ac.uk

ALESSANDRO L. KOERICH

Pontifical Catholic University of Paraná
R. Imaculada Conceição 1155, 80215-901, Curitiba, PR, Brazil
alekoe@ppgia.pucpr.br

CELSO A. A. KAESTNER

Federal University of Technology of Paraná
Av. Sete de Setembro 3165, 80230-901, Curitiba, PR, Brazil
kaestner@dainf.ct.utfpr.edu.br

In this paper we present an analysis of the suitability of four different feature sets which are currently employed to represent music signals in the context of the automatic music genre classification. To such an aim, feature selection is carried out through genetic algorithms, and it is applied to multiple feature vectors generated from different segments of the music signal. The feature sets used in this paper, which encompass time-domain and frequency-domain characteristics of the music signal, comprise: short-time Fourier transform, Mel frequency cepstral coefficient, beat-related features, pitch-related features, inter-onset interval histogram coefficients, rhythm histograms and statistical spectrum descriptors. The classification is based on the use of multiple feature vectors and an ensemble approach, according to time and space decomposition strategies. Feature vectors are extracted from music segments from the beginning, middle and end parts of the music signal (*time-decomposition*). Despite music genre classification being a multi-class problem, we accomplish the task using a combination of binary classifiers, whose results are merged to produce the final music genre label (*space decomposition*). Experiments were carried out on two databases: the Latin Music Database, which contains 3,227 music pieces categorized into ten musical genres; the ISMIR'2004 genre contest database which contains 1,458 music pieces categorized into six popular western musical genres. The experimental results have shown that the feature sets have different importance according to the part of the music signal from where the feature vectors are extracted. Furthermore, the ensemble approach provides better results than the individual segments in most cases. For high-dimensional feature sets, the feature selection provides a compact but discriminative feature subset which has an interesting trade-off between classification accuracy and computational effort.

Keywords: Music classification; feature selection; audio processing.

1. Introduction

Music genres can be defined as categorical labels created by humans to identify or characterize the style of music. In spite of the lack of standards, assigning a genre to a music piece is difficult, due to human perception subjectiveness. However music genre is an important descriptor which is widely used to organize and manage large digital music databases and electronic music distribution (EMD) [1, 30, 42]. Furthermore, on the Internet which contains large amounts of multimedia content, musical genres are frequently used in search queries [8, 18].

Nowadays the standard procedure for sorting and organizing music content is based on meta information tags such as the ID3 tags, which are usually associated with music coded in the MPEG-1 Audio Layer 3 (MP3) audio-specific compression format [14]. The ID3 tags are a section of the compressed MP3 audio file that contains meta information about the music. This metadata includes song title, artist, album, year, track number and music genre, besides other information about the file contents. As of 2009, the most widespread standard tag formats are ID3v1 and ID3v2. Although the ID3 tags contain relevant information for indexing, searching and retrieving digital music, they are often incomplete or inaccurate. For this reason, a tool that is able to classify musical genres in an automatic fashion relying only on the music contents will play an important role in any music information retrieval system. The scientific aspect of the problem is also an issue, since automatic music genre classification (AMGC) can be posed, from a pattern recognition perspective, as an interesting research problem: the music signal is a highly dimensional complex time-variant signal and the music databases can be very large [2].

Any approach that deals with automatic music genre classification has to find an adequate representation of the music signal to allow further processing through digital machines. For such an aim, a feature extraction procedure is applied to the music signal to obtain a compact and discriminant representation in terms of a feature vector. Then, it becomes straightforward to tackle this problem as a classical classification task in a pattern recognition framework [28]. Typically a music database contains thousands of pieces from dozens of manually-defined music genres [1, 23, 35], characterizing a complex multi-class classification problem.

Results on classification, however, depend strongly on the extracted features and their ability to discriminate the classes. It has been observed that beyond a certain point, the inclusion of additional features leads to a worse rather than better performance. Moreover, the choice of features to represent the patterns affects important aspects of the classification such as accuracy, required learning time, and the necessary number of samples. Such a problem refers to the task of identifying and selecting a proper subset of original feature set, in order to simplify and reduce the effort in preprocessing and classifying, while assuring similar or higher classification accuracy than the complete feature set [3, 6].

In this paper we present an analysis of the suitability of four feature sets which are currently employed to represent music signals in the context of AMGC. To such an aim, feature selection is carried out through genetic algorithms (GA). The features employed in this paper comprise short-time Fourier transform, Mel frequency cepstral coefficients (MFCC), beat and pitch related features [42], inter-onset interval histogram coefficients (IOIHC) [13], rhythm histograms (RH) and statistical spectrum descriptors (SSD) [24, 31, 32]. We also use a non-conventional classification approach that employs ensemble of classifiers [7,16], and which is based on *time* and *space* decomposition schemes that produce multiple feature vectors from a single music signal. The feature selection algorithm is applied to the multiple features vectors allowing a comparison of the relative importance of the features according to the segment of the music signal from where it was extracted, the feature set itself, as well as an analysis of the impact of the feature selection on the music genre classification. Principal Component Analysis (PCA) procedure is also considered for comparison purposes. The experiments were carried out on two databases: ISMIR'2004 database [4, 15], and Latin Music Database (LMD) [38].

This paper is organized as follows. Section 2 presents the AMGC problem formalization and summarizes related works in feature selection. Section 3 presents the time/space decomposition strategies used in our AMGC system. Section 4 describes the different feature sets used in this work as well as the feature selection procedure based on GA. Section 5 describes the databases used in the experiments as well as the results achieved while using feature selection over multiple feature vectors from different feature sets. Finally, the conclusions are stated in the last section.

2. Problem Definition and Related Work

Sound is usually considered as a mono-dimensional signal representing the air pressure in the ear canal [33]. In digital audio, the representation of the sound is no longer directly analogous to the sound wave. The signal must be reduced to discrete samples of a discrete-time domain. Therefore, the continuous-time signal, denoted as $y(t)$, is sampled at time instants that are multiple of a quantity T, called the sampling interval. Sampling a continuous-time signal $y(t)$ with sampling interval T produces a function $s(n) = y(nT)$ of the discrete variable n, which represents a digital audio signal [33].

A significant amount of acoustic information is embedded in such a digital music signal. This spectral information can be represented in terms of features. From the pattern recognition point of view we assume that a digital music signal, denoted as $s(n)$, is represented by a set of features. If we consider d features, $s(n)$ can be represented by a d-dimensional feature vector denoted as $\bar{x}$ and represented as

$$\bar{x} = [x_1, \ldots, x_d]^T \in \Re^d \tag{1}$$

where each component $x_i \in \Re^d$ represents a vector component extracted from $s(n)$.

We shall assume that there are c possible labeled classes organized as a set of labels $\Omega = [\omega_1, \ldots, \omega_c]$ and that each digital music signal belongs to one and only one class. Considering that our aim is to classify music according to its genre, then the classification problem consists in assigning a musical genre $\omega_j \in \Omega$ which better represents $s(n)$. This problem can be framed from a statistical perspective where the goal is to find the musical genre ω_j that is most likely, given a feature vector $\bar{x}$ extracted from $s(n)$; that is, the musical genre with the largest posterior probability, denoted as $\hat{\omega}$

$$\hat{\omega} = \arg\max_{\omega_j \in \Omega} P(\omega_j|\bar{x}) \tag{2}$$

where $P(\omega_j|\bar{x})$ is the *a posteriori* probability of a music genre ω_j given a feature vector $\bar{x}$. This probability can be rewritten using Bayes' rule

$$P(\omega_j|\bar{x}) = \frac{P(\bar{x}|\omega_j)P(\omega_j)}{P(\bar{x})} \tag{3}$$

where $P(\omega_j)$ is the *a priori* probability of the musical genre, which is estimated from frequency counts in a data set. The probability of data occurring $P(\bar{x})$ is unknown, but assuming that the genre $\omega_j \in \Omega$ and that the classifier computes the likelihoods of the entire set of possible hypotheses (all musical genres in Ω), then the probabilities must sum to one

$$\sum_{\omega_j \in \Omega} P(\omega_j|\bar{x}) = 1. \tag{4}$$

In such a way, estimated *a posteriori* probabilities can be used as confidence estimates [41]. Then, we obtain the posterior $P(\omega_j|\bar{x})$ for the music genre hypotheses

$$P(\omega_j|\bar{x}) = \frac{P(\bar{x}|\omega_j)P(\omega_j)}{\sum_{\omega_j \in \Omega} P(\bar{x}|\omega_j)P(\omega_j)}. \tag{5}$$

Feature selection can be easily incorporated in this description. Assuming a subset of d' features, where $d' < d$, then $\Re^{d'}$ is a projection of $\Re^d$. Let us denote $\bar{x}'$ as a projection of the feature vector $\bar{x}$, then we want to select an adequate $\bar{x}'$ such that it simplifies the decision

$$\hat{\omega} = \arg\max_{\omega_j \in \Omega} \frac{P(\bar{x}'|\omega_j)P(\omega_j)}{\sum_{\omega_j \in \Omega} P(\bar{x}'|\omega_j)P(\omega_j)}. \tag{6}$$

Also, since $\bar{x}'$ has a lower dimension than $\bar{x}$, it can be computed faster than $\bar{x}$.

