Effective Training Methods for Automatic Musical Genre Classification

Eren Atsız, Erinç Albey and Enis Kayış
Industrial Engineering, Özyeğin University, Istanbul, Turkey

Keywords: Music Information Retrieval, Genre Classification.

Abstract: Musical genres are labels created by human and based on mutual characteristics of songs, which are also called musical features. These features are key indicators for the content of the music. Rather than predictions by human decisions, developing an automatic solution for genre classification has been a significant issue over the last decade. In order to have automatic classification for songs, different approaches have been indicated by studying various datasets and part of songs. In this paper, we suggest an alternative genre classification method based on which part of songs have to be used to have a better accuracy level. Wide range of acoustic features are obtained at the end of the analysis and discussed whether using full versions or pieces of songs is better. Both alternatives are implemented and results are compared. The best accuracy level is 55% while considering the full version of songs. Besides, additional analysis for Turkish songs is also performed. All analysis, data, and results are visualized by a dynamic dashboard system, which is created specifically for the study.

1 INTRODUCTION

In order to understand the characteristics of music, musical genres are used as labels. Identifying the depth nature of music is a multifaceted concept. Musical genres are not depended on harsh rules, they are open boundaries, which are affected by societies, cultures and diverse factors. Automatic music genre classification has attracted much attention recently. Genre classification is not only intriguing for academic studies but also in real life. Today, recommendation engines work directly based on music genre classification. Therefore, classifying genres accurately in a wide extent is crucial to come up with a successful recommendation system. The best example for this matter can be Spotify's working strategy and its attention for defining/classifying genres. Nevertheless, categorizing musical genres is a complicated task and diverse approaches have been proposed by scientists, engineers, and musicologists. In particular, the field of music information retrieval (MIR) offers diverse approaches to examine genre classification.

In this paper, we propose an alternative genre classification method, using the acoustic characteristics of music. We use Acoustic Brainz[1] for acoustic analysis of the songs. For each song, their genres are taken as the ground truth from the GTZAN dataset. The acoustic analysis has been conducted for these songs in order to obtain their various audio features. Besides acoustic analysis, one of the most important points in the paper while realizing genre classification is deciding on whether a song should be used as full version or in a piece for train and test. In this study, both alternatives are implemented. Considering pieces of songs are a bit more complex than using full versions since it is crucial to decide on which time interval is picked to be more accurate while analyzing. In this paper, it is decided to have pieces of 30 seconds, time interval between 30th and 60th seconds. Selection of songs are in English, based on GTZAN dataset, which has 10 genres that are listed as follows; blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. In addition to the main analysis with English songs, further study is also conducted on a set of Turkish songs.

[1] https://acousticbrainz.org/

Atsiz, E., Albey, E. and Kayış, E.
Effective Training Methods for Automatic Musical Genre Classification.
DOI: 10.5220/0007933202750280
In *Proceedings of the 8th International Conference on Data Science, Technology and Applications (DATA 2019)*, pages 275-280
ISBN: 978-989-758-377-3

2 LITERATURE REVIEW

Identifying musical genres has been given great attention in different academic fields and a very popular topic in MIR studies over the decades (McKay and Fujinaga, 2006). The main and mutual goal of various approaches is to classify music genres in an accurate way as providing an appropriate summary of the music information. These approaches that have been applied based on diverse methodologies are innumerable to list yet a survey of Correa and Rodrigues examine studies from the numerous past years and revealed diverse approaches to a great extent. According to Correa and Rodrigues (2016), classification problems can be handled based on five diverse perspectives from the MIR community. These are audio-content based, symbolic-content based, lyrics-based, community meta-data based and hybrid approaches (Silla and Freitas, 2009)

Audio content-based approach examines digital audio signals to discover music features while symbolic content-based perspective gathers features from different symbolic data forms that present music in a high-level abstraction. These content-based approaches have drawn great attention lately in MIR studies. A review of audio-based classification is revealed by Fu and others. Music classification methods, inclusive discussion about related principal studies, descriptions of them based on music features, schemes of classification and performance accuracies are presented. The review includes five major areas: genre and mood classification, artist identification, instrument recognition and music annotation (Fu et al., 2011).

