Automatic Music Classification into Genres

Gjorgji Madjarov[1], Goran Pesanski[2], Daniel Spasovski[2], and Dejan Gjorgjevikj[1]

[1] Faculty of Computer Science and Engineering
[2] Faculty of Electrical Engineering and Information Technologies
{gjorgji.madjarov,dejan.gjorgjevikj}@finki.ukim.mk
pesanski_goran@yahoo.com,sdaniel506@yahoo.com

Abstract. Musical genres are categorical labels created by humans to characterize pieces of music. Although music genres are inexact and can often be quite arbitrary and controversial, it is believed that certain song characteristics like instrumentation, rhythmic structure, and harmonic content of the music are related to the genre. In this paper, the task of automatic music genre classification is explored. Multiple features based on timbral texture, rhythmic content and pitch content are extracted from a single music piece and used to train different classifiers for genre prediction. The experiments were performed using features extracted from one or two 30 second segments from each song. For the classification, two different architectures flat and hierarchical classification and three different classifiers (kNN, MLP and SVM) were tried. The experiments were performed on the full feature set (316 features) and on a PCA reduced feature set. The testing speed of the classifiers was also measured.The experiments carried out on a large dataset containing more than 1700 music samples from ten different music genres have shown accuracy of 69.1% for the flat classification architecture (utilizing one against all SVM based classifiers). The accuracy obtained using the hierarchical classification architecture was slightly lower 68.8%, but four times faster than the flat architecture.

Keywords: music, genre, classification, flat, hierarchical

1 Introduction

A music genre is a conventional category that identifies pieces of music as parts of a set of rules and conventions.Although the artistic nature of music means that these classifications are often arbitrary and controversial, and some genres may overlap, this human made labels help people to better organize their music collections, based on their individual music perception and cognition, choose the music radio station to listen to, and plays important role in electronic music distribution.Music can be divided into different genres in several ways. There are many music genres,such as classical, rock, pop, disco, etc. The music genre, because of its flexibility, is constantly exposed to changes and as result many fused genres are produced that creates a genre classification hierarchy. It is determined that many of the elements that belong to an audio signal can be used as features that are needed for music classification. These features include the spectral characteristics of the audio signal, timbre, pitch, tempo, energy distribution, rhythm or other content[1][2][3][4]. The majority of the research projects are focused exactly on providing better methods for feature extraction[4][5][6]. Though the concept of musical genre might not be well defined, recent approaches that use audio feature extraction combined with machine learning techniques have achieved promising results[1][2][6][7]. Nowadays, the digital music databases, mostly located on the Web, become more popular both for professional and private purposes. The need for automatic organization and classification increases every day. Manual or even semi-automated annotation of each music file is impractical, expensive and time consuming approach. This problem inspires the computer scientist and researches, together with the music workers, to work on a solution. The automatic music genre classification is a big challenge for many scientists and researchers. Different classification techniques for automatic music genre classification, such as Support Vector Machines (SVM)[1][2][7][8], Artificial Neural Networks (ANN)[6][9], Hidden Markov Models (HMMs)[10] and Gaussian Mixture Models (GMMs)[11] have been used by different researches. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals)[12][13] as a framework for audio processing and speech analysis with specific emphasis on music information retrieval applications was used for the task of feature extraction in our research.

In this paper, we address the problem of automatic music genre classification in ten different genres. More specifically, three different sets of features represented by timbral texture, rhythmic content and pitch content are used for classifying ten different music genres.Two classification architectures (flat and hierarchical) are experimentally evaluated,using three different types of base classifiers: SVM [8], Multilayer Perceptron (MLP) [14] and k - Nearest Neighbours (kNN) [15].

Section 2 presents the feature extraction process and the features used in our experiments. Section 3 describes the datasets used in the experiments and the experimental setup. The experimental results are presented and discussed in Section 4. Finally, the conclusions are given in Section 5.

