Statistical Models in Music Genre Classification

Tiago Filipe Beato Mourato de Matos
Técnico Lisboa, UL
Lisboa, Portugal
tiagofbmm@gmail.com

ABSTRACT

Music genre classification is a subjective process. As such, there is a long way to go until a highly reliable mechanism is built in order to classify new songs into their belonged groups. Although many efforts have been made to achieve such goal, the variability and continuous evolution of the musical culture has always been a barrier for physicists and statisticians that study this subject.

Due to its complexity, which involves culture influences sometimes (not so) highly structured, the need for retrieving musical attributes is the main issue on which this study has been focused. Since the creation of a random sample of music examples from each initially considered genres, to the proper characteristics retrieval, many aspects were refined during the process, with the main objective of having a refined classifier. The first attribute selection alongside with the physical approach where a more specific set of attributes including MFCC (mel-frequency-cesptral-coefficients) and FBE (filterbank energy) were retrieved, became an extremely useful way to fulfill the main objectives established for this project.

In this process, the KNN (K-nearest-neighbours) algorithm, along with the Euclidean Distance and a proper set of attributes from the referred sources will reach a 79% success rate in classifying musical examples. We conclude that not all of the attributes initially retrieved were actually usefull for classification, and that the union of the best attributes from both sources got the best results.

On the other hand, the comparison of elements using principal components analysis and a graphical representation of their similarities has become very supportive of the conclusions made from the data sources.

Keywords

Music, Data Mining, Clustering, Principal Component Analysis.

1. INTRODUCTION

Classifying a song within any given genre is a subjective and transient process. The central question is that the limits of each musical genre are not uniquely defined. Consequently, the classification can turn out to be not as easy as we think it could be, but something misleading and sensitive to human perception differences. Pointing that out, there are many websites which list the genre of a song and one may categorize it as Rock while the other classifies it as Heavy Metal. Furthermore as many genres overlap and some also have sub-genres, there is no definitive way of finding the genre of a song, but there are many methods which can be tried.

The two main problems which are the base motivation for this thesis are the classification of a random subset of songs in terms of their musical genre, as well as clustering them into groups in an unsupervised way. To achieve such an objective, some different tools will be used, like the Principal Component Analysis and other statistical methods of classification and clustering, which can be framed in the Data Mining structure.

It will be seen that the two major approaches, one by physics and the other by statistics, have different characteristics, where the classification reaches a success rate of nearly 80%. Together with that, some surprising genre similarities will be shown, as well as some that could be expected *a priori*. The obtained results will be validated using different cross-validation methods, and using different sets on which the classifiers can be tested.

2. LITERATURE REVIEW

Through this review of a part of the related work in this subject, it can be seen that many of the most recent ones have the characteristic of being mainly focused on the physical and electrical interpretation of music. As mentioned before, the task of classifying a song in its genre can be a complex task. George Tzanetakis and Perry Cook (Tzanetakis and Cook, 2002) focus their work on those signal forms, in order to retrieve the MFCC. These coefficients play an important role in this process, as they represent a part of the frequency spectrum of a sound, based on a triangular cosine transform in log-mel scale. Pointing their relevance, MFCC are commonly used for speech recognition for phone numbers or by associating personal recognition by voice.

The next step in their work regarded the estimation of the

probability density function of the vector of characteristics obtained, using the EM method to estimate the associated parameters.

Regarding another study (Clark, Park and Guerard, 2012), although the objective of creating an automatic mechanism for musical classification, the vector of attributes analyzed was totally different, with measures from a musical distribution site, the Echonest. In this work, the authors used some algorithms of machine learning, such as Neural Networks and Growing Neural Gas for classifying a set of songs in their genres, having a maximum sucess rate of 70%.

3. DATASET CONSTRUCTION

Having the previous classification work in mind, it got obvious that the creation of a useful dataset that would clearly classify the songs that we were analizing would have great influence in the sucess rate. According to the subjective limits referred, mixed groups such as Pop-Rock, Punk-Rock or Heavy-Metal would not be desired to build a classifier. It was a suspect that the more distinguishable groups of songs would be those that have a longer and strongly established history, such as Rock, Reggae, Rap or Classic. Having these 4 groups in consideration, the task begun with the creation of a random sample of each of those genres. To achieve that, Lastfm site was used to generate random sets of songs of each genre.

With the (artist,song) dataset established, the retrieval of the main attributes that characterizes them was the next step. For that problem, a Java script was created to acess an URL with respect to each of the songs, in another website, the Echonest. In this website, it is possible to gather an API key to retrieve the information required about every song. This application simulates the way humans hear the songs, envolving the psycoaccoustic principles, musical perception and a physical and cognitive recognition of music. The vector of attributes gathers information of Timbre, Rhythm and Loudness is in Table 1.

Table 1: Musical attributes and domain.

Characteristic	Domain
Energy	$(0, 1)$
Tempo	$\mathbb{R}_0{}^+$
Speechiness	$(0, 1)$
Key	$\{0, 1, 2, \ldots, 11\}$
Duration	$\mathbb{R}_0{}^+$
Mode	$\{0, 1\}$
Time Signature	$\mathbb{N}$
Loudness	$\mathbb{R}_0{}^-$
Danceability	$(0, 1)$

These attributes were measured in a dataset with distribution shown in Table 2.

Table 2: Dataset Genre Distribution.

Genre	Number of Elements
Classic	53
Rap	48
Reggae	52
Rock	52

4. CLASSIFICATION WITH FIRST ATTRIBUTES SET

The core of this work in terms of classification was the KNN algorithm. This one is a so called lazy learner, which in case demands a high computacional ability.

