Music Genre Recognition in the Rough Set-Based Environment

Piotr Hoffmann[✉] and Bożena Kostek

Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory,
Gdańsk University of Technology, 80-233 Gdańsk, Poland
phoff@sound.eti.pg.gda.pl, bokostek@audioacoustics.org

Abstract. The aim of this paper is to investigate music genre recognition in the rough set-based environment. Experiments involve a parameterized music database containing 1100 music excerpts. The database is divided into 11 classes corresponding to music genres. Tests are conducted using the Rough Set Exploration System (RSES), a toolset for analyzing data with the use of methods based on the rough set theory. Classification effectiveness employing rough sets is compared against k-Nearest Neighbors (k-NN) and Local Transfer function classifiers (LTF-C). Results obtained are analyzed in terms of global class recognition and also per genre.

Keywords: Music processing · Rough sets · Genre recognition · k-Nearest Neighbors · RSES system · LTF-C

1 Introduction

There exist many methods that can be used for data storage, analysis and classification. The main feature of these methods should be universality and efficiency. With regard to universality a system should allow for collecting and storing various data sets, regardless of the processes and phenomena described in them. The effectiveness of the system should enable users to make data analysis and classification easily, and to control these processes. Handling data efficiently requires storing them in tables with objects (rows) and attributes (columns), describing single instances. Data in their nature could be imprecise, uncertain and/or incomplete, thus this requires special preparation so that they can be processed, mined and classified. Another important issue is to select significant components present in the tables to provide their discernibility within the classes. For the analysis of data with characteristics described above the rough set-based methods are very useful, as they generate interpretable results in the form of reducts and rules. One of the well-known systems for the analysis and classification of data is the Rough Set Exploration (RSES) system that returns extracted rules and reducts acquired from the rough set-based analysis [3, 4]. It is worth to emphasize that the RSES system is a software tool that enables to carry out large-scale computational experiments related to the analysis of array data using the rough set theory [13, 16].

In this publication, the authors illustrate the process of recognizing music genres using the rough set theory. One of the main objectives of the data analysis shown in this paper is to uncover underlying causes or factors and to determine the relationship between objects (audio tracks belonging to music genres) in the case study related to the Music Information Retrieval (MIR) domain [5]. The classification is carried out on a set of music descriptors using the methods available in the RSES software. In addition, data are to undergo a pre-processing of feature vector parameters employing the PCA method [1, 15] and parameter weighting. Lastly, a comparison of genre classification results employing two sets of music excerpts is provided.

2   Data Preprocessing

The theory of rough sets was created in the early 80s of the twentieth century. Its main use is for synthesizing and analyzing data sets efficiently. Methods based on the rough set theory have been used, among others, in data mining and knowledge discovery in complex tasks of classification and computer decision support systems [13]. Currently, it is one of the fastest growing methods within the artificial intelligence domain. In the rough set theory a requirement that a data set needs to have a clearly defined boundaries is discarded [2, 17]. The scope of rough sets is defined by the lower and upper approximations of tabular data, obtained experimentally.

The difference between the upper and lower approximation is the border area, which includes all cases, that cannot be seamlessly classified on the basis of the current knowledge. The lower approximation set contains all objects for which there is no doubt that they are representatives of this set in view of knowledge. Objects that cannot be excluded that they are representatives of this set belong to the upper approximation. Boundary of the set are all of the objects for which it is not known whether or not they are representatives of a given set. The larger area border set, the more objects in it are less precise. The theory of rough sets allows the processing of both quantitative and qualitative data tables, called decision tables.

The basic structure of the data in the information systems using rough set theory is a table. All data are grouped in tables according to the principle that the rows of tables are objects, and attributes are columns. The formula (1) presents the information system [17].

$$SI = \ < U, A, V, f >$$

(1)

where:

U - non-empty, finite set of objects,
A - non-empty finite set of attributes,
V - set of attribute values,
f - function of information, which is the Cartesian product of a set of objects and a set of attribute values.

In Music Information Retrieval systems tables filled with music descriptors constitute the information system. In such systems, each track is parameterized and then stored in

a table [5, 8]. A special case of information systems are decision tables, which describes cases (also called examples or objects) using conditional attribute values and a decision. Attributes are independent variables while the decision is a dependent variable, which means that the conditional attributes determine the value of a decision.