2 Feature Extraction and Datasets

2.1 Feature Extraction

The main focus of our research is the classification methods and techniques for music genre classification and not the feature extraction. As mentioned in the previous section, the feature extraction was performed using the MARSYAS [3] tool. Two different types of features were extracted. The first set consists of 31 timbral texture features extracted from each music sample including: time-domain Zero-Crossings (1), SpectralCentroid (1), Rolloff (1), Flux (1), Chroma (14) and Mel-FrequencyCepstralCoefficients MFCC (13) over a texture window

[3] http://marsyas.info/

of 1 sec. The second set consists of 48 Spectral features: SpectralFlatnessMeasure (24) and SpectralCrestFactor (24). Each feature extracted by MARSYAS is represented by 4 separate values,so the actual length of the feature vector is four times the number of features. Each music sample is represented by 79 features each represented by 4 values.

2.2 Datasets

In the paper we concentrate on analyzing the performance of different classification techniques for the problem of automatic music genre classification. Our goal was to classify music files in wav format according to their genre. We used the same ten music genres, used by Tzanetakiset al. [1][2]. The selected genres include: Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock. 2760 music songs (23 hours of audio data) were collected and used in the experiments. All songs were stored as 22.5 kHz mono, 352 kbps wave files. For splitting the song wave files before the feature extraction WavSplit 1.2.1 for Linux [16] was used.

Three different datasets were used in this analysis. All three datasets are composed of 1000 instances for training and 1760 instances for testing. The exact distribution of the training and testing instances regarding the genres are shown in Table 1.

Table 1. The distribution of the training and testing instances regarding the genres

	Blues	Classical	Country	Disco	Hip-Hop	Jazz	Metal	Pop	Reggae	Rock
Training	100	100	100	100	100	100	100	100	100	100
Testing	167	94	101	159	240	139	150	179	131	400
Total	267	194	201	259	340	239	250	279	231	400

Each instance from the first dataset is represented by 30 second segment (between the 30th and 60th second) of the actual music song,referred as music sample. Each music sample was described by timbral texture features (124 features)and spectral features (192 features) mentioned in the previous section, which results with 316 features in total per sample.This data set is denoted as one sample features data set. Unlike the first dataset, in the second dataset,each song was represented by two music samples (both 30 seconds), the first beginning at the 30% of the duration of the song and the second at 60% of songs duration. Features (timbral texture and spectral) are extracted separately from each sample of the song and then the resulting features are concatenated in a single feature vector representing that instance. We denote the second data set as two sample features data set. In order to reduce the length of the feature vector obtained by concatenating the features from the two parts of the song as discussed before, in the third dataset we performed the PCA (Principal Component Analysis) feature selection method [17][18]. In this manner, the length of an instance in the third dataset was reduced from 612 features to 151 features.

3 Experimental Setup and Results

3.1 Experimental Setup

The comparison of the classification methods (SVMs Support Vector Machines, MLP Multilayer Perceptron and kNN k Nearest Neighbours) was performed using their implementations in Weka [19]. For training the SVMs, we used the SMO implementation. In particular, we used SVMs with a radial basis kernel. The kernel parameter gamma and the penalty C, for each combination of dataset and method, were determined by 10-fold cross validation using only the training set. The values 2^{-15}, 2^{-13}, , 2^{1}, 2^{3} were considered for gamma and 2^{-5}, 2^{-3}, ,2^{13}, 2^{15} for the penalty C. The number of neighbours in the kNN method for each dataset was determined from the values 1 to 9 with step 2. The Neural Networks are represented by MLP with 25 neurons in the hidden layer and value for the validation threshold of 10. After determining the best parameters values for each method on every dataset by 10-fold cross validation, the classifiers were trained using all available training examples and were evaluated by recognizing all test examples from the corresponding dataset. Two different architectures (flat and hierarchical) are considered and explored in our work. The following subsections include the brief description and the results obtained from each classifying architectures.

3.2 Experimental Results

Flat Classification The flat classification addresses the problems where the predefined classes are separately treated and there is no structure defining the relationships among them (or that structure is not considered even if it exists). According to this,we do not take into account the possible relationships between the classes for the purpose of the flat classification.