Let us consider the main characteristic of this algorithm, which is the principle of finding similarities between the objects to be classified. The similarity between each pair of elements is measured regarding its p attributes. In order to classify an element with the same attributes, the algorithm finds the k closer members to the object, according to the defined similarity. In its simplest form, the analyzed object is then classified according to the majority of the classification of the k closer elements.

After testing several dissimilarity functions to compare the elements and every set of attributes, the results pointed to choose the Euclidean distance.

In addition, the final set of attributes chosen to build the classifier were simply the numerical ones, except the Tempo. These tests were performed regarding an index measure of good classification computed by *cross validation* in the algorithm KNN.

Proceeding in the same way, the single parameter k of the KNN algorithm was determined regarding the "leave-one-out" method. The results for the classification are in Table 3.

Table 3: Sucess Rate estimated by *Cross-Validation* with Euclidean Distance.

CV/k	1	2	3	4	5	6	7	8	9	10
50:50	66.2	65.7	67.2	68.35	68.6	68.3	69.2	70.2	69.9	70.7
70:30	68.4	68.3	66.6	67.8	69.0	67.9	70.9	70.4	71.2	69.7
75:25	69.3	67.9	67.9	66.7	69.7	67.8	68.9	70.5	68.9	70.0
80:20	69.4	66.4	67.6	69.3	67.5	68.8	69.8	69.0	70.2	71.8
90:10	69.9	70.8	69.4	68.8	67.4	69.9	69.61	70.0	70.5	69.1
L-O-O	69.3	71.7	68.8	68.8	66.3	71.21	69.3	70.2	71.7	70.7
CV/k	11	12	13	14	15	16	17	18	19	20
50:50	69.6	70.3	69.8	69.5	69.6	69.4	69.0	68.6	69.6	68.9
70:30	69.8	71.9	69.5	71.3	69.5	70.8	69.7	70.9	70.5	69.7
75:25	71.3	71.8	71.4	71.5	70.2	71.9	70.5	70.4	70.8	71.7
80:20	70.0	70.6	71.3	72.0	71.8	71.7	71.1	69.6	71.2	71.2
90:10	72.3	72.1	73.4	73.8	72.9	72.6	71.5	72.7	73.4	70.9
L-O-O	70.2	70.7	70.2	72.7	73.7	72.2	72.2	73.2	73.2	72.7

Table 3 has the cross-validation success rate in different partitions, where every random generation of partition was performed 100 times, being the estimated result the mean of the sucess rate obtained (except "leave-one-out", for obvious reason).

4.1 ROC Curves

As mentioned before, there are different indicators to analyze the results obtained by classification. Although all of them have a probabilistic interpretation and represent different aspects of the process, the ROC (Receiving Operating Characteristic) curves play an unifying role of those. Their major relevance become clear as a decision tool in medical terms (Zweig and Campbell, 1993).

The main objective is to determine a factor (in this case k) that supplies the best relation between the so called true positives rate (TPR) and the false positives rate (FPR).

It is important to establish that in a binary classifier, it will divide de dataset in two groups, where the positiveness refers to the existence of the analyzed property, whilst the negativeness refers to its absence. Beyond the plot observation of the two mentioned measures, the AUC (area under the curve) is also a good measure of the classifier's performance, so that the random case would be represented by $AUC = 0.5$ and the perfect one by $AUC = 1$ (Marzban, 2004). In the current work, we have four groups to divide the dataset, so, as an example we have:

Let us take the happening C="Classic Song" and NC="Non Classic Song". Then we will have the following sets:

1. True Positives (TP)= C|C = the classical songs classified as such.

2. False Positives (FP)=C|NC= the non classical songs that are classified as classical songs.

3. True Negatives (TN)=NC|NC= the non classical songs classified as such.

4. False Negatives (FN)=NC|C= the classical songs that were classified as non classical ones.

This kind of analysis is usually condensed in the confusion matrix:

Table 4: Confusion Matrix

Predicted /Real	Real Positive	Real Negative
Predicted Positive	True Positives (TP)	False Positives (FP)
Predicted Negative	False Negatives (FN)	True Negatives (TN)

Therefore, the point was to maximize the TPR and minimize at the same time the FPR, having a distinct analysis for each of the classes involved. By the results of the ROC curves of each genre, the k was finally chosen to be $k = 15$.

5. RESULTS FOR FIRST DATASET

Considering the estimated parameter $k = 15$, the following results on Tables 5 to 8 represent the classification performance in the 205 songs dataset.

Table 5: Classic

Class/Real	Pos.	Neg.
Pos.	45	8
Neg.	7	145

Table 6: Rap

Class./Real	Pos.	Neg.
Pos.	35	18
Neg.	13	139

Table 7: Reggae

Class./Real	Pos.	Neg.
Pos.	27	16
Neg.	25	137

Table 8: Rock

Class./Real	Pos.	Neg.
Pos.	42	14
Neg.	11	138

Through the matrixes in the Tables 5 to 8, it is understandable that the majority of errors of bad classification is closely related with the connection between Rap and Reggae. Beyond the confusion matrixes shown, there are other ways to show the results of classification, such as **accuracy**, **specificity**, **precision** and **sensitivity** (see Table 9).

Table 9: Classifier performance measures ($j = 0, 1$).