Lyrics-based approaches create a classification based on lyric information of music as using text mining techniques. Lyric-based classification can be considered as a natural language processing (NLP) problem. In this problem type, the purpose is to assign labels and meaning to texts. That means genre classification of the lyrical text in this case. Traditional approaches have benefitted from n-gram models and algorithms like Support Vector Machines (SVM) or k-Nearest Neighbour (k-NN), and Naive Bayes (NB) (Tsaptsinos, 2017).

Similarly, community meta-data based approaches aim to get online musical information related to songs while benefitting from web-mining techniques. Hybrid-based systems utilize features arising from previous approaches.

In this paper, genre classification is studied according to the content-based approach. The key step of the content-based approach is to overview of audio features. The aim is to predict genres/subgenres of songs while using related music features that automatically computed from audios (Bogdanov et al, 2018). Besides the content-based approach, the dissimilatory point of the paper is giving importance to the usage of songs. Examining specific parts and full versions of songs separately are providing a diverse method to genre classification based on features of songs.

The study of Basili, Serafini, and Stellato (2004) enables a great exemplification. Their work uses different machine learning algorithms to classify music genres into “widely recognized genres” based on trained examples (Basili et al., 2004). Diverse combinations of musical features are used to decide which can bring the most accurate results. In this case, musical features include instruments, instrument classes, meter & time changes and note extension/range. Using these features, different algorithms are implemented to classify genres according to them, accuracies are compared at the end.

Additionally, Tzanetakis and Cook (2002) study this subject with supervised machine learning approaches; his work can be defined as a pioneer study. As using different models such as Gaussian Mixture model and k-nearest neighbor classifiers, they introduced 3 sets of features. These features are categorized as timbral structure, rhythmic content and pitch content.

In light of these approaches, we perform an acoustic analysis for songs and obtain their various audio features. In addition to the approach based on features, songs are used in diverse methods in order to compare results. We focus on how to train and test songs, both full versions and part of songs are utilized in train and test process. In the end, results are compared in order to have the greatest accuracy. This distinctive perspective is to determine the role of using way of the audio files in a genre classification.

3 METHODOLOGY

Our methodology consısts of two parts. Fırst, we conduct an acoustıc analysıs to generate acoustıc features of a gıven song. The second part uses acoustıc features for a set of songs to fınd a model and ıts parameters to predıct genres of these songs. We then use thıs model to predict genre classıfıcatıon probabılıtıes for a second set of songs. For both parts, we consider two dıfferent settıngs; either use the full versıon of the songs or use a specıfıc part of the song (i.e., 30-60 second ınterval, which is the standard in the literature). Thus we study four dıfferent scenarios ın total (see Table 1) and compare the performance of these scenarıos.

Table 1: Scenarios based on usage of songs.

Scenario	Acoustic Analysis	Classification
1	Full Song	Full Song
2	Full Song	Part of a Song
3	Part of a Song	Full Song
4	Part of a Song	Part of a Song

The acoustic analysis in Acoustic Brainz is conducted for both full versions of songs and 30-60 seconds intervals at the beginning. The aim is to create acoustic features of all songs in the dataset with the help of low-level analysis procedure. These features are very various such as loudness, dynamic complexity, spectral energy, dissonance and so on. Then, a model and related parameters are provided by Acoustic Brainz as a result of the analysis. As providing that, Acoustic Brainz runs the SVM model with different parameter combinations and tries to find the best alternative, which completes the training process in the system. The output of this process is a history file that is a source for further predictions.

In order to predict genre classification probabilities for a second set of songs, we use the history file including the best combination of SVM parameters. At this point, the test process starts. Acoustic analysis of songs is turned into a high-level analysis via Acoustic Brainz. The significant quality of high-level analysis is providing the possibilities of belonging to any genres for a related song. To exemplify, possibilities of being blues, rock, pop genres etc. (for all 10 genres) are listed with regarding possibilities for a new song; the summation of possibilities equals to 1. Also, it is significant that the high-level data can be constructed based on the acoustic features of songs without any need to reach to the original audio files. Therefore, after conducting a low-level analysis with a training dataset, training solution is gained as a source for further predictions. The low-level data carries the descriptors for the acoustics which are basically characterizing dynamics, loudness and spectral information of a sound/signal, as well as rhythmical information such as beats per minute and tonal information such as scales. On the contrary, the high-level data has information about moods, vocals, music types, especially genres as in our case. Genres are identified automatically as using acoustic features as base sources with the help of trained classifiers.