Based on the decision table only, it is impossible to directly know the relationship between the conditional attributes and decision describing objects. Therefore, it is necessary to further process them to extract dependencies. Attribute reduction is very crucial in the rough set-based data analysis because it is used to induce decision rules without reducing the classification accuracy [19]. In the rough set theory the reduct is generally defined as a minimal subset of attributes that can classify the same domain of objects as unambiguously as the original set of attributes, which means that the reduct is a minimal subset of attributes having the characteristics of the whole collection. For a given information system may have multiple reducts, consisting of a variable number of arguments. The major problem is, however, the identification and removal of attributes that are unnecessary. The process of determining reducts is considered as a bottleneck in the inference systems based on rough sets. On the other hand, reducts and decision making may be acquired for large systems of dozens or hundreds of attributes using genetic algorithms. After preparing reducts it is possible to generate data available in the decision-making system in the form of logical rules used in the classification process. The rules have the conditional form, and their number is equal to the maximum number of objects multiplied by the number of distinguishable reducts. However, not all rules are needed to use or to be implemented in the decision-making process. The reduction method applied in the RSES system was very thoroughly described by its authors [2], thus it will not be recalled here.

The main element of music genre recognition systems is the optimized parameter input. The extracted feature vector should have a detailed description of parameterized samples and preserve a very good separability. Taking into account these assumptions feature vector containing 173 elements has been prepared. The vector includes parameters associated with the MPEG 7 standard [12] and the melcepstral parameters [8–10]. The list of parameters include: Spectral flatness Measure Spread Spectrum Audio, Audio Spectrum Envelope, Spectral Centroid, Temporal centroid. Full list of parameters was shown in the study [18]. Frequency band used for the parameter is in the range from 63 to 8000 Hz. The prepared feature vector is used to describe each signal frame.

173-element vector generates a very large amount of information describing a song. Consequently, this leads to an extensive amount of data undergoing classification, which in the context of the usage of e.g. the k-Nearest Neighbors classifier is important. It was therefore decided to apply Principal Component Analysis (PCA) to reduce data redundancy [1]. This is to identify patterns in the data and present them in such a way as to indicate their similarities and differences. The PCA method uses the variance of the data to prepare a new database parameters. The new descriptors are linear combinations of parameters that carry much information about the test set. It was experimentally checked that the PCA method can shorten the given feature vector of 173 descriptors to 19, which significantly reduces the computation time. In addition, the use of the described analysis improves the classification efficiency, which was presented in an earlier paper by the authors [7].

3   Experiments

This Section describes the results of the experiments in which the rough set theory was applied to music genre classification. For this purpose rough set classification learning algorithm provided by RSES was used. Experiments aimed at a comparison between standard classification algorithms - k-Nearest Neighbors and Local Transfer function classifiers [10, 11]. The rule set decision algorithm based on conditional rules calculates the attributes of the new object, which is essential for decisions related to the content of reducts. Then, it looks for rules that match attribute values, if there are no matching rules, the result is the most common decision, or the least expensive decision. In the case that multiple rules match attribute values, they may indicate a number of decisions, then the vote should be taken that selects the answer that appears most often [6].

The k-Nearest Neighbors algorithm is the simplest one, and as such is very commonly used for classification. We utilized the minimal distance that uses the Euclidean distance function. Its aim is to predict the class membership of objects. The decision is based on the k-closest objects. An object is classified by the majority vote [14].

LTF-C algorithm is a neural network employed for classification tasks with the architecture similar to the radial network (RBF), but different training algorithms. It consists of two layers of neurons. The first layer - hidden - contains neurons with Gaussian transfer function, that detects cluster of patterns of the same class in the training data. Each neuron of this layer is assigned a class that tries to detect the cluster [10]. The second layer consist of linear neurons that segregate responses of hidden neurons

Table 1.   Number of excerpts in the Synat and GZTAN databases

Genre:	Synat	GZTAN
Pop	100	100
Rock	100	100
Country	100	100
R&B	100	–
Rap & Hip-Hop	100	100
Classical	100	100
Jazz	100	100
Dance & Dj	100	100
NewAge	100	–
Blues	100	100
Hard Rock & Metal	100	100
Reggae	–	100

according to the assigned classes and add them by formulating a final answer network structure [10].