Measure	Probability	Estimate
Accuracy	$Pr(\hat{Y}(\mathbf{X}) = 0 \mid Y = 1)$ and $Pr(\hat{Y}(\mathbf{X}) = 1 \mid Y = 0)$	$\frac{FP+FN}{n}$
Sensitivity	$Pr(\hat{Y} = j \mid Y = j)$	$\frac{TP}{TP+FN}$
Precision	$Pr(Y = j \mid \hat{Y}(X) = j)$	$\frac{TP}{TP+FP}$
Specificity	$Pr(\hat{Y}(X) \neq j \mid Y \neq j))$	$\frac{TN}{TN+FP}$

In these four genres we got the results shown below.

Table 10: Classic

Accuracy	0.9268
Specificity	0.9477
Precision	0.8490
Sensitivity	0.8654

Table 11: Rap

Accuracy	0.8488
Specificity	0.8854
Precision	0.6604
Sensitivity	0.7292

Table 12: Reggae

Accuracy	0.8
Specificity	0.8954
Precision	0.6279
Sensitivity	0.5192

Table 13: Rock

Accuracy	0.8780
Specificity	0.9079
Precision	0.7500
Sensitivity	0.7925

Analysing the indicators set in Tables 10 to 13, we can see that for each genre we have a high proportion of correctly classified elements. Furthermore, in accuracy terms, it is clear that the greatest genre in that measure is Classic, followed by Rock, Rap and finally Reggae. On the other hand, the specificity values are very close between the different genres, nearly 90%. Referring to precision, it behaves in a similar way as accuracy, with Reggae having the smallest proportion of True Positives. Finally, in sensitivity terms, Classic is the winner, whilst Reggae is by far the worst one.

6. UNSUPERVISED CLASSIFICATION

6.1 K-Means

Diverging from the approach taken till now, we will skip to another group of methods of classification: unsupervised classification methods. Starting by the K-Means algorithm, its executional base is settled on the principle of having a dataset of dimension n on which we suspect of having k different groups of elements.

In order to apply this method to the musical dataset, we recurred to R functions. Actually, we want to extract four groups from the original set, so we know that the parameter $k = 4$, where k is now the number of groups to extract.

Taking in consideration the variables selection performed in the previous sections, the metric used was the Euclidean distance, calculated in the whole set of numerical variables, except the Tempo. Even in this algorithm application, we have some variants (Redmond and Heneghan, 2005), Forgy, Hartigan-Wong, Lloyd and MacQueen.

To get rid of the differences that can happen due to the randomness of some factors in these algorithms, we will execute it in this way, for each of the mentioned methods, we compute the maximum success rate obtained by 100 runs of K-Means method. The results for the minimum error are in Table 14.

Table 14: Misclassification Error versus K-Means chosen methods.

Method	Misclassification Error Rate
Forgy	0.3073
Hartigan-Wong	0.3073
Lloyd	0.3073
MacQueen	0.3122

6.2 K-Medoids

On the same guideline of the K-Means method, K-Medoids comes as an interesting variant, because this one is not as sensitive to *outliers*[1] as the former one.

The major difference between these two methdos is that the centroids in the K-Means algorithm is one point of the space, whichever it is, and in K-Medoids the centroids are one of the objects in the dataset. The final objective of this algorithm is to minimize the expression:

$$E = \sum_{j=1}^{k} \sum_{x \in C*_j} |x - o_j| \quad (1)$$

Summing up, E will be the sum of the absolute error for every element in the dataset, whilst x is the space point representing each element and o_j the representant of cluster $C*_j$.

Although the *outliers* are less important in the results of this algorithm, its complexity becomes much bigger: $O(k(n-k)^2)$, where k is the number of groups and n the cardinal of the dataset. As mentioned before, using R software and its packages "cluster" and "pam", we have implemented methods to choose the initial centroids in order to have better performance, and we obtained 0.6878 of missclassification error rate.

[1]This concept is not easy to properly define but it is understood as an observation that is noticeably distant from the remaining ones in the sample from where it was selected (Grubbs, 1969)

6.3 Hierarchical Clustering

The hierarchical clustering is about forming groups of objects in a clustering tree.

Most of this type algorithms are known for this aglomerative definition of hierarchical clustering, being their variants in terms of determining the distance between groups (Han and Kamber, 2006). The result of this applications can be seen through a tree structure, where each step of the algorithm shows the grouped elements. There are several distances that can be considered, Ward, single, complete, average linkage.

In the Table 15 and Figure 1, the results and dendrogram (Ward) in respect to the hierarchical clustering application are shown.

Table 15: Misclassification error with hierarchical clustering versus choses distances.

Method	Misclassification error
Ward	0.2732
Single(Min)	0.7318
Complete(Max)	0.5317
Average	0.5366

Through the Figure 1, we can see the configuration of the aglomerative tree to obtain the desired four groups:

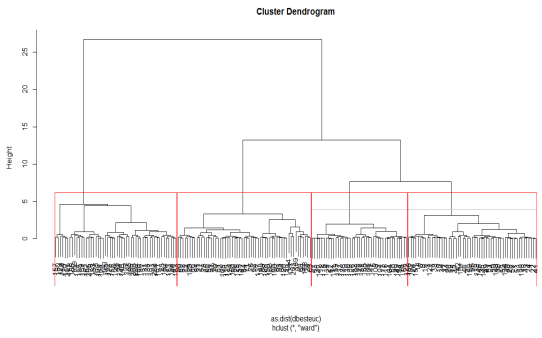


Figure 1: Ward method dendrogram with 4 clusters.