The issue of automatic music genre classification as a pattern recognition problem has been brought up in the work of Tzanetakis and Cook [42]. In this work they use a comprehensive set of features to represent a music piece, including timbral texture features, beat-related features and pitch-related features. These features have become of public use, as part of the MARSYAS framework,[a] an open software platform for digital audio applications. Tzanetakis and Cook have used Gaussian classifiers, Gaussian mixture models and k Nearest-Neighbors (k-NN) classifiers together with feature vectors extracted from the first 30 seconds of the music pieces. They have developed a database named GTZAN which comprises 1,000 samples of

[a] Music Analysis, Retrieval and SYnthesis for Audio Signals, available at http://marsyas.sourge-forge.net/

music pieces from ten music genres (classical, country, disco, hiphop, jazz, rock, blues, reggae, pop, metal). Using the full feature set (timbral + rhythm + pitch) and a ten-fold cross validation procedure, they have achieved correct music genre classification with 60% accuracy.

Most of the current research on music genre classification focuses on the development of new feature sets and classification methods [17, 21–23, 27]. A more detailed description and comparison of these works can be found in [39]. On the other hand, few works have dealt with feature selection. One of the few exceptions is the work of Grimaldi *et al.* [10, 11]. The authors decompose the original problem according to an ensemble approach, employing different feature selection procedures, such as ranking according to the information gain (IG), ranking according to the gain ratio (GR), and principal component analysis (PCA). In the experiments they have used two hundred music pieces from five music genres, together with a k-NN classifier and a five-fold cross validation procedure. The feature vector was generated from the entire music piece using discrete periodic wavelet transform (DPWT). The PCA approach proves to be the most effective feature selection technique, achieving an accuracy of 79% with the k-NN classifier. The space decomposition approach achieved 81% for both the IG and the GR feature selection procedures, showing it to be an effective ensemble technique. When applying a forward sequential feature selection based on the GR ranking, the ensemble achieved is 84%. However, no experiments have been carried out using a standard feature set, like the one proposed by Tzanetakis and Cook [42].

Fiebrink & Fujinaga [9] discuss the use of complex feature representation and the necessary computational resources to compute them. They have employed 74 low-level features available at the jAudio [20]. jAudio is a software package for extracting features from audio files as well as for iteratively developing and sharing new features. Then, these features can be used in many areas of music information retrieval (MIR) research. To evaluate feature selection in the AMGC problem they have employed a forward feature selection (FFS) procedure and also a principal component analysis (PCA) procedure. The experiments were carried out using the Magnatune database (4,476 music pieces from 24 genres) [19] and the results over a testing set indicate that accuracy rises from 61.2% without feature selection to 69.8% with FFS and 71% with PCA.

Yaslan and Cataltepe [44] have also employed a feature selection approach for music genre classification using search methods, such as forward feature selection (FFS) and backward feature selection (BFS). FFS and BFS methods are based on a guided search in the feature space, starting from an empty set and from the entire set of features, respectively. Several classifiers were used in the experiments such as linear and quadratic discriminant classifiers, Naïve-Bayes, and variations of the k-NN classifier. They have employed the GTZAN database and the MARSYAS framework for feature extraction [42]. The experimental results have shown that feature selection, the use of different classifiers, and a subsequent combination of results can improve the music genre classification accuracy.

Bergstra *et al.* [2] use AdaBoost which performs the classification iteratively by combining the weighted votes of several weak learners. The feature vectors were built from several features like fast Fourier transform coefficients, real cepstral coefficients, MFCCs, zero-crossing rate, spectral spread, centroid, rolloff and auto-regression coefficients. Experiments were conducted considering the music genre identification task and the artist identification task of the 2005 Music Information Retrieval EXchange competition (MIREX'05). The proposed ensemble approach have shown to be effective in three music genre databases. The best accuracies in the case of the music genre identification problem vary from 75.10% to 86.92%. This result allowed the authors to win the task of music genre identification in the MIREX'05 competition.

In this paper we present a different approach to analyze the suitability of different feature sets which are currently employed to represent music signals. The proposed approach for feature selection is based on genetic algorithms. The main reason for the use of genetic algorithm in feature selection instead of other techniques such as PCA, is that the use of feature selection mechanisms based on feature transformation might improve the predictive accuracy, but limits the quality of results from a musicological perspective, as it loses potentially meaningful information about which musical qualities are most useful in different contexts, as pointed out by McKay and Fujinaga [26].

3. Music Classification: The Time/Space Decomposition Approach

The assignment of a genre to a given music piece can be considered as a three step process [2]: (a) the extraction of acoustic features from short frames of the audio signal; (b) the aggregation of the features into more abstract segment-level features; and (c) the prediction of the music genre using a class decision procedure that uses the segment-level features as input. We emphasize that if we follow the classical machine learning approach, the decision procedure is obtained from the training/validation/test cycle over a labeled database [28].

The AMGC system is based on standard supervised machine learning algorithms. However, we employ multiple feature vectors obtained from the original music signal according to *time* and *space* decompositions [5, 34, 36]. We follow an ensemble approach in which the final class label for the AMGC problem is produced as follows [25]: (a) feature vectors are obtained from several segments extracted from the music signal; (b) component classifiers are applied to each one of these feature vectors, providing a set of partial classification results; (c) a combination procedure is employed to produce the final class label from these partial classifications.

3.1. *Time decomposition*

Since music is a time-varying signal, *time decomposition* is obtained by considering feature vectors extracted from different temporal parts of the music signal. In this work we employ three segments, one from the beginning, one from the middle and

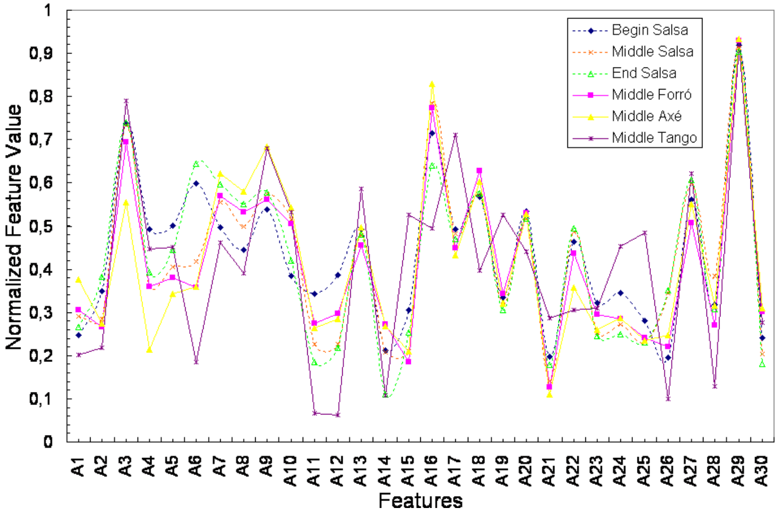

Fig. 1. Average values of over 150 music pieces of the Latin musical genre Salsa for 30 features extracted with MARSYAS from different parts of the music signal and a comparison with average values of three other Latin genres: Forró, Axé, and Tango.

one from the end part of the whole music signal. Each one of these segments is 30-second long, which is equivalent to 1,153 frames in the MP3 file format.