All classification tests were carried out in the RSES environment. Tests were performed on two data sets. The first one, called "Synat", contained 1100 audio excerpts divided into 11 most popular music genres. The second one GZTAN [17], a commercial data set, contained 1000 audio files divided into 10 music genres. Sizes and content of these two sets are presented in Table 1 [7, 8]. The length of each music excerpt is 30 s. Both datasets were created in a similar way to reflect a variety of music genres and contain audio files belonging to the most popular music genres.

In preparing parameters for the classification the data volume was reduced using the method of PCA. The number of parameters after employing PCA was 33, which accounted for 80 % of information retained from the entire feature vector. Figure 1 presents a block diagram of the proposed processing path of music genre according to rough set-based analysis.
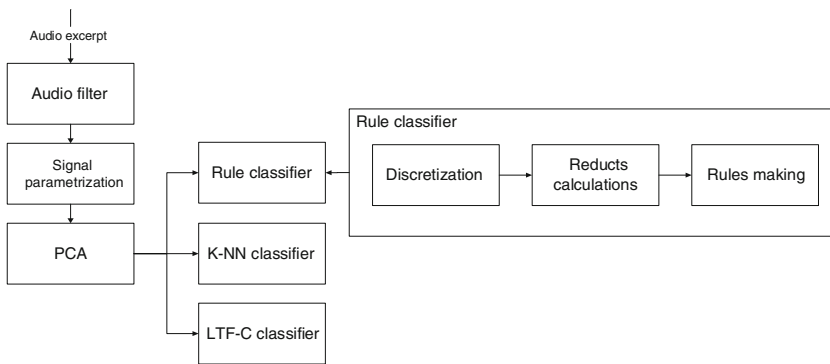
Fig. 1. Block diagram of processing in the experiment

Classification in the testing phase was performed with the default settings of the algorithms used. The RSES system automatically chooses the optimum values of parameters for the most effective results. This confirms the validity of the additional data preparation prior to processing information using the rough set theory. In Table 2 the results of tests conducted are shown. The table shows the results for the feature vectors without the PCA method and with the use of PCA. At the same time, it should be noted that feature vectors not reduced by the PCA method were processed by algorithms for much longer.
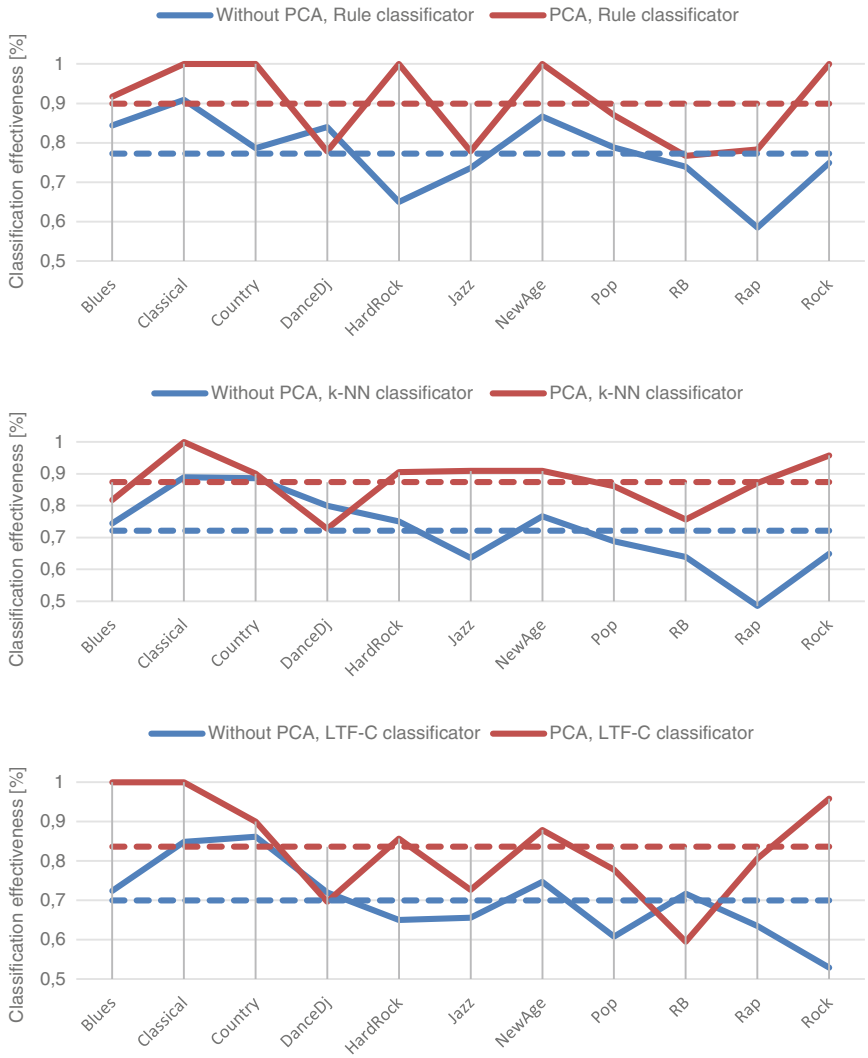
The effectiveness of the classification algorithms reached 70 % when the PCA method was not used and 85 % when it was employed. Analyzing the results obtained for individual genres, it can be noticed that classical music genre is distinguishable among genres as it has got a very good 90 % classification effectiveness for each test set. Similarly, very good results were achieved for rock and hard rock genres with the use of the PCA method.

