Music Classification Using Fourier Transform and Support Vector Machines

Davis Moswedi[1] and Ritesh Ajoodha[2]

[1] University of The Witwatersrand, Johannesburg, South Africa
1876473@students.wits.ac.za
[2] University of The Witwatersrand, Johannesburg, South Africa,
ritesh.ajoodha@wits.ac.za

Abstract. Information retrieval from music is an active research area in computer science. In this paper, we perform a music classification by genre using the subset of the characteristics of the music signal. Features based on magnitude, pitch, and tempo have been found to be informative for classifying musical pieces by genre. We group the features into these categories. These features are calculated from the Fourier transform's magnitude spectrum. By analyzing the data and exploring it, we develop knowledge about features that can be used for classification, and finally using an information ranking classifier to select the best feature. Finally, Support Vector Machines had the best performance with an accuracy of 81.85% when classifying Spotify music into 20 genres.

Keywords: Information retrieval, music classification by genre, Fourier transform's magnitude spectrum, pitch, tempo, support vector machines

1 Introduction

Increasing music production and its shift to online platforms such as Spotify, Google Play Music, iTunes, and many other streaming platforms has increased interest in music information retrieval as a topic in computer science. The music genres are vague because of the evolution of music over time, but customers to the music streaming platforms prefer to listen to music that is categorized by genre, artist, or album. The successful classification of music will improve customer experience and increase revenue for music streaming services.

As music genre is categorized and labeled manually by human beings, this may lead to different interpretations of music genre from individual to individual. This study aims to investigate large databases of music with many unique music genres. Through thorough analysis of music signals and their properties – such as tempo, valence, loudness, and many others – we can demonstrate that some genres belong together rather than as individual categories. Then, music genres can be regrouped together, allowing machine learning models to be trained to provide an accurate automatic classification of music genres.

2 Related Work

Automatic music genre classification is important for organizing, searching, retrieving, and recommending music [8]. There are mainly three steps in music classification: feature extraction, feature selection and fitting machine learning classifiers. Magnitude, tempo, pitch, and chordal features are the most important musical features to classify music by genre [7] [1]. A magnitude-based feature is a timbral marker of music signals like loudness, compactness and pitch. A music signal's rythm can be represented and analyzed using tempo-based features. Pitch-based measures define properties of sound wave signals based on frequency, allowing us to classify musical signals as high or low. Features based on chords examine Chroma, a chord-recognizing feature in music signals [7] [5]. A fast Fourier transform is used to extract these features, which are represented by their means and standard deviations to reduce their natural high dimensions. A widely used music database for classification by genre is the GTZAN which was collected by [7]. The data set consists of 1000 songs that are categorized into 10 genres, with 100 songs in each genre. Table 1 lists some notable music classification algorithms based on this dataset.

Authors	**Accuracy**
Nkambule and Ajoodha	**80.8**%
Ajoodha, Klein, and Rosman	**81**%
Tzanetakis and Cook	**61**%
Lee et al.	**90.6**%
Panagakis, Kotropoulos, and Arce	**91**%
Lau and Ajoodha	**66**%

Table 1. Notable music classification on GTZAN dataset

For accurate music genre classifiers to be implemented, a reliable ground truth is essential. Ambiguity in the ground truth results from: unclear definitions of music genres, inconsistencies in various music sources, similarity between music genres, etc [4]. For this study, we propose the use of Spotify's database, which contains 149 different music genres. In total, 1995 songs are represented in the database; they are not equally distributed across 149 different music genres. To avoid the inconsistencies outlined above, we use hierarchical clustering algorithms to group these music genres. The dendrogram is especially useful, since we can see how the clusters are arranged hierarchically and make an informed decision about the number of clusters we want to use. The dendrogram is more useful when the height reflect the distance between the clusters and the ultra-metric tree inequality holds. In this case, we can use the distance to find reasonably number of clusters.

3 Music Features

There are three main features that are used in our study: magnitude-based, tempo-based, pitch-based, and speech-based. As well as the previously mentioned features, there are also additional features, some of which are derived from them. These features are represented by their means. The mean is calculated using the following equation:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} f_i$$

3.1 Magnitude-based features

A fast Fourier transform of a music signal produces a magnitude spectrum that contains a variety of spectral features that can be used to classify a music by genre [1]. Our study incorporates magnitude-based features with spectral features, such as the following:

1. Loudness: loudness is how the auditory system perceive the pressure of music signals as sound waves, which range from quiet to loud. The use of loudness arise in many fields such as multi-speaker speech, speech recognition and music genre classification.