We argue that this procedure is adequate for the AMGC problem, since it is capable of taking into account the time variation of the music signal which is usual in many music pieces, providing a more accurate indication of the music genre. This phenomena is illustrated in Fig. 1, which presents the average values of 30 features extracted with MARSYAS framework from different music sub-intervals, obtained from 150 music pieces of the genre Salsa, Forró, Axé, and Tango. It is clear that there is a local dependence for some features. A similar behavior was found with other music genres. This local dependence may introduce some bias on the approaches that extract features from a single short segment of the music signal. This variability is a major drawback for the machine learning algorithms employed in the classification, because they have not only to deal with the traditional intra-class and inter-class variability but also with the intra-segment variability.

Finally, time decomposition also allows us to evaluate whether the features extracted from different parts of the music have similar discriminative power, aiding in the selection of the most relevant features to be considered in the task. Figure 2 illustrates the time decomposition process where feature vectors are generated from different segments of the music signal.

3.2. *Space decomposition*

Conventionally, music genre classification is a multi-class problem. However we can also accomplish the classification task using a set of binary classifiers, whose results

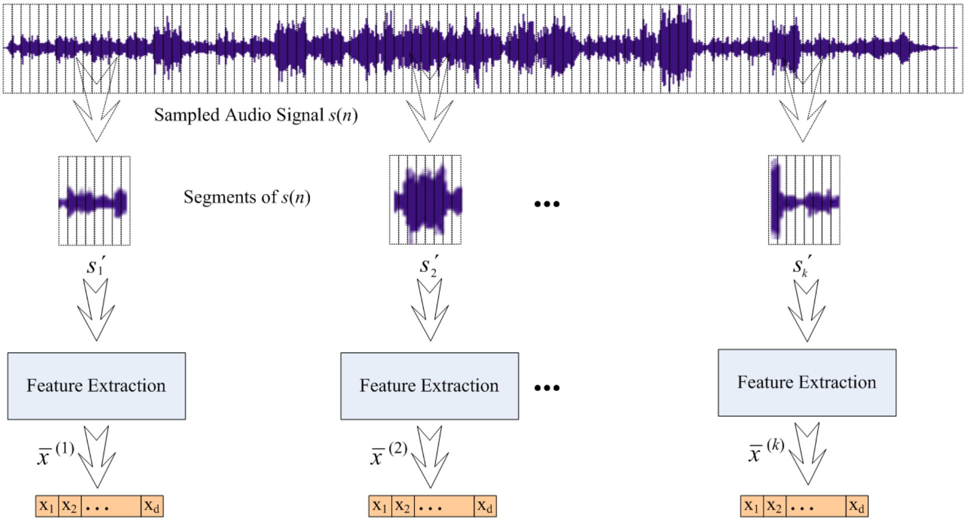

Fig. 2. An overview of the time decomposition approach: extraction of feature vectors from multiple segments of the music signal.

can be merged by a combination procedure in order to produce the final music genre label. Since different features may be used for different classes, the procedure characterizes a *space decomposition* of the feature space. The approach is theoretically justified because in the case of binary problems, the classifiers tend to be simple and effective [25].

Two main space decomposition techniques can be employed: (a) *one-against-all* (OAA) approach, where a classifier is constructed for each class and all the examples in the remaining classes are considered as negative examples of that class; (b) *round-robin* (RR) approach, where a classifier is constructed for each pair of classes, and the examples belonging to the other classes are discarded. Figures 3 and 4 illustrate these two approaches. For an m-class problem (m music genres), a set of m classifiers is generated in the OAA technique, and $m(m-1)/2$ classifiers in the RR case.

Both *time* decomposition and *space* decomposition produce a set of class label results as output of the component classifiers; they are combined according to a decision procedure to produce the final class label.

3.3. *Feature sets*

There is no accepted theory of which features are the most adequate for the music genre classification problem [1, 2]. In our previous work we have employed the MARSYAS framework for feature extraction [39, 40]. Such a framework extracts acoustic features from audio frames and aggregates them into high-level music segments [42]. We now extend our analysis to three other alternative features sets

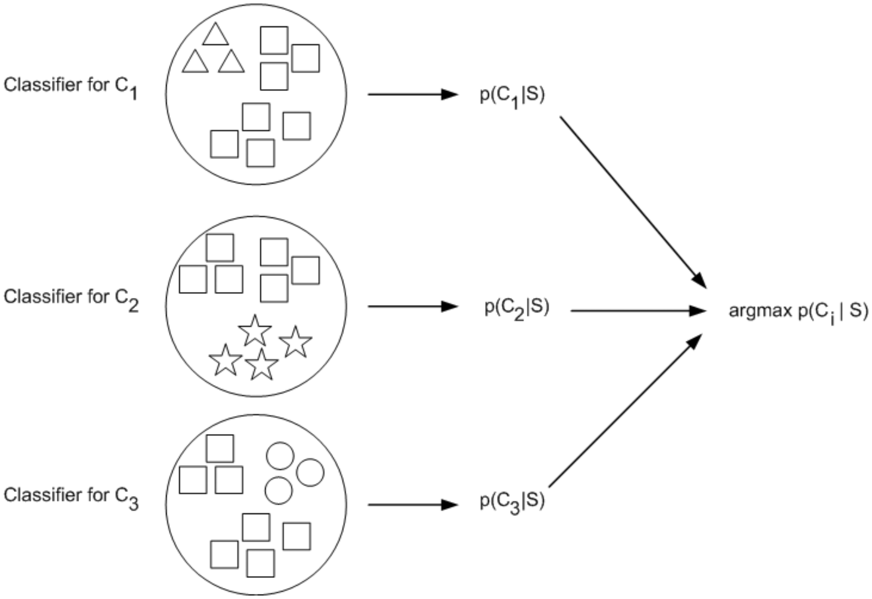

Fig. 3. Illustration of the one-against-all space decomposition approach for three classes and three classifiers.

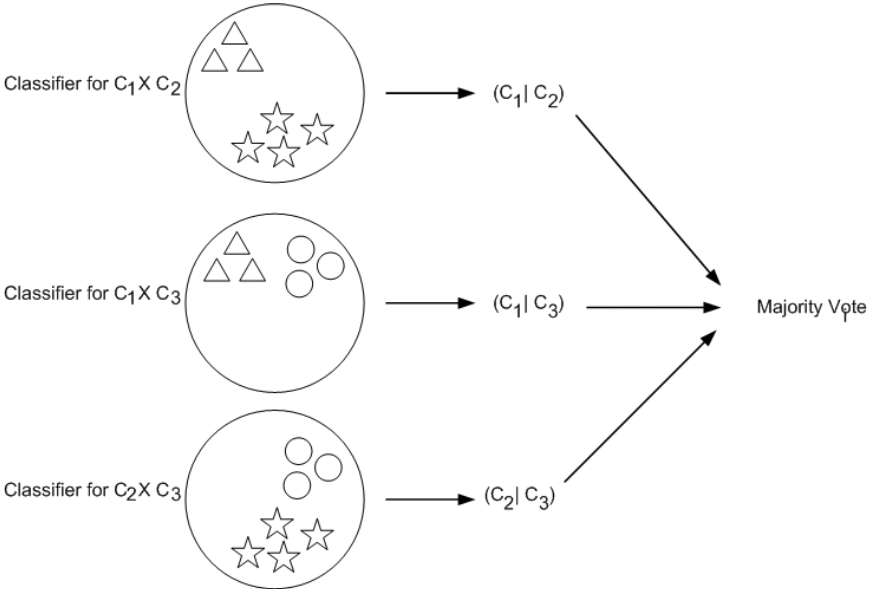
Classifier for $C_1 X C_2$
$(C_1| C_2)$
Classifier for $C_1 X C_3$
$(C_1| C_3)$
Majority Vote
Classifier for $C_2 X C_3$
$(C_2| C_3)$

Fig. 4. Illustration of the round-robin space decomposition approach for three classes and three classifiers.

that have been used to represent music signals: (a) Inset-Onset Interval Histogram Coefficients (IOIHC), that constitutes a pool of features related to rhythmic properties of sound signals computed from a particular rhythm periodic function [12,13]; (b) Rhythm Histogram (RH) features which is a set of features based on psycho-acoustical models that captures flotation on frequency bands which are critical to the human auditory system [24, 31, 32]; (c) Statistical Spectrum Descriptors (SSD) [24], which is an extension of RH features and that employs statistical measures to represent each band frequency.