Genre classification process starts when a training solution (history file) is gained. Thanks to the history file, new audio files' high-level analysis can be started. The acoustic analysis process is not different from the beginning. Each song has its own acoustic features via Acoustic Brainz. However, the system already contains high-level data. Based on the training, selected best parameters from the SVM model, the genre classification of new songs can be realized. For each song, a special file is created (.yaml files). They include acoustic features and genre classifications of the related song. As a result, the outputs of the prediction process are stored in these files. In order to see the classification results clearly, the files are parsed so that their full data is stored. The stored data is utilized for creating confusion matrixes which are beneficial visualizations that represent all results of prediction phase in a clear format.

4 RESULTS

A dataset of selected 200 songs is split into two groups as 160 and 40 songs. The first group, which has 160 songs, is used for training. 16 songs are selected randomly from each genre. The second group of 40 songs is used for testing. Similarly, 4 songs are selected randomly from every 10 genres. In order to start the analysis, 200 songs of a dataset are created in .mp3 version but using songs as .mp3 without any change leads to misclassification in the system. Therefore, song files need to be converted to .wav formats in order to make the system work with accurate results. The process is applied for both full versions and pieces. In addition to file formats, frequency is a significant factor to be reconsidered in order to have a better analysis and outputs. All frequencies are set as 22050 Hz.

For 160 songs, two different trainings are processed and two diverse tests are studied. In the first approach, full versions of songs are considered. For these audio files, all steps explained in the methodology part of the report are completed. As a result of this work, a history file is gained for these 160 songs' full versions. In addition to this training strategy, breaking songs into pieces is considered as a second option. In this second strategy, time intervals between 30th - 60th seconds for every song are taken. Similarly yet separately, the methodology is applied for these 160 different pieces.

First training approach which is using full versions of 160 songs is used in order to test the other 40 songs as mentioned earlier. Testing is realized for these 40 songs while considering not only their full versions but also 30th - 60th second's intervals of them. The results in confusion matrixes are presented separately as below. The numbers indicate the number of songs. 4 songs for each genre are tested. According to scenario 1, three songs out of four, which are blues, are classified also as blues on the test. In other words, blues genre is tested with 75 % accuracy. Another example is that none of the rock songs are classified as rock genre according to the test results. If the big picture is analyzed, testing full version of 40 songs results in 55,0 % accuracy while Scenario 2, testing 30th - 60th second interval of songs, gives 47,5 % accuracy based on the train data created by using songs' full versions. That means testing full versions with full versions' train data is closer to the reality than the other option.

Table 2: Scenario 1.

	blu	cla	cou	dis	hip	jaz	met	pop	reg	roc
blu	3	0	0	0	0	0	0	0	1	0
cla	0	4	0	0	0	0	0	0	0	0
cou	0	0	3	0	0	0	0	0	0	1
dis	0	0	0	3	0	0	0	0	0	1
hip	0	0	0	0	2	0	1	0	0	1
jaz	1	0	0	0	0	1	0	1	0	1
met	0	0	0	1	0	0	2	0	0	1
pop	0	0	0	0	0	0	0	2	0	2
reg	1	0	0	0	0	0	0	0	2	1
roc	0	0	1	0	0	0	2	1	0	0

Table 3: Scenario 2.

	blu	cla	cou	dis	hip	jaz	met	pop	reg	roc
blu	2	0	0	0	1	0	0	0	0	1
cla	0	3	0	0	0	1	0	0	0	0
cou	0	0	3	0	0	0	0	0	0	1
dis	0	0	0	2	0	0	0	1	0	1
hip	0	0	0	0	1	0	1	0	1	1
jaz	2	0	0	0	0	1	0	0	0	1
met	0	0	0	0	0	0	2	0	1	1
pop	0	0	0	0	0	0	0	1	0	3
reg	1	0	0	0	0	0	0	0	2	1
roc	0	0	1	0	0	0	1	0	0	2

The second training approach is resulted from analyzing 30th-60th second's intervals of 160 songs to test both options. In scenario 3, testing full versions with intervals between 30th - 60th seconds train data gives 32,5 % accuracy. On the contrary, in scenario 4, testing pieces of songs provides 52,5 % accuracy, which is better.