3.2 Tempo-based Features

In music classification by genre is helpful to study tempo features since most music holds normal musical arrangements that creates an impression of rhythm [5].

1. Energy: is the measure of discrete-time signal of sound by taking root mean square energy (RMS) of the signal.
2. Tempo: is the speed of the music signal, measured in beats per minute (BPM).

3.3 pitch-based and speech-based features

The pitch of a note is a perceived characteristic included in the frequency of the music. Most music of the same genre has melodies that are a combination of notes from a scale set. It is hard for even humans to distinguish between pitches.

1. Zero Crossing Rate (ZCR): is the frequency of sign changes that happen along a discrete-time signal.
2. Speechiness: In this feature, the attribute detects the presence of spoken words in a track, and the closer to 1.0 the attribute value is, the more speech-like the recording is (e.g. talk show, audio book, poetry).

3.4 Additional Features

1. Danceability: Several song features are used to measure the danceability of a song, including beat strength, tempo stability, and overall tempo. This method generates a score that indicates whether a person is able to dance to the entire song comfortably.
2. Popularity: Song popularity is a measure of how popular a song is and this data is provided by music streaming platforms.

4 Feature Selection

Information gain ranking is a filter algorithm that measures the information gain relative to the dependent variable (class) to evaluate the value of a feature. As you can see, the upper part of Table 2 shows the features that were used, while the lower part shows the features that were discarded. The value of the features was discovered through the Information gain rank algorithm. In Figure 1, The data for liveness and speechness were heavily skewed and dominated by outliers. With the outliers removed and the information rank algorithm applied, both of the features became less valuable or irrelevant as a predictor of music genres.

Features maintained	**Dimension**	**Mutual information**
Acousticness	1	0.797
Beats Per Minute (BPM)	1	0.689
Danceability	1	0.299
Energy	1	0.532
Loudness (dB)	1	0.227
Popularity	1	0.121
Valence	1	0.645
Features discarded	**Dimension**	**Mutual information**
Liveness	1	0.049
Speechness	1	0.060

Table 2. Features maintained are show in the upper region while Features discarded are in the lower region. Column two and three list the feature dimension and feature mutual information respectively.

5 Automatic Music classification by genre

In this section we perform automatic music classification by genre using six off-the-shelf classifiers: Naive Bayes, Logistic Regression, k-nearest neighbours, Random Forest, Multilayer Perceptron, and Support vector machines. Table 3 give accuracy and time taken to build each classifier. The classifiers were build using stratified k-fold (i.e. k=10) cross-validation. It can be seen that Support

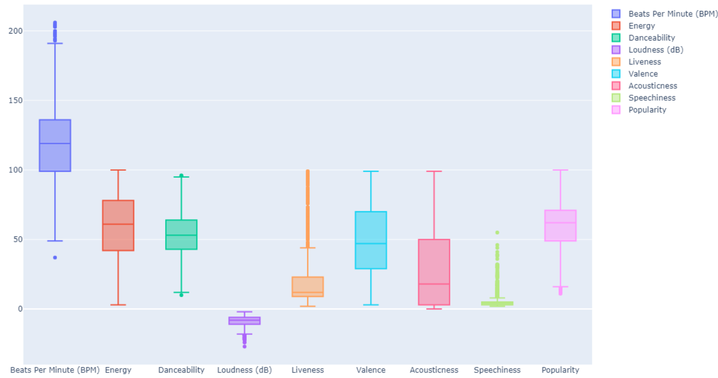

Fig. 1. Data distribution based on a five number summary per feature. The data are normally distributed in most features, except for liveness and speechness, in which there are mostly outliers.

Vector Machines has best accuracy. Figure 2 shows the confusion matrix for 20 generalised Spotify Genres using Support vector machines with stratified 10-fold cross-validation.

Classifier	**Accuracy**	**Time to build (s)**
Naive Bayes	77.01%	0.06
Logistic Regression	77.08%	1.41
K-nearest neighbours	77.56%	0.08
Random Forest	80.22%	2.99
Multilayer Perceptron	80.40%	11.27
Support Vector Machines	81.85%	0.46

Table 3. Model performance

Classifier	**Parameters used**
Naive Bayes	Gaussian Naive Bayes with smoothing
Logistic Regression	solver=newton-cg, penalty=l2, max iterations=100
K-nearest neighbours	k=5, distance metric=minkowski, weighting=distance
Random Forest	number of trees=100, split function=gini
Multilayer Perceptron	hidden layers=1, activation function=relu, solver=lbfgs, max iterations=200
Support Vector Machines	kernel=polynomial, kernel degree=4, coef0=1.0