Table 2. Classification effectiveness of music genres in Synat dataset.

Genre [%]	Rule classification		k-NN		LTF-C	
	No PCA	PCA	No PCA	PCA	No PCA	PCA
Blues	0.844	0.917	0.744	0.818	0.724	1
Classical	0.909	1	0.889	1	0.849	1
Country	0.786	1	0.886	0.9	0.862	0.9
DanceDj	0.84	0.778	0.8	0.727	0.72	0.697
HardRock	0.65	1	0.75	0.905	0.65	0.857
Jazz	0.736	0.778	0.636	0.909	0.656	0.727
NewAge	0.867	1	0.767	0.909	0.747	0.879
Pop	0.788	0.87	0.688	0.861	0.608	0.778
RB	0.739	0.767	0.639	0.757	0.717	0.595
Rap	0.585	0.783	0.485	0.871	0.635	0.806
Rock	0.749	1	0.649	0.958	0.529	0.958
$\overline{\Sigma}$	0.772	0.899	0.721	0.874	0.699	0.836

Table 2 clearly shows the gain after applying the method using the rough set theory. On average it is about 5 % better than while employing two other methods, which should be considered a very good result despite a much longer data processing. Longer data processing in classification systems based on the theory of rough sets is due to the discretization step and generation of reducts. In particular, the step of reduct generating is a very demanding for the available resources. In Fig. 2 the results obtained in the experiments in the form of graphs were summarized. The dashed line presents the average value of the individual results.

The RSES system occurred to be the most effective classification algorithm, the weakest algorithm in ranking was the neural network. It was also the least balanced in its indications, i.e. recognizing individual music genres with a large discrepancy. The k-Nearest Neighbors algorithm was already investigated by the authors in the classification of music genres. The publication cited in here showed that the k-NN algorithm achieved the best results on music parameterized data [7]. However, the application of rough sets in the classification process of music genres with respect to existing algorithms that were used by the authors resulted in about 5 % better performance, which should be considered as a very good result, because this is due to only the change of the classification method without additional processing.

Fig. 2. Classification effectiveness of the algorithms employed

To confirm the results obtained by the authors another experiment was conducted using a commercially available GZTAN database [17]. In the experiment, eight music genres common for both databases were compared in the context of the classification efficiency. The experiment uses the k-NN algorithm. In Fig. 3 results obtained from the experiments for various music genres are shown.