By investigating how are the objects (songs) distributed in the different groups, the left side of the tree referrs to the Classic elements, and when moving to the right, we can see the Rap, Reggae and Rock groups. Moving up in the tree, it can be seen that, althoug there is a common source for Rap and Reggae (historically speaking), they are not in the first cluster that covers both genres. Furthermore, it is clear that Classic is the most prominent genre. These conclusions can also be withdrawed from the Table 16.

Table 16: Genre distribution for each cluster with Ward Method.

Cluster	Classic	Rap	Reggae	Rock
1	84.6%	0%	5.8%	9.6%
2	5.2%	66.7%	28.1%	0%
3	0%	19.5%	63.4%	17.1%
4	9.1%	3.6%	12.7%	74.5%

Having in consideration Table 16, it becomes clear that there is some difficulty in distinguishing Reggae and Rap, being that confusion mostly justified by Reggae because this one turns to look similar to rock in this analysis. The most unequivocal groups are the first and the fourth one, which refer to Classic and Rock.

7. DETERMINISTIC CROSS-VALIDATION

In this section, we will try to solve some of the doubts that came from the previous classification. Again we will rely on **cross validation**, but in this case, we will use it to find out other kinds of similarities that would not be so clear at first sight. What is meant by deterministic cross validation is that we will classify each of the songs of the dataset with a training set constituted by elements that do not belong to the genre in analysis.

The process is as follows:

1. Choose three musical genres.
2. Create a training set having the songs with the chosen genres.
3. Classify the remaining genre with the mentioned training set.

In Table 17 are shown the results ($k = 15$).

Table 17: Similarities between musical genres with KNN ($k = 15$).

Genre to Classify/Results	Classic	Rap	Reggae	Rock
Classic	-	3.85%	9.62%	86.54%
Rap	0%	-	93.75%	6.25%
Reggae	5.77%	48.08%	-	46.15%
Rock	18.87%	7.55%	73.58%	-

Analyzing the Table 17, we can conclude that Classic genre is more similar to Rock than with any other. Furthermore, Rap is very similar to Rap but when we classify Reggae songs, they split between Rap and Rock. Finally, the classification of Rock songs shows us that this genre is closer to Reggae than to the other genres.

8. VISUAL REPRESENTATION

Having some tools that make possible the analysis of the current songs set, a visual representation of what is actually happening and how distributed the elements are is a priority.

8.1 Principle Component Analysis

Principle Component Analysis comes up as form of data reduction (Smith, 2002). This method consists on transforming a set of p dimensional tuples into $q \leq p$ dimensional ones, having the original space to be reduced into a smaller dimension and providing relations between the involved variables that might not be clear *a priori*.

The method is an ortogonal linear transformation of the original data into another coordinate system, in a way that the greatest variance of those projections is assigned to the first coordinate, named the first principal component, being this reasoning used for the remaining coordinates. Contrairily to the simple data reduction, retrieving them from the original set, this process allows that, in general, the new projection keeps the initial properties.

Concretizing this algorithm to our problem, the results as shown for a reduction of data dimension for $q = 2$ and $q = 3$ in the Figures 2 and 3.

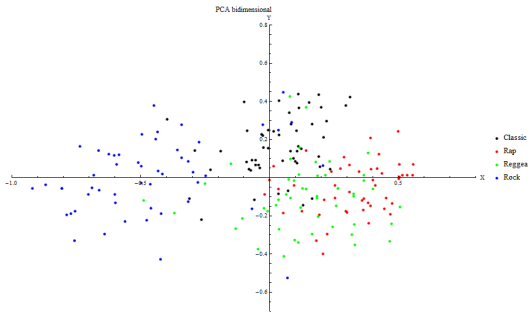
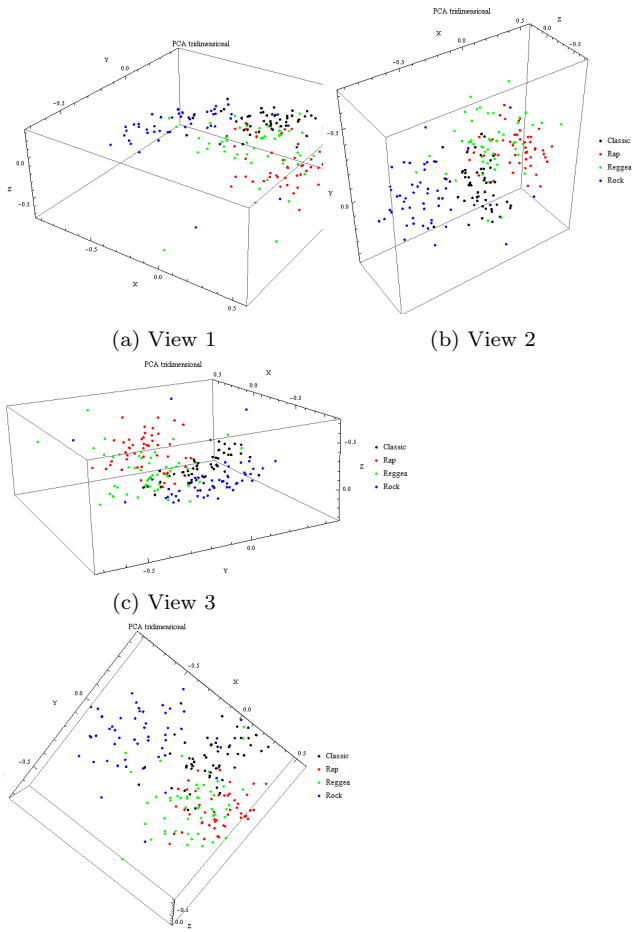


Figure 2: PCA 2-D view with numerical attributes except Tempo.