Table 4. Model implementation details

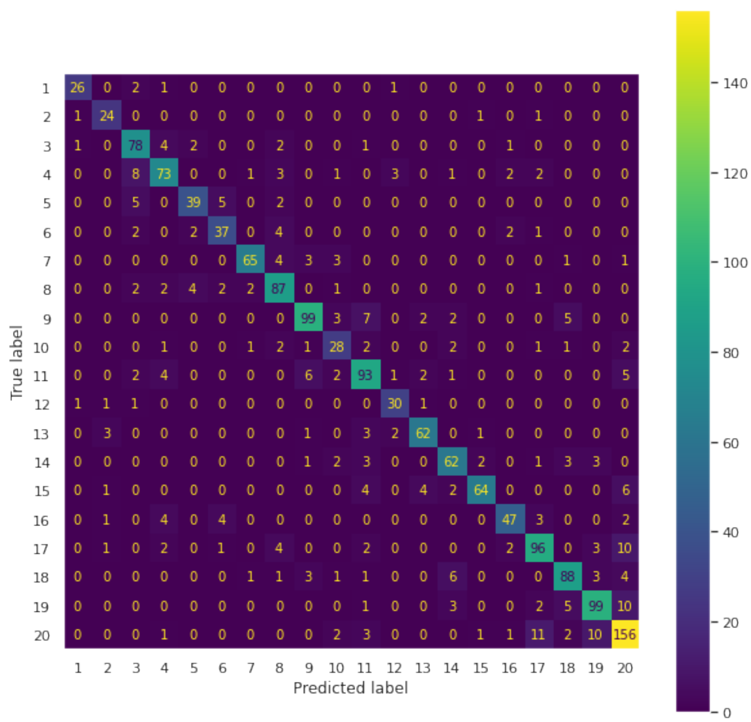

Fig. 2. The confusion matrix for 20 generalised Spotify Genres using Support vector machines with stratified 10-fold cross-validation

6 Conclusion and Recommendations

In this study we performed automatic music classification by genre, it is seen that the models performed nearly the same as the best classification models in the literature. The use of content-based features and small changes to the existing music genre classification models will unlikely result in a significant better classification of music by genre [1]. This statement poses a limitation in improving the accuracy using content-based features hence tackling this problem in a different way may increase the accuracy of the models. The same strategy introduced in this study of grouping genres related to one another can be used in the future and also incorporating cultural features with structural ones. In existing music databases, such as GTZAN, there are fewer music genres, which is not indicative of the real nature of music. The use of databases with a large number of music genres allows researchers to redefine music genres through hierarchical clustering algorithms. The redefinition of music genres will eliminate incorrectly named genres and genres that were categorized as different genres due to regions (i.e. countries) will be reclassified as one. This method could result in better classification accuracy than the current one.

References

[1] Ritesh Ajoodha, Richard Klein, and Benjamin Rosman. "Single-labelled music genre classification using content-based features". In: *2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*. IEEE. 2015, pp. 66–71.

[2] Dhevan Lau and Ritesh Ajoodha. "Music Genre Classification: A Comparative Study Between Deep-Learning And Traditional Machine Learning Approaches". In: (2020).

[3] Chang-Hsing Lee et al. "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features". In: *IEEE Transactions on Multimedia* 11.4 (2009), pp. 670–682.

[4] Cory McKay and Ichiro Fujinaga. "Musical genre classification: Is it worth pursuing and how can it be improved?" In: *ISMIR*. Citeseer. 2006, pp. 101–106.

[5] Tshepo Nkambule and Ritesh Ajoodha. "Classification of Music by Genre using Probabilistic Models and Deep Learning Models". In: (2021).

[6] Yannis Panagakis, Constantine Kotropoulos, and Gonzalo R. Arce. "Music genre classification via sparse representations of auditory temporal modulations". In: *2009 17th European Signal Processing Conference*. 2009, pp. 1–5.

[7] George Tzanetakis and Perry Cook. "Musical genre classification of audio signals". In: *IEEE Transactions on speech and audio processing* 10.5 (2002), pp. 293–302.

[8] S Vishnupriya and K. Meenakshi. “Automatic Music Genre Classification using Convolution Neural Network”. In: *2018 International Conference on Computer Communication and Informatics (ICCCI)*. 2018, pp. 1–4. DOI: 10.1109/ICCCI.2018.8441340.