3.3.1. *MARSYAS features*

The MARSYAS framework for feature extraction implements the original feature set proposed by Tzanetakis & Cook [42]. The features can be split into three groups: beat related, timbral texture and pitch related. The beat-related features (features 1 to 6) include the relative amplitudes and the beats per minute. Timbral texture features (features 7 to 25) account for the means and variance of the spectral centroid, rolloff, flux, the time zero domain crossings, the first five MFCCs and low energy. Pitch-related features (features 26 to 30) include the maximum periods and amplitudes of the pitch peaks in the pitch histograms. We note that most of the features are calculated over time intervals.

A normalization procedure is applied, in order to homogenize the input data for the classifiers: if V_{max} and V_{min} are the maximum and minimum values that appears in all dataset for a given feature, a value V is replaced by V_{new} using Eq. (7).

$$V_{new} = \frac{(V - V_{min})}{(V_{max} - V_{min})}. \tag{7}$$

The final feature vector, outlined in Table 1, is 30-dimensional (beat: 6; timbral texture: 19; pitch: 5). For a more detailed description of the features refer to [37] or [42].

3.3.2. *Inset-Onset Interval Histogram Coefficients (IOIHC)*

In the Inset-Onset Interval Histogram Coefficients (IOIHC), features are related to rhythmic properties of sound signals [12, 13]. The features are computed from a particular rhythm periodicity function (IOIH) that represents normalized salience with respect to the period of inter-onset intervals which are present in the signal. The IOIH is further parameterized by the following steps: (a) projection of the IOIH period axis from linear scale to the Mel scale, of lower dimensionality, by means of a filter; (b) computation of the IOIH magnitude logarithm; and (c) computation of the Inverse Fourier Transform, keeping the first 40 coefficients. These steps produce features analogous to the MFCC coefficients, but in the domain of

Table 1. Description of the feature vector implemented by the MARSYAS framework.

Feature #	Description
1	Relative amplitude of the first histogram peak
2	Relative amplitude of the second histogram peak
3	Ratio between the amplitudes of the second peak and the first peak
4	Period of the first peak in bpm
5	Period of the second peak in bpm
6	Overall histogram sum (beat strength)
7	Spectral centroid mean
8	Spectral rolloff mean
9	Spectral flow mean
10	Zero crossing rate mean
11	Standard deviation for spectral centroid
12	Standard deviation for spectral rolloff
13	Standard deviation for spectral flow
14	Standard deviation for zero crossing rate
15	Low energy
16	First MFCC mean
17	Second MFCC mean
18	Third MFCC mean
19	Fourth MFCC mean
20	Fifth MFCC mean
21	Standard deviation for first MFCC
22	Standard deviation for second MFCC
23	Standard deviation for third MFCC
24	Standard deviation for fourth MFCC
25	Standard deviation for fifth MFCC
26	The overall sum of the histogram (pitch strength)
27	Period of the maximum peak of the unfolded histogram
28	Amplitude of maximum peak of the folded histogram
29	Period of the maximum peak of the folded histogram
30	Pitch interval between the two most prominent peaks of the folded histogram

rhythmic periods rather than in signal frequencies. The resulting coefficients provide a compact representation of the IOIH envelope.

Roughly, lower coefficients represent the slowly varying trends of the envelope. It is our understanding that they encode aspects of the metrical hierarchy, they provide a high level view on the metrical richness, independently of the tempo. Higher coefficients, on the other hand, represent finer details of the IOIH, they provide a closer look at the periodic nature of this periodicity representation and are related to the pace of the piece at hand (its tempo, subdivisions and multiples), as well as to the rhythmical salience (i.e. whether the pulse is clearly established, this is reflected in the shape of the IOIH peaks: relatively high and thin peaks reflect a clear, stable pulse). More details on these features can be found in [13]. Feature values are normalized to the $[0, 1]$ interval. The overall procedure generates a 40-dimensional feature vector that is employed for classification, illustrated in Table 2.

Table 2. Synthetic description of the IOIHC feature vector.

Feature #	Description
1	First coefficient (related to slow trends in the envelope)
2	Second coefficient (...)
...	...
39	Thirty-ninth coefficient (...)
40	Fortieth coefficient (related to periodic nature of the signal)

3.3.3. *Rhythm Histograms (RH)*

In Rhythm Histogram (RH), the set of features is based on psycho-acoustical models that capture rhythmic and other fluctuations on frequency bands critical to the human auditory system [24, 31, 32]. The feature extraction process is composed of three stages. Initially, the specific loudness sensation on 24 critical frequency bands is computed by using a short time fast Fourier transform. Then the resulting frequency bands are grouped to the Bark scale, applying spreading functions to account for masking effects and successive transformation into the Decibel, Phon and Sone scales. The Bark scale is a perceptual scale which groups frequencies to critical bands according to perceptive pitch regions [45]. The step produces a psycho-acoustically modified Sonogram representation that reflects human loudness sensation. In the second step, a discrete Fourier transform is applied to this Sonogram, resulting in a time-invariant spectrum of loudness amplitude modulation per modulation frequency for each individual critical band. These two steps produce, after additional weighting and smoothing steps, a set of features called rhythm pattern [31, 32] indicating occurrence of rhythm as vertical bars, but also describing smaller fluctuations on all frequency bands of the human auditory range. A third step is applied in order to reduce dimensionality: it aggregates the modulation amplitude values of the 24 individual critical bands, exhibiting the magnitude for 60 modulation frequencies between 0.17 and 10 Hz [24]. Similar to the previous feature sets, feature values are normalized.

Since the complete process is applied to several audio segments, the final Rhythm Histogram feature vector is computed as the median of the individual values for each audio segment, generating a 60-dimensional feature vector, indicated in Table 3.

Table 3. Synthetic description of the Rhythm Histogram (RH) feature vector.

Feature #	Description
1	Median of magnitude in modulation frequency (0.17∼0.34 Hz)
2	Median of magnitude in modulation frequency (0.34∼0.51 Hz)
...	...
60	Median of magnitude in modulation frequency (9.93∼10.1 Hz)

3.3.4. *Statistical Spectrum Descriptors (SSD)*

In the Statistical Spectrum Descriptors (SSD) [24], the specific loudness sensation is computed on 24 Bark-scale bands, as in RH. Subsequently the statistical measures mean, median, variance, skewness, kurtosis, minimum and maximum values are computed on each of these critical bands.

The SSD feature set describes fluctuations on the critical bands and captures additional timbral information that is not covered by the previous feature set. The final feature vector for SSD is 168-dimensional and it is able to capture and describe acoustic content very well. Final feature values are normalized to $[0, 1]$. The SSD feature set is illustrated in Table 4, where the 24 Bark band edges are given in Hertz as [0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500].

3.4. *Classification, combination and decision*

In our AMGC system standard machine learning algorithms were employed as individual component classifiers. Our approach is homogeneous, that is, the very same classifier is employed in every music part. In this work we use the following algorithms: decision trees (J48), k nearest neighbor (k-NN), Naïve-Bayes (NB), multilayer perceptron neural network classifier (MLP) trained with the backpropagation momentum algorithm, and support vector machine (SVM) with pairwise classification [28]. The final classification label is obtained from all the partial classifications, according to an ensemble approach, by applying a specific decision procedure. In our case, the combination of the time and space decomposition strategies works as follows:

(1) one of the space decomposition approaches (RR or OAA) is applied to all three segments of the time decomposition approach (i.e. beginning, middle and end);

Table 4. Synthetic description of the Statistical Spectrum descriptors (SSD) feature vector.