Analyzing the Figures 2 and 3 we can have a greater perception of the data distribution, by projecting the first two and three principal components, with the variability percentages being, 63.31% (2 first components) and 77.73% (3 first components).

Through the PCA, we can realize some important characteristics of the specific data. In terms of musical genres, and supporting the former conclusions, it is possible to verify that the distribution of points in the bidimensional case shows great difference between Rock and the remaining ones. On the other hand, although the separation is not that clear, in Figure 2 is revealed that the Classic elements are distinguishable from the remaining genres. Referring to Rap and Reggae, its projection in two dimensions also shows a straight relation between these two genres, which can be justified by its historical common source.

Furthermore, by the 3-D view of the projections, the conclusions remain the same as the former, having the Classic elements a compact form, whilst Rap and Reggae are now more easily distinguishable.



(a) View 1 (b) View 2

(c) View 3

(d) View 4

Figure 3: PCA 3-D view with numerical attributes except Tempo.

9. VALIDATION AND MODEL EXTRAPOLATION

9.1 Validation

In order to obtain another measure of the classification quality, a different set of songs was classified, having 20 elements of each of the considered genres. Proceeding in the same way then before, with the same set of attributes and by creating the distance matrix again with Euclidean distance and using KNN with $k = 15$, the results obtained are on the Tables 18 to 25.

Class./Real	Pos.	Neg.
Pos.	16	3
Neg.	4	57

Table 18: Classic

Class./Real	Pos.	Neg.
Pos.	16	3
Neg.	4	57

Table 19: Rap

Class./Real	Pos.	Neg.
Pos.	13	7
Neg.	7	53

Table 20: Reggae

Class./Real	Pos.	Neg.
Pos.	14	8
Neg.	6	52

Table 21: Rock

Accuracy	0.9125
Specificity	0.95
Precision	0.842
Sensitivity	0.8

Table 22: Classic

Accuracy	0.9125
Specificity	0.95
Precision	0.842
Sensitivity	0.8

Table 23: Rap

Accuracy	0.825
Specificity	0.883
Precision	0.65
Sensitivity	0.65

Table 24: Reggae

Accuracy	0.825
Specificity	0.867
Precision	0.636
Sensitivity	0.7

Table 25: Rock

As it can be confirmed by the last tables, the results are very similar to the obtained ones before.

9.2 Extrapolation

Having done the validation of the built model with a new set of songs, another objective of this work is to achieve the limitations of the musical genres, namely between the original genres and a new set, Blues, Jazz, Metal and Pop.

The procedure will run in a similar way to the former ones. The training set will be formed by the same 205 songs from the 4 inital genres, and a set of 10 elements of each new genre will be classified. Again with KNN, Euclidean distance and the same set of attributes, the results are in Table 26.

Table 26: Similarities between the new genres with KNN.

Genre to Classify/Results	Classic	Rap	Reggae	Rock
Blues	20%	10%	20%	50%
Jazz	60%	20%	10%	10%
Metal	0%	0%	0%	100%
Pop	10%	10%	40%	40%

From Table 26, it can easily be seen that all the elements of genre Metal are classified as rock, what suggests great similarity between them. Beyond that, Jazz seems closer to Classic than to any of the other initial genres. Although Blues and Pop do not have the same clear evidence, we can figure out that the majority of the Blues songs are closer to the Rock ones, whilst in Pop there is an equal tendency for Rock and Reggae.

In order to get the same quality of visual representation, PCA was applied to the data and the results can be seen in Figures 4 and 5.

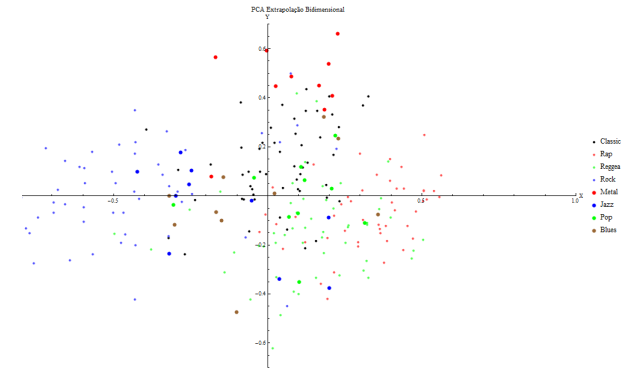
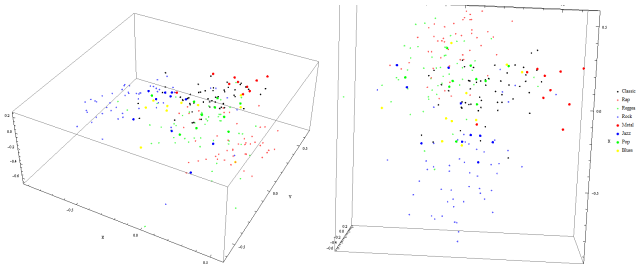
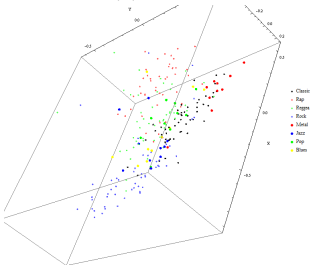


Figure 4: PCA 2-D with new musical genres.