The 10-genre classification is performed by classifying the music samples in their appropriate genre, using the classifiers mentioned previously. One performance evaluation measure (accuracy)and the testing time measured in seconds were used to estimate the performance of the different classifiers.

Table 2 shows the results of the three classifiers on the first and the second dataset. One instance from the first dataset was represented by only one music sample per song, described by 316 features, while an instance from the second dataset was represented by two music samples per song, described by 316 features each.The first column of the table describes the classification genres. The other columns show the accuracy of the classifiers per genre. The first group of three columns shows the performance obtained on the first dataset, while the second group of three columns shows the performance obtained on the second dataset. The last two rows present the overall accuracy of the classifiers per dataset and the testing times accordingly.The best prediction results are achieved by SVM for both datasets.The MLP classifier showed similar performance results, but its testing time is about two times longer than the SVM classifier for the first dataset and 1.8 times longer for the second dataset. For some particular genres

MLP is even more accurate than the SVM. This is the case for the half of the genres for the first dataset and 4 out of 10 genres for the second dataset. Blues is the genre that decreases the overall accuracy of the MLP. The KNN classifier, compared to the SVM and MLP, provides lower accuracy for almost every genre.

Table 2. Classification accuracy comparison between different classifiers (accuracy in %)

	One sample features			Two sample features		
	SVM	KNN	MLP	SVM	KNN	MLP
Blues	50.90	33.53	36.53	43.11	32.34	35.93
Classical	91.49	80.85	88.30	91.49	82.98	86.17
Country	64.36	67.33	70.30	70.30	68.32	70.30
Disco	64.15	62.89	60.38	62.89	64.78	64.78
Hip-Hop	84.58	79.58	84.58	89.58	81.25	85.42
Jazz	70.50	48.92	72.66	80.58	49.64	71.22
Metal	82.00	75.33	84.67	87.33	82.00	88.67
Pop	32.40	22.35	36.31	32.40	24.02	29.05
Reggae	67.18	66.41	70.23	63.36	68.70	70.99
Rock	68.50	44.50	66.50	72.00	40.00	70.75
Accuracy(%)	67.16	55.51	66.19	69.09	55.91	67.05
Time(s)	34	29	63	79	62	151

Table 3 shows more detailed information about the musical genre classifier performance in the form of a confusion matrix. In a confusion matrix, the columns correspond to the actual genre and the rows to the predicted genre.The relative distribution of the values in the confusion matrix for the two other classifiers is very similar. These matrices show that Classical music instances are classified with the highest accuracy. On the other hand, Blues, Pop and Rock happen to be the genres that are most often confused with the others. For example, Blues is often mixed with Country, Rock and Pop music. Pop is mixed with Disco, Rock and Country music and etc.

Table 4 shows the confusion matrix for the second dataset. It can be noticed that the number of correctly classified instances increased, especially in Jazz, Rock and Hip-Hop genres. This led to improving the overall percentage of correctly classified instances using all classifiers, SVM being the best with classification accuracy of 69.1%.

For the third dataset (where PCA dimensionality reduction was performed), only the performance of the SVM classifier was measured. As we expected, the testing time of the classifier was shortened, while the accuracy slightly decreased. In particular, classifying the whole test set required 25 seconds, and the obtained accuracy was 65.75%.

Hierarchical Approach to Music Genre Classification Hierarchical classification refers to assigning samples to a suitable class from a hierarchical class

Table 3. Music genre confusion matrix using SVM classifier on test dataset

	Blues	Classical	Country	Disco	Hip-Hop	Jazz	Metal	Pop	Reggae	Rock
Blues	**85**	2	18	13	0	9	2	15	5	17
Classical	5	**86**	1	1	0	1	0	0	0	0
Country	25	1	**65**	1	0	2	0	1	2	4
Disco	4	0	9	**102**	5	3	0	27	0	9
Hip-Hop	1	0	4	12	**203**	3	1	15	0	1
Jazz	11	9	0	5	1	**98**	12	0	0	3
Metal	3	5	1	1	0	3	**123**	3	0	11
Pop	20	2	23	34	6	3	5	**58**	1	27
Reggae	11	0	6	14	1	3	1	2	**88**	5
Rock	12	9	23	49	1	3	19	10	1	**274**