Feature #	Description
1	Mean of the first critical band (0∼100 Hz)
2	Median of the first critical band (0∼100 Hz)
3	Variance of the first critical band (0∼100 Hz)
4	Skewness of the first critical band (0∼100 Hz)
5	Kurtosis of the first critical band (0∼100 Hz)
6	Min-value of the first critical band (0∼100 Hz)
7	Max-value of the first critical band (0∼100 Hz)
...	...
162	Mean of the twenty-fourth critical band (12000∼15500 Hz)
163	Median of the twenty-fourth critical band (12000∼15500 Hz)
164	Variance of the twenty-fourth critical band (12000∼15500 Hz)
165	Skewness of the twenty-fourth critical band (12000∼15500 Hz)
166	Kurtosis of the twenty-fourth critical band (12000∼15500 Hz)
167	Min-value of the twenty-fourth critical band (12000∼15500 Hz)
168	Max-value of the twenty-fourth critical band (12000∼15500 Hz)

(2) a local decision considering the class of the individual segment is made based on the underlying space decomposition approach: the majority vote for the RR and rules based on the *a posteriori* probability given by the specific classifier of each case for the OAA;
(3) the decision concerning the final music genre of the music piece is made based on the majority vote of the predicted genres from the three individual time segments.

Majority vote is a simple decision rule, only the class labels are taken into account and the one with more votes wins

$$\hat{\omega} = \operatorname*{maxcount}_{i \in [1,3]} \left[\arg\max_{\omega_j \in \Omega} \; P_{D_i}(\omega_j | \bar{x}^{(i)}) \right] \tag{8}$$

where i denotes the index of the segment, feature vector, and classifier and P_{D_i} denotes the *a posteriori* probability provided at the output of classifier D_i. We assume that maxcount returns the most frequent value of a multiset.

4. Feature Selection

The feature selection (FS) task is defined as the choice of an adequate subset of original feature set with the aim of simplifying or reducing the effort in the further steps, such as preprocessing and classification, while maintaining or even improving the final classification accuracy [3, 6]. In the case of the AMGC problem, feature selection is an important implementation issue, since computing acoustic features from a long time-varying signal is a time-consuming task.

Feature selection methods are often classified into two groups: the *filter* approach and the *wrapper* approach [29]. In the filter approach the feature selection process is carried out independently, as a preprocessing step, before the use of any machine learning algorithm. In the wrapper approach a machine learning algorithm is employed as a sub-routine of the system, with the aim of evaluating the generated solutions. In both cases the FS task can be modeled as an heuristic search: one must found a minimum size feature set that maintains or improves the music genre classification performance.

We emphasize that our system deals with several feature vectors, according to time and space decompositions. Therefore, the FS procedure is employed independently in the feature vectors extracted from all music segments, allowing us to compare the relative importance of the features according to the part of the music signal from where they were extracted.

The proposed approach for feature selection is based on the genetic algorithm paradigm, which recognized as an efficient search procedure for complex problems. Our procedure follows a standard GA paradigm [28].

Individuals (chromosomes) are n-dimensional binary vectors, where n is the maximum size for the feature vector (30 for MARSYAS, 40 for IOIHC, 60 for RH and 168 for SSD). They work as a binary mask, acting on the original feature

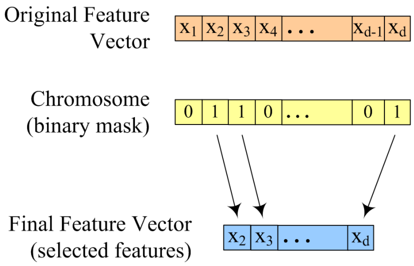

Fig. 5. The feature selection procedure for one individual in the GA procedure.

vector in order to generate the reduced final vector, composed only by the selected features, as shown in Fig. 5. Fitness of the individuals are directly obtained from the classification accuracy of the corresponding classifier, according to the wrapper approach.

The global feature selection procedure is as follows:

(1) each individual works as a binary mask for an associated feature vector: a value 1 indicates that the corresponding feature is used, 0 that it must be discarded;
(2) initial assignments of 0's and 1's are randomly generated to create initial masks;
(3) a classifier is trained, for each individual, using the selected features;
(4) the generated classification structure — for each individual — is applied to a validation set to determine its accuracy, which is considered as the fitness value of this individual;
(5) we proceed elitism to conserve the top ranked individuals; crossover and mutation operators are applied in order to obtain the next generation.

In our FS procedure we employ 50 individuals in each generation, and the evolution process ends when it converges, that is, there is no significant change in the population in the successive generations, or when a fixed maximum number of generations is achieved. The top ranked individual — the one associated to the highest accuracy in the final generation — indicates the selected feature set.

5. Experiments

This section presents the experiments and the results achieved on music genre classification and feature selection. The main goal of the experiments is to evaluate if the features extracted from different parts of the music signal have similar discriminative power for music genre classification. Another goal is to verify if the ensemble-based method provides better results than the classifiers taking into account features extracted from single segments.

Our primary evaluation measure is the classification accuracy. Experiments were carried out using a ten-fold cross-validation procedure, that is, the presented results are obtained from ten randomly independent experiment repetitions.

Two databases were employed in the experiments: the Latin Music Database (LMD) and the ISMIR'2004 database. The LMD is a proprietary database composed of 3,227 music samples in MP3 format originated from music pieces of 501 artists [37, 38]. Three thousand music samples from ten different Latin musical genres (Tango, Salsa, Forro, Axe, Bachata, Bolero, Merengue, Gaucha, Sertaneja, Pagode). The feature vectors from this database are available to researchers in the webpage www.ppgia.pucpr.br/∼silla/lmd/. In this database music genre assignment was manually made by a group of human experts, based on the human perception on how each music is danced. The genre labeling was performed by two professional teachers with over ten years of experience in teaching ballroom Latin and Brazilian dances. The experiments were carried out on stratified training, validation and test datasets. In order to deal with balanced classes, 300 different song tracks from each genre were randomly selected.

The ISMIR'2004 genre database is a well-known benchmark collection that was created for the music genre classification task of the ISMIR 2004 Audio Description contest [4, 15]. Since then, it has been used by the Music IR community. It contains 1,458 music pieces categorized into six popular western music genres: classical (604 pieces), electronic (229), jazz and blues (52), metal and punk (90) and world music (244).

5.1. *Experiments with MARSYAS features*

The initial experiments employ the MARSYAS framework features. Tables 5 to 7 present the results obtained with the feature selection procedure applied to the beginning, middle and end music segments, respectively [37]. Since we are evaluating the feature selection procedure, it is also important to measure performance without the use of any FS mechanism. Such an evaluation corresponds to the baseline (BL) column presented in the tables. Columns 3 and 4 also show the results for OAA and RR space decomposition approaches without feature selection. Columns BL + GA, OAA + GA and RR + GA present the corresponding results with the GA feature selection procedure.

We can outline some conclusions based on Tables 5 to 7: (a) GA feature selection method with the RR space-time decomposition approach produces for J48 and 3-NN better accuracy results than the other options; (b) GA FS seems to be ineffective

Table 5. Classification accuracy (%) using MARSYAS features and space decomposition for the beginning segment of the music (S_{beg}).

Classifier	BL	OAA	RR	BL + GA	OAA + GA	RR + GA
J48	39.60	41.56	45.96	44.70	43.52	**48.53**
3-NN	45.83	45.83	45.83	51.19	51.73	**53.36**
MLP	53.96	52.53	**55.06**	52.73	53.99	54.13
NB	44.43	42.76	44.43	**45.43**	43.46	45.39
SVM	—	23.63	**57.43**	—	26.16	57.13

Table 6. Classification accuracy (%) using MARSYAS features and space decomposition for the middle segment of the music (S_{mid}).

Classifier	BL	OAA	RR	BL + GA	OAA + GA	RR + GA
J48	44.44	44.56	49.93	45.76	45.09	**50.86**
3-NN	56.26	56.26	56.26	60.02	60.95	**62.55**
MLP	**56.40**	53.08	54.59	54.73	54.76	49.76
NB	47.76	45.83	47.79	50.09	48.79	**50.69**
SVM	—	38.62	**63.50**	—	32.86	59.70

Table 7. Classification accuracy (%) using MARSYAS features and space decomposition for the end segment of the music (S_{end}).