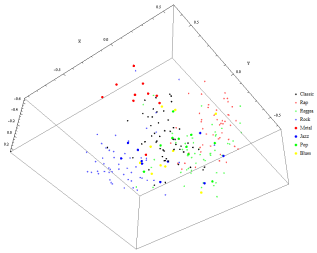


(a) View 1

(b) View 2



(c) View 3



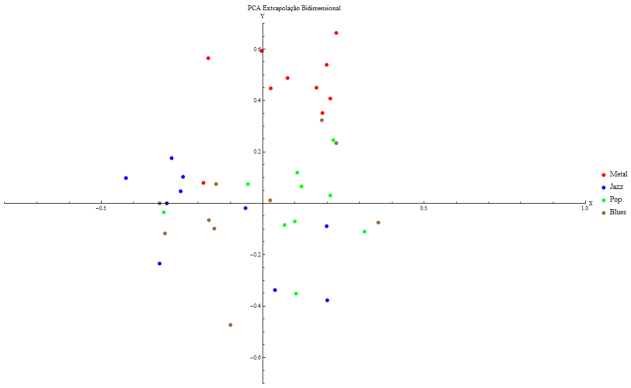
(d) View 4

Figure 5: PCA 3-D view with new musical genres.

Analising the obtained results through this classification method, some comments are needed. In respect to the Figure 4, the songs represented by thicker red dots are, on the contrary to the results on Table 26, closer to the elements of the Classic set, being the second closer set the Rock one. Referring to Jazz genre, it can be seen that they are closer to Jazz than the remaining groups. In terms of Pop, its elements are above the elements of Reggae genre, according with Table 26. Finally, Blues comes up to have similarities with Rock and Classic, fact that agrees with the info in Table 26.

On the other hand, the similarity between the new genres can be seen through Figure 6.

Figure 6: PCA 2-D representation with new genres



Referring to the Figure 6, it shows that Metal is the more detached group and that Jazz and Blues are close genres too. To finalize, Pop also shows as a compact space and distant from other genres.

10. ALTERNATIVE ATTRIBUTES RETRIEVAL

As mentioned on the previous sections, music can be analyzed as an electric signal, with all its implications. Beyond the retrieved attributes, we also want to gather specific information about its electrical characteristics, what will provide us the ability to analyze attributes visually represented by Figure 7.

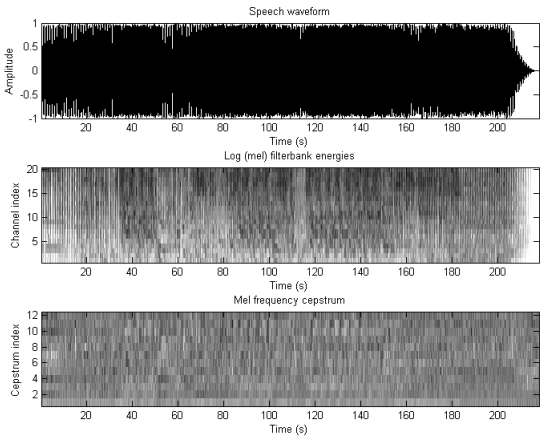


Figure 7: Sound Features Graphics.

Despite the fact that the first graphic in Figure 7 shows us a representation of amplitude versus time, its analysis

is not prosperous to the results for which we are aimed. On the other hand, the remaining ones were analyzed and used as sound characteristics, helping the classification task. Whilst the second graphic on Figure 7 referrs to the so called "filterbank energies", the third one referrs to the MFCCs.

10.1 Filterbank Energies

The signal spectrum gives us the energy distribution as function of its frequencies. An example of this concept is the idea implemented in nowadays equalizers. This instrument allows us to adjust our sound machines the way that more suits our desires, by increasing certain band frequencies influence, whilst decreasing others.

The main objective of these devices is to separate energy from a frequency region of a signal's spectrum, using a "*bandpass filter*". Its functionality resides on retrieving a specific frequency band of a signal, eliminating the remaining ones (Cassidy and Smith III, 2005). By defining f_{cl} and f_{ch} as the inferior and superior frequency limits:

$$BW(\text{Band width}) = f_{\text{cl}} - f_{\text{ch}}$$

This way, a "*filterbank*" is a system that divides the input signal $x(t)$ into $x_1(t), x_2(t), \ldots$, where each one corresponds to a different spectrum region, as shown in Figure 8

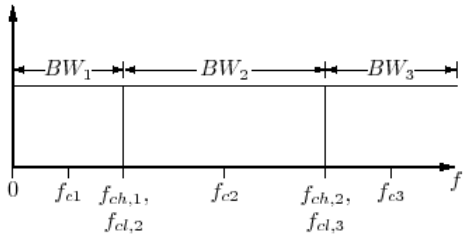


Figure 8: Exemplo de filterbank.

In our case, we will have 20 coefficients which characterize the energy in frequency intervals between $(300 - 3700)$Hz.

10.2 MFCC

The extraction process of MFCCs or "*filterbank energy*" coefficients is non-invertible, so that it is a transformation where there exists some information loss. This information loss, despite being an undeniable fact, is mainly justified by the enormous computacional complexity that would envolve to gather extremely rigorous elements extraction. By increasing the eficiency of this process, we would have to increase the number of coefficients to retrieve, what would naturally increase the sound complexity, without having a proof of its utility on this process.

The MFCCs are coefficients that characterize small intervals of time in a sound (Logan, 2005). Their sucess is mainly due to the fact that they represent the amplitude of a spectrum in a compact way. The use of MFCCs along with the FBEs will give us a way to analyze the signal's characteristics in their main two branches: amplitude and frequency.

