Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2021, 8(4):150-156

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

A Music Genre Classification and Feature Discovery Method based on XGBoost

Yikang Kang

School of Information, Renmin University of China, Beijing, China
Email: yikang_2000@126.com

Abstract We proposed a method to measure music similarity, and proved that songs within the same genre are more similar than songs between genres by calculating Between-genre similarity and Within-genre similarity. Then we proposed a method that combines the results of multiple binary classifiers based on XGBoost to tackle the problem of music genre classification. By comparing the results with five other machine learning algorithms, we proved that the method based on XGBoost is the best-performing one in terms of prediction range and accuracy. Finally, we further proposed XGBoost feature discovery algorithm to find the most important features that distinguish a genre from other genres.

Keywords Music Genre Classification; Feature Discovery; XGBoost Algorithm

1. Introduction

Music comes in many different types and styles ranging from Classic music to Reggae and Vocal. Nowadays, companies may use music classification for many purposes, either to be able to place recommendations to their customers or simply as a product. Determining music genres is the first step in that direction.

However, music genre classification has always been a challenging task. Sounds never evoke the same emotions in different people. Music genres are hard to systematically and consistently describe due to their inherent subjective nature. Thus, merely defining a music genre based on the subjective feelings (e.g. cheerful, sorrowful…) it carries to people is not effective and convincing. Machine Learning techniques have proved to be quite useful in extracting trends and patterns from a large pool of data. Therefore, analyzing the songs based on their digit signatures for some factors such as tempo and energy, along with the technique of machine learning can tackle the problem of music classification.

In this article, we shall study how to develop a method for music genre classification and then explore the most important features that distinguish a music genre from other genres.

Related Research

Focusing on the characteristic segment method and classification method, Hu JK [1] proposed the music segmentation method based on the feature parameters of discrete cosine transform domain and the music classification method based on Lcs. Juliano [2] evaluated the impact of frame selection on automatic music genre classification and also presented a novel texture selector based on K-Means aimed to identify diverse sound textures within each track. Du [3] paid more attention to the music features combining with statistical features and proposed a hierarchical structure for the music classification. Yue [4] proposed a sparse feature extraction method based on sparse decomposition and multiple instrument component dictionaries, which could get sparse features that would be used independently & with high interpretability through in-depth analysis of the sparse coefficient vector. Jiang [5] studied the association between the timbre perception feature of music and the texture of images. He also established the matching model of timbre and texture. JIANG [6] introduced the experimental steps and the analysis process for getting the timbre perceptual eigenvalues of each timbre material.

The above-mentioned works have some contributions to the problem of music classification, but their music genre classification methods have some limitations which harm the classification precision.

Our contributions are as follows:

- We proposed a method to measure music similarity and proved that songs within the same genre are more similar than songs between genres by calculating between-genre similarity and within-genre similarity, indicating the feasibility of classifying a song into its music genre based on its features.
- We divided a multi-class classification problem into multiple binary classification problems and combined the results to improve accuracy.
- For each song, we get a bucket of predicted genres instead of a single predicted genre to better suit the problem of music classification and improve accuracy and reliability.
- We used the XGBoost Algorithm to classify the music genres. By comparing the results with five other machine learning algorithms, we proved that XGBoost is the best-performing algorithm in terms of prediction range and the accuracy of the prediction.
- In order to further find the most important features that distinguish a genre from the other genres, we further applied the XGBoost Algorithm to do feature discovery.

2. Data and Data Preprocessing

Our work is based on the datasets provided in problem D of the 2021 Interdisciplinary Contest in Modeling (ICM). The dataset provided contains 98340 songs across 21 genres. The music genre and various song features of each song are provided in detail.

First, in order to base on enough samples, we filtered songs that belong to a music genre that includes less than 1000 song records in the dataset. After the filtering, 93632 songs belonging to 13 music genres (Blues, Classical, Country, Electronic, Folk, International, Jazz, Latin, Pop/Rock, R&B, Reggae, Stage& Screen, and Vocal) are left, as shown in table 1.

Table 1: The number of songs in each music genre

Music Genre	**num of songs**
Blues	1256
Classical	2352
Country	7608
Electronic	1332
Folk	1397
International	1460
Jazz	6942
Latin	4674
Pop/Rock	47513
R&B	10340
Reggae	1421
Stage & Screen	1058
Vocal	6279
sum	**93632**

Data Description

We chose 14 features, which can be divided into three categories: characteristics of the music, type of vocals and other info. The definitions and descriptions of the 14 features are listed below:

- Characteristics of the music:

- dance-ability: A measure of the music's suitability for dancing.

- energy: A measure representing the perception of intensity and activity.
- valence: A measure indicating the musical positiveness conveyed by a track.
- tempo: The overall estimated tempo of a track in beats per minute (BPM).
- loudness: The average loudness of a track in decibels (dB).
- mode: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track.
- key: The estimated overall key of the track.

- Type of vocals:

- acousticness: A measure of whether the track is without technology enhancements.
- instrumentalness: A measure of whether a track contains no vocals.
- liveness: Detects the probability that the track was performed live.
- speechiness: Detects the presence of spoken words in a track.
- explicit: Detects explicit lyrics in a track

- Other info:

- duration_ms: The duration of the track in milliseconds.
- popularity: The popularity of the track.
- year: The year of release of a track.

3. Music Similarity

When comparing different music, there are numerous dimensions of similarity, thus the notion of music similarity is complicated. In this part, we explore whether songs belonging to the same genre are more similar in terms of the 14 features compared to songs belonging to different genres. If the conclusion is true, it makes sense to use the 14 features to decide the music genre of a song.