Classifier	BL	OAA	RR	BL + GA	OAA + GA	RR + GA
J48	38.80	38.42	45.53	38.73	38.99	**45.86**
3-NN	48.43	48.43	48.43	51.11	51.10	**53.49**
MLP	48.26	**51.96**	51.92	47.86	50.53	49.64
NB	39.13	37.26	39.19	**39.66**	37.63	39.59
SVM	—	28.89	54.60	—	28.22	**55.33**

for the MLP classifier, since its best results are obtained with the complete feature set; (c) in the case of the NB classifier GA FS produces the better results without space decomposition in S_{beg} and S_{end}, and with the RR approach in S_{mid}; (d) the best results for the SVM classifier are achieved with the RR approach, and GA FS increases accuracy only in the S_{end} segment. This classifier also presents the best overall result using the RR space decomposition in S_{mid} without feature selection.

Analogously, Table 8 presents global results using time and space decompositions, for OAA and RR approaches, with and without feature selection. We emphasize that this table encompasses the three music segments (beginning, middle and end).

Table 8 shows that the RR + GA method improves classification accuracy for the classifiers J48, 3-NN and NB. Also, the OAA and OAA + GA methods present similar results for the MLP classifier, and only for the SVM classifier the best results are achieved without FS. These results also indicate that space decomposition and feature selection are more effective for classifiers that produce simple separation surfaces between classes, like J48, 3-NN and NB, in contrast with the results achieved

Table 8. Classification accuracy (%) using MARSYAS features and global time and space decomposition.

Classifier	BL	OAA	RR	BL + GA	OAA + GA	RR + GA
J48	47.33	49.63	54.06	50.10	50.03	**55.46**
3-NN	60.46	59.96	61.12	63.20	62.77	**64.10**
MLP	59.43	**61.03**	59.79	59.30	60.96	56.86
NB	46.03	43.43	47.19	47.10	44.96	**49.79**
SVM	—	30.79	**65.06**	—	29.47	63.03

with the MLP and SVM classifiers, which can produce complex separation surfaces. This situation corroborates to our hypothesis on the use of space decomposition strategies.

As previously mentioned, we also want to analyze if different features sets have the same importance according to the segment from where they are extracted from the music signal. Table 9 shows a schematic map indicating the features selected in each music segment. In this table we employ a binary BME mask — for (B)eginning, (M)iddle and (E)nd time segments — where 0 indicates that the feature was not selected while 1 indicates that it was selected by the FS procedure in the corresponding time segment.

In order to evaluate the discriminative power of the features, the last column in this table indicates how many times the corresponding feature was selected in the experiments (max 15 selections). Although this evaluation can be criticized, since different features can have different importance according to the employed classifier, we argue that this counting gives an idea of the global feature discriminative power.

Table 9. Selected features (BME mask) for the MARSYAS feature set.

Feature	3-NN	J48	MLP	NB	SVM	#
1	000	001	010	101	111	7
2	000	000	010	010	011	4
3	000	001	010	011	000	4
4	000	111	010	111	001	8
5	000	000	110	101	100	5
6	111	101	111	111	110	13
7	011	110	110	000	100	7
8	001	111	110	000	111	9
9	111	111	111	111	111	15
10	110	011	111	111	111	13
11	100	001	111	001	110	8
12	011	010	111	011	111	11
13	111	011	111	111	111	14
14	001	010	101	000	011	6
15	011	111	111	111	111	14
16	111	111	111	111	111	15
17	111	100	111	111	111	13
18	111	111	111	111	111	15
19	111	010	111	111	111	13
20	011	010	110	101	101	9
21	111	111	111	101	111	14
22	111	110	111	111	111	14
23	111	111	111	100	111	13
24	011	000	111	001	011	8
25	111	011	101	111	111	13
26	000	010	100	111	111	8
27	000	111	000	101	101	7
28	111	111	011	111	111	14
29	000	100	000	000	101	3
30	000	011	000	111	000	5

For example, features 6, 9, 10, 13, 15, 16, 17, 18, 19, 21, 22, 23, 25 and 28 are important for music genre classification. We remember that features 1 to 6 are beat related, 7 to 25 are related to timbral texture, and 26 to 30 are pitch related.

5.2. *Experiments with other feature sets*

We also conduct some experiments using the alternative feature sets described in Secs. 3.3.2 to 3.3.4. Since the SVM classifier presents the best results in the previous experiments, we have limited the further experiments to this specific classifier.

Table 10 summarizes the results with all feature sets. In this Table columns are related to the employed feature set, with and without GA FS. MS stands for the application of SVM in the MARSYAS feature set, previously presented, for comparison purposes. Rows indicate the application of the SVM algorithm individually to each time segment (S_{beg}, S_{mid}, S_{end}) and also the final majority vote result, obtained from the time decomposition approach.

In general, the GA FS procedure did not improve significantly the classification accuracy for the SVM classifier, as occurred in the previous experiments. We emphasize that the SSD feature set presents superior performance in all cases. Corresponding values with GA FS in SSD are just a little below, indicating that the procedure can be useful depending on the application. One can argue if in this case we can also analyze the relative importance of the features. In the last three feature sets (IOIHC, RH and SSD) the feature vectors are composed by successive coefficients obtained from a complex transformation applied to the audio signal. This situation is different from the MARSYAS case, where most of the features have a specific semantic meaning. Therefore, we consider that carrying out a detailed analysis similar to the one in Table 9 is meaningless. On the other hand feature selection can be employed to reduce computational effort. In Table 11 we present the number of features selected by the GA in each additional experiment for the

Table 10. Classification accuracy (%) for SVM applied to alternative feature sets, with and without GA feature selection.

Segment	MS	IOIHC	RH	SSD	MS + GA	IOIHC + GA	RH + GA	SSD + GA
S_{beg}	57.43	47.30	49.80	**75.70**	57.13	46.90	48.93	74.70
S_{mid}	63.50	53.27	54.63	**82.33**	59.70	52.83	51.67	81.87
S_{end}	54.60	26.43	52.10	**79.97**	55.33	26.60	50.90	79.93
Maj vote	65.06	52.53	56.97	**84.70**	63.03	52.47	55.40	83.93

Table 11. Number and percentage of features selected in the GA feature selection experiments with SVM on the different feature sets.

Segment	MS + GA	IOIHC + GA	RH + GA	SSD + GA
S_{beg}	24 (80%)	23 (58%)	48 (80%)	99 (59%)
S_{mid}	22 (73%)	26 (65%)	47 (78%)	111 (66%)
S_{end}	24 (80%)	29 (73%)	52 (86%)	103 (62%)

different feature sets. Recall that the original feature set sizes are 30, 40, 60, and 168 for MARSYAS, IOIHC, RH and SSD respectively.

Overall, we note that from 58% to 86% of the features were selected. In the MARSYAS and RH feature sets the average percentual of features selected is roughly 80%. In the SSD feature set which, is the one with the highest dimension, on average only 62% of the features were selected. This reduction can be useful in practical applications, especially if we consider that the corresponding fall in accuracy (Table 10) is less than 1%.

5.3. *Experiments with PCA feature construction*

We conduct experiments in order to compare our FS approach based on GA with the well-known PCA feature construction procedure that is used by several authors for FS [9–11, 44]. As in the previous section, we restrict our analysis to the SVM classifier, and we use the WEKA data mining tool with standard parameters in the experiments, i.e. the new features account for 95% of the variance of the original features.

Table 12 presents the accuracy results for the SVM classifier in the Latin Music Database, for the different feature sets using PCA for feature construction. Results without FS were maintained for comparison purposes. In correspondence, Table 13 presents the number of features constructed by the PCA procedure in each additional experiment.

A comparison between the GA and the PCA feature selection methods can be done by inspecting Tables 10 and 12 (for accuracy) and Tables 11 and 13 (for the number of features). We conclude that the SSD feature set produces the best results without FS in all cases. The MS feature set is in second place. GA FS and PCA procedures produce similar results: the first one is superior for the SSD and IOIHC feature sets, and it is slightly inferior for MS and RH feature sets. In all cases the

Table 12. Classification accuracy (%) for SVM applied to all feature sets, with and without PCA feature construction.

Segment	MS	IOIHC	RH	SSD	MS + PCA	IOIHC + PCA	RH + PCA	SSD + PCA
S_{beg}	57.13	47.30	49.80	**75.70**	58.20	45.17	49.37	70.33
S_{mid}	59.70	53.27	54.63	**82.33**	62.43	49.10	53.50	77.10
S_{end}	55,33	26.43	52.10	**79.97**	60.93	23.93	52.43	74.90
Maj vote	63.03	52.53	56.97	**84.70**	65.46	48.63	57.07	79.90

Table 13. Number and percentual of features using the PCA feature construction method with SVM on different all feature sets.