The results obtained for the GZTAN database compared to the SYNAT database are 7 % lower. Similarly to the database SYNAT in the case of database GZTAN an increase in the effectiveness of the recognition genres after PCA use can be noticed. Significant differences in the recognition of music genres were found in the case of DanceDj genre,
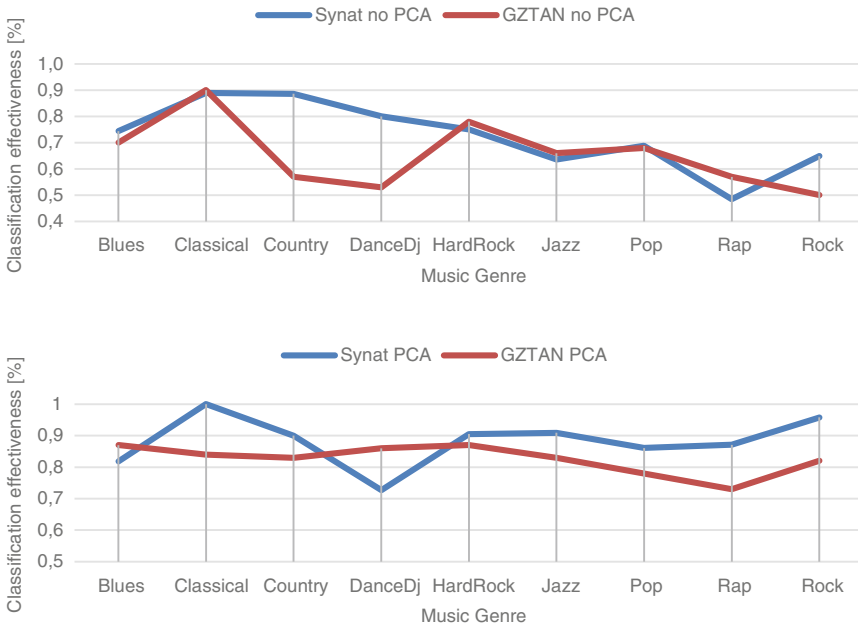
Fig. 3. Classification effectiveness of the k-NN algorithm using two different data sets.

which can be caused by a different description of DanceDj genre in GZTAN database. The reason for the lower classification efficiency is probably a greater variety of songs in the GZTAN database. Tracks occurring in the database fully describe the variety of music genres. Furthermore the database contains also recordings with reduced quality, which may have an adverse effect on the effectiveness of the parameterization of the proposed solution.

In order to confirm the statistical significance of the obtained results the T-Student test was performed. Values were calculated for each analyzed genre. To reject the null hypothesis of statistical insignificance of the results T parameter should be higher than 2.228. The critical value is based on the value from the T-Student distribution table for 10 degrees of freedom. In Table 3 the exact values of the parameter T for all experiments are shown. Seven of eight results of experiments can be considered as statistically significant. Small variations obtained in these results show the correctness of the results in the statistical sense. The only experiment that did not show statistical significance of the results was carried out based on the GZTAN database without PCA.

Table 3. Summary of the results of T-Student Test

	Rule classification		*k*-NN		LTF-C		*k*-NN GZTAN	
	No PCA	PCA	No PCA	PCA	No PCA	PCA	No PCA	PCA
T-value	2.422	2.825	2.463	2.387	2.439	2.557	2.261	2.446

To summarize, it may be concluded that rule decision algorithm gives in most cases better results than minimal distance or neural algorithms. But this is associated with a significantly longer duration of data processing. Moreover, in the case of a trial experiment conducted by the authors when a very large database with more than 30000 parameterized music excerpts was used, the system was not able to calculate reducts due to lack of the application memory resources. The test confirms earlier findings on the computational complexity of the process of generating reducts. The additional use of the PCA method in the decision process further improved the effectiveness of the classification decision by reducing the number of attributes.

4   Conclusions

In the experiments conducted, the authors used the RSES application for testing the effectiveness of recognizing music genres using the rough set theory. The application enables to efficiently carry out data analysis and generates classification tables. The authors examined also the effect of using the PCA method in data processing according to the theory of rough sets. Experiments were conducted on two different data sets.

The classification effectiveness achieved in the experiments is very good (above 85 %) for the whole data set. In other publications related to music genre classification [10, 11, 20] when *k*-NN and rule classification algorithms were used, the results were worse. The main reason for such good results obtained by the authors may be a unique parameterization module which was applied along with the PCA dimensionality reduction. Very good results have been achieved through the effective parameterization, accurately describing the test set. The data after analysis with the PCA method were very well separable and were described by a much smaller number of data without losing much information. Calculating reducts from the PCA-reduced data improves the classification effectiveness up to 10 %. Performed statistical analysis confirmed the statistical significance of results. Most of experiments can be defined as statistically significant.