Table 4. Music genre confusion matrix using concatenated features from two samples

	Blues	Classical	Country	Disco	Hip-Hop	Jazz	Metal	Pop	Reggae	Rock
Blues	**72**	0	23	11	0	17	3	16	2	22
Classical	1	**86**	1	2	0	2	0	0	0	2
Country	16	1	**71**	4	1	4	0	1	1	2
Disco	6	0	7	**100**	8	4	2	25	0	7
Hip-Hop	1	0	0	6	**215**	4	0	11	0	3
Jazz	11	1	1	2	1	**112**	9	0	1	1
Metal	3	3	1	1	0	2	**131**	1	0	8
Pop	23	3	34	27	9	3	2	**58**	1	19
Reggae	8	0	2	21	5	3	0	3	**83**	6
Rock	18	2	29	32	0	6	19	7	0	**288**

space [8]. By utilizing the previously defined hierarchical architecture,the classification problem can be decomposed into a smaller set of problems. In this approach the classification is accomplished with the cooperation of classifiers built at each level of the tree.One of the obvious problems with the top-down approach is that a misclassification at a parent class may force a sample to be misrouted before it can be classified into the correct child classes.

In many classification experiments, the hierarchical approach can lead to better results in the multi-class classification process. Fig. 1 shows the 2-level hierarchy that we considered in the experiments. Each test instance is passed through the hierarchical architecture of classifiers resulting in an instance classified in one of the 10 music genres.

The first level of the hierarchy consists of 4 nodes illustrating the most distinctive groups of music genres, based on the confusion matrices obtained from the flat classification. Classical music, as the most distinctive genre in flat classification, represents one node in the hierarchy. The other three nodes contain groups of genres that are similar to each other and often mutually misclassified by the classifiers. Hierarchical classification architectures of the three different classifiers discussed previously were trained and applied to the music genre classification problem.

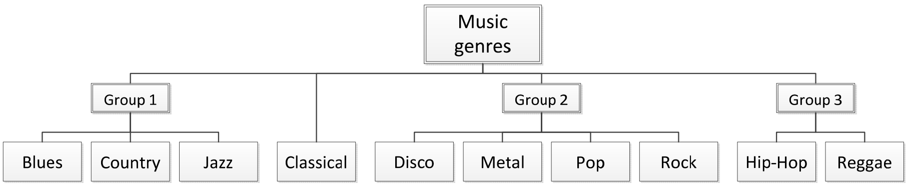

Fig. 1. 2-level music genre hierarchy

For the hierarchical architecture we present only the results of the best classifier. Table 5 shows the results at the 1^{st} and 2^{nd} level obtained by the hierarchy of SVM classifiers. It can be seen that the overall accuracy is slightly lower (only 0.2%) compared to the flat classification, but the testing time is significantly improved (more than four times). It can also be noted that for particular genres as Classical, Rock and Pop the accuracy is improved, especially for the experiments where concatenated features are used.

Table 5. Results from the hierarchy provided by the SVM classifier (accuracy in %)

	1st level		2nd level	
	One sample features	Two sample features	One sample features	Two sample features
Classical	87.23	92.55	87.23	92.55
Blues	76.17	75.18	43.11	41.32
Country			60.40	63.37
Jazz			64.75	74.10
Disco	86.26	87.73	60.38	61.64
Metal			83.33	88.67
Pop			31.84	35.75
Rock			71.00	73.50
Hip-Hop	81.67	82.21	86.25	88.33
Reggae			70.23	66.41
Accuracy(%)	83.01	83.92	66.25	68.81
Time(s)	7	15	9	19

4 Conclusions and Future Work

In this paper, we address the problem of automatic music genre classification.Three different sets of features represented by timbral texture, rhythmic content and pitch content were used for classifying ten different music genres. Two classification architectures (flat and hierarchical) were evaluated experimentally, using three different types of base classifiers: SVM, Multilayer Perceptron and k Nearest Neighbours.