Table 4: Scenario 3.

	blu	cla	cou	dis	hip	jaz	met	pop	reg	roc
blu	0	0	0	1	0	1	0	0	2	0
cla	0	1	0	0	0	0	0	0	3	0
cou	0	0	4	0	0	0	0	0	0	0
dis	0	0	0	4	0	0	0	0	0	0
hip	0	0	0	3	0	0	0	0	1	0
jaz	0	1	1	0	0	0	0	0	2	0
met	0	0	0	3	0	0	1	0	0	0
pop	0	0	0	3	0	0	0	0	0	1
reg	0	0	0	1	0	0	0	0	3	0
roc	0	0	0	3	0	0	0	0	1	0

Table 5: Scenario 4.

	blu	cla	cou	dis	hip	jaz	met	pop	reg	roc
blu	2	0	0	0	0	1	0	0	0	1
cla	0	3	0	0	0	0	0	0	1	0
cou	0	0	3	0	0	0	0	0	1	0
dis	0	0	1	3	0	0	0	0	0	0
hip	1	0	0	0	1	0	1	1	0	0
jaz	2	0	0	0	0	1	0	1	0	0
met	0	0	1	1	0	0	2	0	0	0
pop	0	0	0	0	0	0	0	2	0	2
reg	1	0	0	1	0	0	0	0	2	0
roc	0	0	2	0	0	0	0	0	0	2

In a nutshell, two history files which are created based on different trainings are benefited in order to practice four different test scenarios to predict 40 songs' genres. It can be interpreted from the results that the combination of the train and test data is highly important. In other words, better results are gained when the same versions of songs are used in both train and test data.

4.1 Analysis of the Turkish Songs

Up to this point, only English songs from GTZAN are analyzed and tested. Differently from this, Turkish songs are also studied directly with GTZAN's training data. In the Turkish songs' analysis, two diverse tests are implemented. As it is applied before, full versions and intervals between 30th -60th seconds are tested separately. For Turkish songs, rather than selecting different genres from diverse artists, a specific genre and artist is decided to be concentrated on for testing. For that purpose, one of the most well-known artists, Sezen Aksu, from Turkish music industry is selected as the sole artist. Besides the genre classification approaches explained before, another analysis is performed in this regard. Since Sezen Aksu's songs have left important marks in history, the genre changes within the years are considered as significant indicators in order to have insight about the alterations of the Turkish music industry.

The testing is implemented in the exact same setting with the main part of the study yet the only difference of this part is related to the training data. GTZAN dataset is used directly for training data creation. The library has 1000 songs with specific intervals, which are all used without any change. For the test data, 313 songs of Sezen Aksu, from 1977 to 2017 years, are utilized. In that case, one single training data generated by using direct GTZAN dataset is used for testing two alternatives, which are using full versions of Sezen Aksu songs and intervals between 30th -60th seconds. Alternatively from the previous section, we visualize results in line charts, not with the confusion matrixes. The reason is to be able to see changes throughout the years clearly. Each year, all Sezen Aksu songs of that year are tested and the results for pop and country genres are presented as follows. In the last 40 years, from 1977 to 2017, the highest number of pop songs belong to 1998 when songs are tested as full versions. However, when part of songs are used in testing, the highest numbers are in 1989 and 2005. From these charts, distribution of a specific genre throughout 40 years can be seen. Also, two scenarios can be compared.

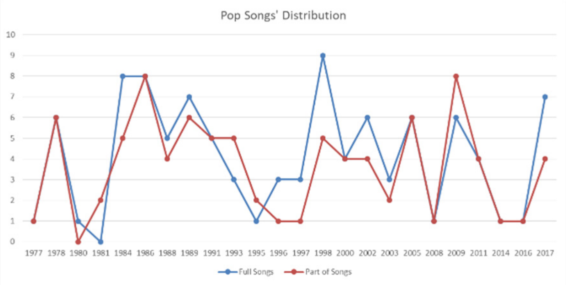

Figure 1: Number of pop songs belong to Sezen Aksu from year 1977 to 2017.