Segment	MS + PCA	IOIHC + PCA	RH + PCA	SSD + PCA
S_{beg}	19 (63%)	19 (48%)	41 (68%)	45 (27%)
S_{mid}	18 (60%)	16 (40%)	43 (72%)	45 (27%)
S_{end}	19 (63%)	31 (78%)	43 (72%)	46 (27%)

ensemble approach demonstrates to be an adequate procedure: its results are better than the ones for S_{beg}, S_{mid}, S_{end} individual segments. Concerning the number of features, as expected, the PCA method produces a more compact representation, using from 27% to 78% of the original features; the GA FS, on the other hand, selects from 58% to 86% of the original features. However, as pointed out by by McKay and Fujinaga [26], from a musicological perspective the GA offers a more interesting result, as it presents which features are important to the user. On the other hand, with the PCA, this information is lost during the feature transformation process for constructing the more compact (i.e. with lower dimensionality) feature set.

5.4. *Experiments with ISMIR database*

Several experiments were conducted using the ISMIR database. In the experiments, the performance of music genre classification on the different feature sets (MARSYAS, IOIHC, RH and SSD) with and without FS is evaluated.

The results for classification accuracy are presented in Tables 14 and 15, for GA and PCA procedures, respectively. Tables 16 and 17 present the number of selected features (for the GA) and constructed features (for the PCA).

Similar to the previous experiments, the best results are produced using the SSD feature set without feature selection, except for the middle segment in which

Table 14. Classification accuracy (%) for SVM applied to all feature sets, with and without GA feature selection in the ISMIR database.

Segment	MS	IOIHC	RH	SSD	MS + GA	IOIHC + GA	RH + GA	SSD + GA
S_{beg}	66.57	45.00	57.55	**71.20**	67.38	44.94	57.70	70.08
S_{mid}	71.86	49.71	62.84	76.12	71.50	51.61	63.68	**76.48**
S_{end}	67.54	42.61	59.12	**72.72**	66.92	43.04	59.29	72.05
Maj vote	71.44	45.00	60.81	**77.21**	71.27	44.79	61.34	76.74

Table 15. Classification accuracy (%) for SVM applied to all feature sets, with and without the PCA feature construction method in the ISMIR database.

Segment	MS	IOIHC	RH	SSD	MS + PCA	IOIHC + PCA	RH + PCA	SSD + PCA
S_{beg}	66.57	45.00	57.55	**71.20**	62.61	44.71	55.91	67.11
S_{mid}	71.86	49.71	62.84	**76.12**	68.43	49.99	60.95	73.53
S_{end}	67.54	42.61	59.12	**72.72**	63.26	42.75	57.64	69.81
Maj vote	71.44	45.00	60.81	**77.21**	67.37	44.71	59.34	73.93

Table 16. Number and percentage of features selected in the GA feature selection experiments with SVM on different feature sets in the ISMIR database.

Segment	MS + GA	IOIHC + GA	RH + GA	SSD + GA
S_{beg}	25 (83%)	30 (75%)	40 (67%)	104 (62%)
S_{mid}	20 (66%)	20 (50%)	32 (53%)	94 (56%)
S_{end}	24 (80%)	18 (45%)	30 (50%)	90 (54%)

Table 17. Number and percentage of features using the PCA feature construction method with SVM on different feature sets in the ISMIR database.

Segment	MS + PCA	IOIHC + PCA	RH + PCA	SSD + PCA
S_{beg}	20 (67%)	31 (78%)	37 (62%)	43 (26%)
S_{mid}	19 (63%)	28 (70%)	36 (60%)	44 (26%)
S_{end}	19 (63%)	34 (85%)	35 (58%)	43 (26%)

SSD + GA is better; MS feature set results are in the second position. In the ISMIR database the results obtained with GA FS are slightly superior to the ones with PCA in most cases, for all the considered feature sets. Concerning the ensemble approach, in these experiments we have obtained better results using only the middle segment of the music piece. Finally, considering the number of selected features, the PCA method creates a feature set with 26% to 85% of the original feature set, while the GA FS method selects from 45% to 83% of the same set. We consider that these results are consistent with the ones presented in the previous section.

6. Concluding Remarks

In this paper we evaluate a feature selection procedure based on genetic algorithms in the automatic music genre classification task. We also use an ensemble approach according to time and space decompositions: feature vectors are selected from different time segments of the music, and one-against-all and round-robin composition schemes are employed for space decomposition. From the partial classification results originated from these views, a unique final classification label is provided.

All the experiments were conducted in a large database, the Latin Music Database, with more than 3,000 music pieces from ten musical Latin genres [37,38]. Preliminary experiments were carried out using the MARSYAS feature set. In this case we have employed several classification paradigms and heuristic combination procedures to produce the final music genre label. Additional experiments employing other feature sets such as the IOIHC feature set, the RH feature set and the SSD feature set we also conducted. For these feature sets, only the SVM classifier was employed since it was the classification algorithm that has achieved the best results in the preliminary experiments.

An extensive set of experiments were carried out to evaluate the feature selection procedure which is based on the genetic algorithm paradigm. In the proposed approach each individual works as a mask that selects the set of features to be used in the classifier construction. The fitness of the individuals is based on its classification accuracy, according to the wrapper approach. The framework encompasses classical genetic operations (elitism, crossover, mutation) and stopping criteria. Additional experiments with PCA were conducted for comparison purposes.

In the preliminary experiments, the results achieved with the feature selection have show that this procedure is effective for J48, k-NN and Naïve-Bayes classifiers; for MLP and SVM the FS procedure does not increases classification accuracy

(Tables 5 to 8). These results are compatible with the ones presented in [44]. This conclusion is also confirmed in the experiments carried out using three different feature sets. In this case, the SSD feature set, composed by a series of statistical descriptors of the signal spectrum, was the feature set that has presented by far the best results in terms of accuracy. We also conduct experiments using the ISMIR 2004 Audio Description contest dataset. The results of these experiments are, in general, consistent with those obtained with the LMD database (see Tables 14 to 17).

Another conclusion that can inferred from the initial experiments is that the MARSYAS features have different importance in the classification, according to their origin music segment (Table 9). It can be seen, however, that some features are present in almost every selection, showing they have a strong discriminative power in the classification task. In the case of the alternative feature sets (IOIHC, RH, SSD) the feature selection procedure did not increase the classification accuracy (Tables 10 to 11). We argue that this occurs because these feature sets are composed of a series of coefficients obtained from an unique signal representation, as shown in Tables 2 to 4. Therefore, it is expected that all the features have a similar discriminative power.

We emphasize that the use of the time/space decomposition approach represents an interesting trade-off between classification accuracy and computational effort; also, the use of a reduced set of features implies a smaller processing time. This point is an important issue in practical applications, where an adequate compromise between the quality of a solution and the time to obtain it must be achieved. Indeed, the most adequate feature set, the music signal segment from where the features are extracted, the number and the duration of the time segments, the use of space decomposition strategies and the discovery of the more discriminative features still remain open questions for the automatic music genre classification problem.

Acknowledgments

The authors would like to thank CAPES — a Brazilian research-support agency (process number 4871-06-5) and Mr. Breno Moiana for his invaluable help with the experiment infrastructure. We would also like to thank Dr. George Tzanetakis, Dr. Fabien Gouyon and Mr. Thomas Lidy for providing the feature extraction algorithms used in this work.

References

[1] J. J. Aucouturier and F. Pachet, Representing musical genre: A state of the art, *Journal of New Music Research* **32** (2003) 83–93.

[2] J. Bergstra, N. Casagrande, D. Erhan, D. Eck and B. Kégl, Aggregate features and ADABOOST for music classification, *Machine Learning* **65**(2–3) (2006) 473–484.

[3] A. Blum and P. Langley, Selection of relevant features and examples in machine learning, *Artificial Intelligence* **97** (1997) 245–271.