10.3 Inclusion of FBE and MFCC

In order to include these elements regarding every song, their process was implemented, using a Matlab procedure. The specifications of the parameters are the following:

- Tw = 25 (Frame duration);
- M = 20 (number of "*filterbank energies*");
- C = 12 (number of MFCCs);
- LF = 300 (lower frequency limit);
- HF = 3700 (higher frequency limit);

Note that these are the standard values used on musical analysis. As a result of the new attributes retrieval process, we got 33 new music coefficients that define the songs. Together with the former 5 attributes from the previous model, we got a 38 vector characterizing each song, with the original variables Energy, Speechiness, Duration, Loudness, Danceability, 13 MFCCs and 20 FBEs.

By introducing these new elements, the classification results were not better that the ones obtained before, by using cross-validation with KNN method (Euclidean distance) and $k = 15$.

One reason for this insuccess can be the increase of the ratio between variables and objects. On the other hand, as we are dividing the amplitude and energy spectrums in intervals, some of those might be good for the classification goal whilst others might not. This way, in order to find which coefficients should be included, greedy selection algorithms were created, such as forward and backward ones. The main principles involved were the KNN algorithm with the usual Euclidean distance, and with those elements, at each step cross-validation was used.

The results of these process, by choosing at each step whether there is a coefficient that increases the sucess in classification, can be seen on Table 27.

Table 27: Max success by number of added coefficients (KNN with cross-validation and Euclidean distance.

Number of coefficients	Max success rate
1	76.0%
2	77.1%
3	78.0%
4	78.5%
5	79.0%
6	79.5%

With the forward algorithm to select variables, the best obtained result is to add the MFCC $[1, 2, 3, 7, 11]$ and FBE [20], applying KNN with $k = 11$.

10.4 Results

Having chosen the new elements to be included in the model, we can see that the success rate increases to 79.5%.

Following the same schema as the one with the former attribute set, the confusion matrixes are shown on Tables 31 to 35.

Table 28: Confusion matrix for Classic.

Class./Real	Pos.	Neg.
Pos.	44	0
Neg.	1	37

Table 29: Confusion matrix for Rap.

Class./Real	Pos.	Neg.
Pos.	37	11
Neg.	11	146

Table 30: Confusion matrix for Reggae.

Class./Real	Pos.	Neg.
Pos.	35	11
Neg.	17	142

Table 31: Confusion matrix for Rock.

Class./Real	Pos.	Neg.
Pos.	47	16
Neg.	16	136

Table 32: Evaluation measures for Classic.

Acuracy	0.9415
Specificity	0.9739
Precision	0.9167
Sensitivity	0.8462

Table 33: Evaluation measures for Rap.

Acuracy	0.8927
Specificity	0.9299
Precision	0.7708
Sensitivity	0.7708

Table 34: Evaluation measures for Reggae.

Acuracy	0.8634
Specificity	0.9281
Precision	0.7609
Sensitivity	0.6731

Table 35: Evaluation measures for Rock.

Acuracy	0.8927
Specificity	0.8947
Precision	0.7460
Sensitivity	0.8868

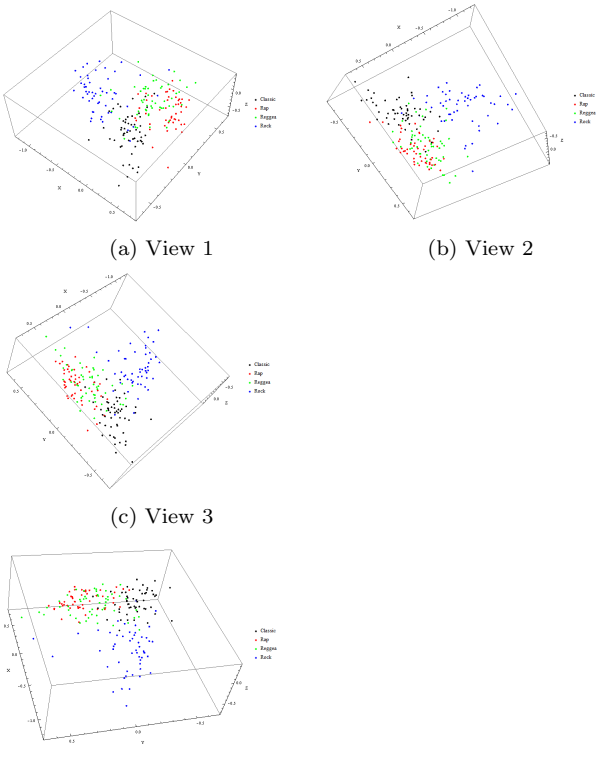
Through all these results, we can see an increase on the sucess rate, as well as the measures for the classifier. For comparison with the former attributes set, the K-Means and K-Medoids reached a better success rate with K-Medoids, 68% but a poorer result with K-Means, 58%.

On the other hand, the hierarchical classification with Ward distance shows a result of 71% of success rate, with the distribution by cluster being the show in Table 36.

Table 36: Music elements distribution by cluster.

Cluster	Classic	Rap	Reggae	Rock
1	79.7%	0%	8.5%	11.9%
2	3.1%	9.2%	18.5%	69.2%
3	0%	25%	71.4%	3.6%
4	5.7%	66.1%	28.3%	0%

Again classifying for each genre the music elements through a training set which is its complementary on the dataset D, the results of this can be seen on Table 37.



(a) View 1 (b) View 2

(c) View 3

(d) View 4

Figure 10: PCA 3-D representation of the objects.

Table 37: Similarities between musical genres.