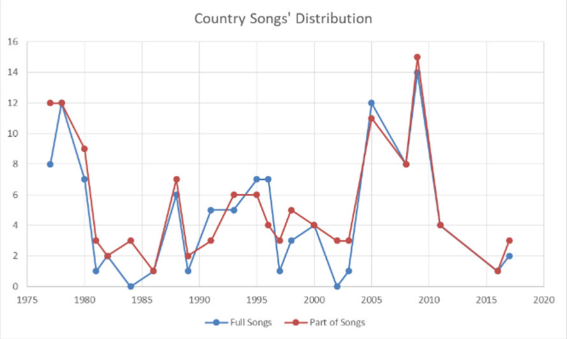

Figure 2: Number of country songs belong to Sezen Aksu from year 1977 to 2017.

5 CONCLUSIONS

Considering the indefinite disposition of musical genres, genre classification can be carried out by studying on technics about specific algorithms as in this paper. Although the area has wide open boundaries, performing genre classification and getting results according to it can bring about much more reasonable determinations rather than identifying genres based on forecasts argued by a human. Therefore, genre classification with the effective usage of Acoustic Brainz is performed with diverse train and test data. Acoustic characteristics of music are considered as significant figures, which basically forms the primary structure of the train data.

In order to perform the analysis, 200 songs are gathered from GTZAN dataset, separated into two groups for train and test process. These songs' genres are given as the ground truth and forecast study is performed based on this acknowledgement. In the first place, the low-level analysis procedure is applied and train data is constructed. Afterward, the high-level analysis is applied for estimating new songs' genres. The key point is that the training data is created as considering acoustic features of the given songs and the best combination of SVM parameters. Once the training data is gained, there is no requirement to implement low-level analysis for the selected songs for GTZAN. In other words, train data holds the qualification of a solid source for further analysis.

The test process is completed based on two different train data which are created with the same ground logic but different approaches. One of the training data is constructed as considering full versions of songs while the other is based on the intervals of 30th - 60th seconds. For both approaches, not only full versions of songs belong to test data is tested but also the pieces. According to these four main tests and related results, it is observed that using the same version of songs for train and test process give better results. In this work, when full versions of songs are tested with the train data that is created with also full versions, 55% accuracy is achieved in classifying genres, which is the greatest among four tests. In addition to all, genre classification is also studied on Turkish songs. A specific artist, Sezen Aksu, is selected for this study. In this regard, genre classification results are indicated. Also, genre changes between years are observed.

Table 6: Accuracy Table.

Scenario	Acoustic Analysis	Classification	Accuracy
1	Full Song	Full Song	55.0 %
2	Full Song	Part of a Song	47.5 %
3	Part of a Song	Full Song	32.5 %
4	Part of a Song	Part of a Song	52.5 %

Finally, in order to save the steps and record the developments of analysis, a web-based dashboard is created by using R Programming. In its construction process, several R packages are utilized but the one that plays the most important role is R-Shiny. This web-based dashboard contains all data used in the analysis, confusion matrixes and results. The significant perks of the dashboard application are its user-friendly screens, beneficial visual expressions. It also gathers all steps and materials of work under a single roof. Therefore, every song used in the analysis is accessible any time and genre classification results related to them can be seen and plotted. Besides, all datasets and results can be downloaded thanks to this dynamic program.

REFERENCES

Basili, R., Serafini, A. & Stellato, A., 2004. Classification of Musical Genre: A Machine Learning Approach. Proceedings of the 2004 International Symposium on Music Information Retrieval at Universitat Pompeu Fabra, 5, 505-8.

Bogdanov D, Porter A, Urbano J, Schreiber H. (2018). *The MediaEval 2018 AcousticBrainz genre task: Content-based music genre recognition from multiple sources.* Paper presented at: MediaEval'18; 2018 Oct 29-31; Sophia Antipolis, France.

Corrêa, D. and Rodrigues, F. (2016). A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, 60, pp.190-210.

Fu, Z., Lu, G., Ting, K. and Zhang, D. (2011). A Survey of Audio-Based Music Classification and Annotation. *IEEE Transactions on Multimedia*, 13, pp.303-319.

McKay, C. & Fujinaga, I. (2006), Musical genre classification: Is it worth pursuing and how can it be improved?, *in* 'ISMIR' , pp. 101-106 .

Silla, C. and Freitas, A. (2009). Novel top-down approaches for hierarchical classification and their application to automatic music genre classification. *2009 IEEE International Conference on Systems, Man and Cybernetics*.

Tsaptsinos, A. (2017). *Lyrics-based music genre classification using a hierarchical attention network.* CoRR, vol. abs/1707.04678.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), pp.293-302