SVM one-versus-all showed the best predictive performance comparing to the kNN and MLP classifiers. The SVM classifier showed predictive accuracy of 67.16% for the case where single 30 second music segment per song was used to build the model. The performance increased for additional 2% when features extracted from two 30 second segments from each song were used, but this also slowed down the prediction by more than 2 times.

On the other hand, the hierarchical approach, that was justified based on the similarities between the musical genres associated with the rhythm, harmony and pitch, showed significant improvements in the testing time comparing to the flat classification approach(more than four times), while showing only slightly lower (0.2%) predictive accuracy.

Future work will involve further analysis of the feature space, genre group dependent selective extraction and combination of different types of features on the second level of the classification hierarchy,examination of alternative classification schemes, and incorporation of more audio classes.We will also try to transform this problem into multi-label one and solve it by commonly used multi-label classification techniques.

References

1. George Tzanetakis, P.C.: Music genre classification of audio signals. IEEE Transactions on speech and audio processing **10** (2002) 293–302
2. George Tzanetakis, Georg Essl, P.C.: Automatic music genre classification of audio signals. In: 2nd Annual International Symposium on Music Information Retrieval. (2001)
3. Tao Li, Mitsunori Ogihara, Q.L.: A comparative study on content-based music genre classification. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. (2003) 282 289
4. Thomas Lidy, A.R.: Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: Proceedings of the 6th International Conference on Music Information Retrieval. (2005) 34–41
5. Michael Scott Cuthbert, Christopher Ariza, L.F.: Feature extraction and machine learning on symbolic music using the music21 toolkit. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. (2011) 387–392
6. Neumayer, R.: Musical genre classification using a multi-layer perceptron. In: Proceedings of the 5th Workshop on Data Analysis (WDA'04), Elfa Academic Press (2004) 51–66
7. Ran Tao, Zhenyang Li, Y.J.: Music genre classification using temporal information and support vector machine. In: ASCI Conference. (2010)
8. Chih-Wei Hsu, Chih-Chung Chang, C.J.L.: A Practical Guide to Support Vector Classication. Department of Computer ScienceNational Taiwan University, Taipei 106 (2010)
9. Aliaksandr Paradzinets, Hadi Harb, L.C.: Multiexpert system for automatic music genre classification. Technical report, Ecole Centrale de Lyon - Departement MathInfo (2009)
10. Karpov, I.: Hidden markov classification for musical genres. Technical report, Rice University (2002)

11. Andre Holzapfel, Y.S.: A statistical approach to musical genre classification using non-negative matrix factorization. IEEE Transactions on Audio, Speech, and Language Processing **16** (2008) 424–434
12. Tzanetakis, G.: Marsyas (2012) Software tool for Music Analysis, Retrieval and Synthesis for Audio Signals (Version 0.4.5).
13. Tzanetakis, G.: Marsyas User Manual. (2012)
14. Bishop, C.M.: Neural Networks for Pattern Recognition. 1 edn. Oxford University Press, USA (1996)
15. Bay, S.D.: Nearest Neighbor Classification from Multiple Feature Subsets. Intelligent Data Analysis **3** (1998) 191–209
16. Weihmann, T.: Wavsplit (2002) Software which splits large WAV files at given time positions (Version 1.2.1.).
17. Joakim Anden, S.M.: Multiscale scattering for audio classification. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. (2011) 657–662
18. Philippe Hamel, Simon Lemieux, Y.B.D.E.: Temporal pooling and multiscale learning for automatic annotation and ranking of music audio. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. (2011) 729–734
19. University of Waikato, N.Z.: Weka (1997) Machine learning software written in Java (Version 3.6.).