Genre to Classify/Results	Classic	Rap	Reggae	Rock
Classic	-	3.846%	48.077 %	48.077 %
Rap	0%	-	87.5%	12.5%
Reggae	7.692%	55.77%	-	36.54%
Rock	5.660%	24.53%	69.81%	-

On Figure 9 we can observe the disposal of the music elements with PCA representation.

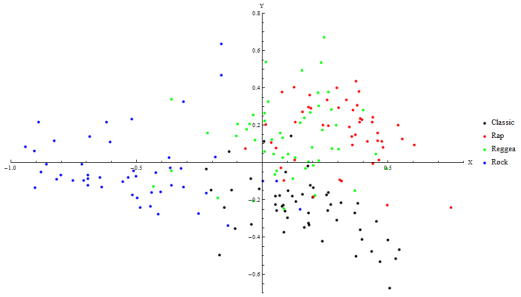


Figure 9: PCA 2-D representation of the objects.

11. VALIDATION AND EXTRAPOLATION OF THE MODEL

11.1 Validation

After the determination of the new elements to be included in the model, the MFCC and the "*filterbank energies*", we will proceed to the model validation with the same test set used before, formed by 10 elements of each genre. The obtained results are shown on Tables 38 to 41.

Table 38: Confusion matrix for Classic.

Class./Real	Pos.	Neg.
Pos.	18	4
Neg.	2	56

Table 39: Confusion matrix for Rap.

Class./Real	Pos.	Neg.
Pos.	16	3
Neg.	4	57

Table 40: Confusion matrix for Reggae.

Class./Real	Pos.	Neg.
Pos.	12	7
Neg.	8	53

Table 41: Confusion matrix for Rock.

Class./Real	Pos.	Neg.
Pos.	15	5
Neg.	5	55

Analysing the data results, it can be seen that the sucess rate increases from 73.8% to 76.3 % by using the new attributes, having the success measures on Tables 42 to 45.

Table 42: Evaluation measures for Classic.

Acuracy	0.925
Specificity	0.933
Precision	0.818
Sensitivity	0.9

Table 43: Evaluation measures for Rap.

Acuracy	0.9125
Specificity	0.95
Precision	0.842
Sensitivity	0.8

Table 44: Evaluation measures for Reggae.

Acuracy	0.813
Specificity	0.883
Precision	0.65
Sensitivity	0.6

Table 45: Evaluation measures for Rock.

Acuracy	0.875
Specificity	0.917
Precision	0.75
Sensitivity	0.75

11.2 Extrapolation

On what respects to the extrapolation task, it is not easy to say which is a right or wrong answer. Eventhough, the empirical musical knowledge allows us to establish that there exists a greater similarity between Rock and Metal than between Metal and any other of the current used genres. With that in mind, the Table 46 shows more doubtable results:

Table 46: Similarities between musical genres with new attributes.

Genre to Classify/Resuls	Classic	Rap	Reggae	Rock
Blues	70%	10%	0%	20%
Jazz	50%	0%	10%	40%
Metal	40%	10%	10%	40%
Pop	60%	0%	0%	40%

12. CONCLUSIONS

Analysing the full process of this work, some conclusions can be retrieved. Since the necessity of obtaining a random musical dataset as possible to the elements extraction from a repository in order to create our own database, the challenges were many, what widered the horizons through which the study had to proceed. This way we started from the system creation and finished on its analysis.

Having established the elements to observe, its analysis revealed some interesting facts, being some of them easily noticeable from common sense about musical genres, whilst others were new and inovative. Although some information obtained was not new, the results are not to be neglected, showing some unexpected facts about musical similarities.

By applying similar classification methods to two different attribute sets in order to classify them into musical genres, disapoiting results were obtained by using the unrefined attributes from MFCC and FBEs. Beyond that fact, the junction are refinement of the whole attribute set took the efforts to a superior result in terms of success rate in classification. Although the mentioned sucess rate reached 79%, it can not be compared to other products with similar functionality, which take a small piece of a song and search for a correspondency on an existing dataset.

Referências

Cassidy, R. and Smith III, J. (2005). Auditory filter bank lab. `https://ccrma.stanford.edu/realsimple/aud_fb/aud_fb.pdf`.

Clark, S., Park, D. and Guerard, A. (2012). Music genre classification using machine learning techniques. `http://web.cs.swarthmore.edu/~meeden/cs81/s12/papers/AdrienDannySamPaper.pdf`.

Grubbs, F. (1969). Procedures for detecting outlying observations in samples. *Technometrics* **11**.

Han, J. and Kamber, M. (2006). *Data Mining Concepts and Techniques.* San Francisco: Morgan Kaufmann.

Logan, B. (2005). Mel frequency cepstral coefficients for music modelling. `http://www.citeulike.org/pdf_options/user/asterix77/article/2212484?fmt=pdf`.

Marzban, C. (2004). The roc curve and the area under it as performance measures. `http://journals.ametsoc.org/doi/pdf/10.1175/825.1`.

Redmond, S. and Heneghan, C. (2005). A method for initialising the k-means clustering algorithm using kd-trees. *Elsevier Science*.

Smith, L. (2002). A tutorial on principal components analysis. `www.ce.yildiz.edu.tr/personal/songul/file/../principal_components.pdf?`

Sumeet, D. and Xian, D. (2001). *Data Mining and Machine Learning in Cybersecurity.* CRC Press.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *Transactions on speech and audio processing* **10**.

Zweig, M. and Campbell, G. (1993). Receiver-operating charecteristic (roc) plots: A fundamental evaluation tool in clinical medicine. `http://www.clinchem.org/content/39/4/561.full.pdf+html`.