Performed test show how important in decision making process is the data preprocessing step. Discretization of tabular data, reducts and rules calculation require operation time and are resource-absorbing, but for smaller sets data processing based on rough sets can be more effective, regardless of the resources involved.

References

1. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdisc. Rev. Comput. Stat. **2**, 433–459 (2010)

2. Bazan, J.G., Nguyen, H.S., Nguyen, T.T., Skowron, A., Stepaniuk, J.: Decision rule synthesis for object classification. In: Orłowska, E. (ed.) Incomplete Information: Rough Set Analysis, vol. 13, pp. 23–57. Physica - Verlag, Heidelberg (1998)

3. Bazan, J., Szczuka, M.S., Wróblewski, J.: A new version of rough set exploration system. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 397–404. Springer, Heidelberg (2002)

4. Kostek, B.: Music Information Retrieval in Music Repositories. In: Skowron, A., Suraj, Z. (eds.) Rough Sets and Intelligent Systems - Professor Zdzisław Pawlak in Memoriam. ISRL, vol. 42, pp. 463–489. Springer, Heidelberg (2013)

5. Kostek, B.: Perception-Based Data Processing in Acoustics, Applications to Music Information Retrieval and Psychophysiology of Hearing. Cognitive Technologies. Springer, Heidelberg (2005)

6. Kostek, B.: Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg (1999)

7. Hoffmann, P., Kostek, B.: Music data processing and mining in large databases for active media. In: Ślęzak, D., Schaefer, G., Vuong, S.T., Kim, Y.-S. (eds.) AMT 2014. LNCS, vol. 8610, pp. 85–95. Springer, Heidelberg (2014)

8. Kostek, B., Hoffmann, P., Kaczmarek, A., Spaleniak, P.: Creating a reliable music discovery and recommendation system. In: Bembenik, R., Skonieczny, L., Rybiński, H., Kryszkiewicz, M., Niezgódka, M. (eds.) Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation. SCI, vol. 541. Springer, Switzerland (2013)

9. Kostek, B., Kupryjanow, A., Zwan, P., Jiang, W., Raś, Z.W., Wojnarski, M., Swietlicka, J.: Report of the ISMIS 2011 contest: music information retrieval. In: Kryszkiewicz, M., Rybinski, H., Skowron, A., Raś, Z.W. (eds.) ISMIS 2011. LNCS, vol. 6804, pp. 715–724. Springer, Heidelberg (2011)

10. Kotropoulos, C., Benetos, E., Panagakis, E.: Music genre classification: a multilinear approach. In: ISMIR (2008)

11. Mlynek, D., Zoia, G., Scaringella, N.: Automatic genre classification of music content. IEEE Signal Process. Mag. **23**(2), 133–141 (2006)

12. MPEG 7 standard. http://mpeg.chiariglione.org/standards/mpeg-7

13. Pawlak, Z.: Rough sets. Int. J. Comput. Inform. Sci. **11**(5), 341–356 (1982)

14. Schnitzer, D., Flexer, A., Widmer, G.: A fast audio similarity retrieval method for millions of music tracks. Multimedia Tools Appl. **58**, 23–40 (2012)

15. Shlens, J.: A Tutorial on Principal Component Analysis, Version 2, 10 December 2005

16. Skowron, A., Polkowski, L. (ed.): Rough Sets in Knowledge Discovery, vols. 1 and 2, Physica Verlag, Heidelberg (1998)

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

18. Tzacheva, A.A., Bell, K.J.: Music information retrieval with temporal features and timbre. In: An, A., Lingras, P., Petty, S., Huang, R. (eds.) AMT 2010. LNCS, vol. 6335, pp. 212–219. Springer, Heidelberg (2010)

19. Wróblewski, J.: Covering with reducts - a fast algorithm for rule generation. In: Polkowski, L., Skowron, A. (eds.) RSCTC 1998. LNCS (LNAI), vol. 1424, pp. 402–407. Springer, Heidelberg (1998)

20. Zheng, J., Oussalah, M.: Automatic System for Music Genre Classification, University of Birmingham, Electronics, Electricial and Computer Engineering (2006)