[4] P. Cano, E. Gómez, F. Gouyon, P. Herrera, M. Koppenberger B. Ong, X. Serra, S. Streich and N. Wack, ISMIR 2004 Audio Description Contest, Technical Report MTG-TR-2006-02, Music Technology Group, Pompeu Fabra University, 2006.

[5] C. H. L. Costa, J. D. Valle Jr. and A. L. Koerich, Automatic Classification of Audio Data, in *Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics*, The Hague, Holland, 2004, pp. 562–567.

[6] M. Dash and H. Liu, Feature selection for classification, *Intelligent Data Analysis* **1** (1997) 131–156.

[7] T. G. Dietterich, Ensemble methods in Machine Learning, in *Proc. of the 1st Int. Workshop on Multiple Classifier System*, Lecture Notes in Computer Science 1857, 2000, pp. 1–15.

[8] J. S. Downie and S. J. Cunningham, Toward a theory of music information retrieval queries: System design implications, in *Proc. of the 3rd Int. Conf. on Music Information Retrieval*, Paris, France, 2002, pp. 299–300.

[9] R. Fiebrink and I. Fujinaga, Feature Selection Pitfalls and Music Classification, in *Proc. of the 7th Int. Conf. on Music Information Retrieval*, Victoria, CA, USA, 2006, pp. 340–341.

[10] M. Grimaldi, P. Cunningham and A. Kokaram, A Wavelet Packet representation of audio signals for music genre classification using different ensemble and feature selection techniques, in *Proc. of the 5th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, Berkeley, CA, USA, 2003, pp. 102–108.

[11] M. Grimaldi, P. Cunningham and A. Kokaram, An evaluation of alternative feature selection strategies and ensemble techniques for classifying music, in *Workshop on Multimedia Discovery and Mining, 14th European Conference on Machine Learning, 7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Dubrovnik, Croatia, 2003.

[12] F. Gouyon, P. Herrera and P. Cano, Pulse-dependent analysis of percussive music, in *Proc. of the 22th Int. AES Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002.

[13] F. Gouyon, S. Dixon, E. Pampalk and G. Widmer, Evaluating rhytmic descriptions for music genre classification, in *Proc. of the 25th Int. AES Conference on Virtual, Synthetic and Entertainment Audio*, London, UK, 2004.

[14] S. Hacker, *MP3: The Definitive Guide* (O'Reilly Publishers, 2000).

[15] Audio Description Contest, Website, URL http://ismir2004.ismir.net/ISMIR_Contest.html, 2004.

[16] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3) (1998) 226–239.

[17] K. Kosina, *Music Genre Recognition*, MSc. dissertation, Fachschule Hagenberg, June 2002.

[18] J. H. Lee and J. S. Downie, Survey of music information needs, uses, and seeking behaviours: preliminary findings, in *Proc. of the 5th Int. Conf. on Music Information Retrieval*, Barcelona, Spain, 2004, pp. 441–446.

[19] Magnatune, "Magnatune: MP3 Music and Music Licensing (Royalty Free Music and License Music)," [Web site] 2006, Available: http://www.magnatune.com

[20] D. McEnnis, C. McKay, I. Fujinaga and P. Depalle, jAudio: A Feature Extraction Library, in *Proc. of the 6th Int. Conf. on Music Information Retrieval*, London, UK, 2005, pp. 600–603.

[21] M. Li and R. Sleep, Genre Classification via an LZ78-Based String Kernel, in *Proc. of the 6th Int. Conf. on Music Information Retrieval*, London, UK, 2005, pp. 252–259.

[22] T. Li, M. Ogihara and Q. Li, A Comparative study on content-based Music Genre Classification, in *Proc. of the 26th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, 2003, pp. 282–289.

[23] T. Li and M. Ogihara, Music Genre Classification with Taxonomy, in *Proc. of IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Philadelphia, PA, USA, 2005, pp. 197–200.

[24] T. Lidy and A. Rauber, Evaluation of feature extractors and psycho-acoustic tranformations for music genre classification, in *Proc. of the 6th Int. Conference on Music Information Retrieval*, London, UK, 2005, pp. 34–41.

[25] H. Liu and L. Yu, Feature Extraction, Selection, and Construction, in *The Handbook of Data Mining*, ed. Nong Ye (Lawrence Erlbaum Publishers, 2003), pp. 409–424.

[26] C. McKay and I. Fujinaga, Musical Genre Classification: Is it worth pursuing and how can it be?, in *Proc. of the 7th Int. Conf. on Music Information Retrieval*, Victoria, CA, USA, 2006, pp. 101–106.

[27] A. Meng, P. Ahrendt and J. Larsen, Improving Music Genre Classification By Short-Time Feature Integration, in *Proc. of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, 2005, pp. 497–500.

[28] T. M. Mitchell, *Machine Learning* (McGraw-Hill, 1997).

[29] L. C. Molina, L. Belanche and A. Nebot, Feature Selection Algorithms: a Survey and experimental Evaluation, in *Proc. of the IEEE Int. Conference on Data Mining*, Maebashi City, JP, 2002, pp. 306–313.

[30] E. Pampalk, A. Rauber and D. Merkl, Content-based Organization and Visualization of Music Archives, in *Proc.of the ACM Multimedia*, Juan-les-Pins, France, 2002, pp. 570–579.

[31] A. Rauber, E. Pampalk and D. Merkl, Using Psycho-acoustic transformations for music genre classification, in *Proc. of the 3rd Int. Conference on Music Information Retrieval*, Paris, France, 2002, pp. 71–80.

[32] A. Rauber, E. Pampalk and D. Merkl, The SOM-enhanced JukeBox: organization and visualization of music collections based on perceptual models, *Journal of New Music Research* **32**(2) (2003) 193–210.

[33] D. Rocchesso, *Introduction to Sound Processing*, 1st ed. (Università di Verona, 2003).

[34] C. N. Silla Jr., C. A. A. Kaestner and A. L. Koerich, Time-Space Ensemble Strategies for Automatic Music Genre Classification, in *Proc. of the Brazilian Symposium on Artificial Intelligence*, Ribeirão Preto, SP, Brazil, Lecture Notes in Computer Science 4140, 2006, pp. 339–348.

[35] C. N. Silla Jr., C. A. A. Kaestner and A. L. Koerich, The Latin Music Database: a database for the automatic classification of music genres (*in portuguese*). *Proc. of 11th Brazilian Symposium on Computer Music*, São Paulo, SP, Brazil, 2007, pp. 167–174.

[36] C. N. Silla Jr., C. A. A. Kaestner and A. L. Koerich, Automatic Music Genre Classification using Ensemble of Classifiers, in *Proc. of the IEEE Int. Conference on Systems, Man and Cybernetics*, Montreal, Canada, 2007, pp. 1687–1692.

[37] C. N. Silla Jr., *Classifiers Combination for Automatic Music Classification (in portuguese)*, MSc dissertation, Graduate Program in Applied Computer Science, Pontifical Catholic University of Paraná, 2007.

[38] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner, The Latin Music Database, in *Proc. of the 9th Int. Conference on Music Information Retrieval*, Philadelphia, PA, USA, 2008, pp. 451–456.

[39] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner, A machine learning approach to automatic music genre classification, *Journal of the Brazilian Computer Society* **14**(3) (2008) 7–18.

[40] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner, Feature Selection in Automatic Music Genre Classification, in *Proc. of the 10th Int. Symp. on Multimedia*, Berkeley, CA, USA, 2008, pp. 39–44.

[41] A. Stolcke, Y. Konig and M. Weintraub, Explicit word error minimization in n-best list rescoring, in *Proc. of Eurospeech 97*, Rhodes, Greece, 1997, pp. 163–166.

[42] G. Tzanetakis and P. Cook, Musical genre classification of audio signals, *IEEE Transactions on Speech and Audio Processing* **10** (2002) 293–302.

[43] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, San Francisco, 2005).

[44] Y. Yaslan and Z. Cataltepe, Audio Music Genre Classification Using Different Classifiers and Feature Selection Methods, in *Proc. of the Int. Conference on Pattern Recognition*, Hong Kong, China, 2006, pp. 573–576.

[45] E. Zicker and H. Fastl, *Pcychoacoustics — Facts and Models*, Springer Series on Information Sciences 22 (Springer, Berlin, 1999).