In order to solve the issue, we first give the following definitions:

- Song similarity: the similarity of two songs
- Between-genre similarity: the average song similarity of all pairwise combinations of songs from two certain genres
- Within-genre similarity: the average song similarity of all pairwise combinations of songs from a certain genre

3.1 Measures of Song Similarity

To tackle the problem of measuring music similarity, we use the Pearson correlation coefficient as a measurement of song similarity.

Pearson correlation coefficient:

Given two variables X (the 14 features of one song) and Y (the 14 features of the other song), Pearson correlation coefficient (PCC) is the covariance of two variables, divided by the product of their standard deviations.

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

In which cov is the covariance, σ_X is the standard deviation of X, σ_Y is the standard deviation of Y, μ_X is the mean of X, μ_Y is the mean of Y, E is the expectation.

3.2 Between-genre similarity and within-genre similarity

We calculated between-genre similarities and within-genre similarities among the 13 music genres. The results are depicted in a 13*13 genre similarity matrix M (for there are 13 music genres in total), where M[i][j] indicates the genre similarity (if i=j: M[i][j] means Within-genre similarity; else: M[i][j] means Between-genre similarity).

The results show that although there are few exceptions, Within-genre similarities (figures on the diagonal line) are generally larger compared with between-genre similarities, proving that songs within the genre are more similar than songs between genres.

The above study of similarity shows the feasibility of classifying a song into its music genre based on its 14 features, laying a solid foundation for further classifying music genres based on the 14 features of songs.

4. Music Genre Classification and Feature Discovery Model

4.1 XGBoost Algorithm

We assume that X is the song features set (including the 14 features), Y is the music genre that the song belongs to. The given training data set given is $S = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\}$, in which $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots\ldots, x_i^{(p)})$ is an input sample, p is the number of features (14 in this case), i= 1,2, …, n, n is the number of samples.

After the original training samples are processed by feature engineering techniques such as vacant value processing, outlier processing, and data standardization, the training data set is input into the XGBoost model.

XGBoost（eXtreme Gradient Boosting）is an improvement of gradient boosting algorithm and an optimized machine learning algorithm, which can quickly and accurately solve many data science problems. Newton’s method is used to expand the loss function to the second-order when solving the extreme value of the loss function. Besides, in order to balance the decline of the objective function and model complexity, the regularization term is added to the loss function. According to the XGBoost algorithm, the objective function of the music genre classification model can be defined as [7]:

$$\mathrm{obj}^{(j)} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$$

In which y_i represents the actual music genre labels, $\hat{y}_i$ represents the predicted label by the model, $\sum_i l(y_i, \hat{y}_i)$ indicates the loss function, and $\sum_k \Omega(f_k)$ reflects the complexity of the model:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda||\omega||^2$$

In which T represents the number of the nodes of a tree, while $||\omega||^2$ is L2 regularization.

The key idea of the music genre classification model based on XGBoost is to obtain the optimal value of the objective function by optimizing the objective function step by step.

First, the trees are added continuously, and a tree is generated by feature splitting. A new function is used to fit the residual of the last prediction and obtain the optimal value of the objective function;

Second, when the training is over, K trees are obtained in total. When predicting the score of a sample, the features of the sample correspond to leaf nodes in each tree, and each leaf node has a score;

Finally, the score of each tree is added to get the predicted value of the sample.

4.2 Music genre classification method based on XGBoost

In order to tackle the problem of music genre classification, we proposed a method based on XGBoost.

There are 13 music genres in total (as shown in table 1). If we use the 14 features of each song to directly predict which genre it belongs to, the precision rate of this 13-class classification problem is relatively low. Through our experiments, the precision rates of the 13 music genres range from 15% to 60%. In order to provide more reliable and accurate genre predictions, we supplemented the single multi-classification model with 13 separate binary XGBoost classifiers $C_i (i = 1,2, \ldots, 13)$. C_i is used to classify whether a song belongs to the particular genre i or the 12 other genres. If a song belongs to music genre i, then the label of the song would be assigned 1. If a song does not belong to music genre i, then 0 would be assigned to the label of the song.

Since the sample size varies across different music genres, we apply the random under-sampling technique to adjust the ratio between the different genres represented. Specifically, we randomly remove samples with label 0 until there are equal amounts of samples with label 0 and label 1.

Next, we divide the adjusted data set into a training dataset, a validation dataset, and a testing dataset. We train the model on the training dataset, adjust the model parameters using the validation dataset and then make predictions on the testing dataset to compare with their true labels.

After 13 binary classifiers are trained, we further explore the classification results. We used the stratified sampling technique based on the total number of song records of each music genre to select 29000 final testing samples. We input the 14 features of each sample into each trained binary classifier. Each classifier would then predict whether the sample belongs to its music genre or not and output 1 for belonging and 0 for not belonging. As a result, this process produces a bucket for each sample containing all the corresponding genres of binary classifiers that output 1.

Then, for each sample, we summarize the number of music genres in the prediction bucket (criterion 1), and whether the bucket contains the genre that the sample actually belongs to (criterion 2). These two criteria indicate the prediction genre range and the accuracy of the prediction respectively.

Table 2: The partial results of binary XGBoost classifiers

Song Identifier	Real Genre Identifier	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12	C 13	Criterion 1	Criterion 2
23206	7	1	1	0	1	0	0	**1**	0	0	0	0	0	0	4	1
36685	9	1	0	1	0	0	0	0	0	**1**	1	0	0	0	4	1
42900	9	1	0	1	0	0	0	0	0	**1**	0	0	0	0	3	1
44317	9	0	0	0	0	0	0	0	0	**1**	0	0	0	0	1	1
57490	9	0	0	1	0	0	0	0	0	**1**	1	0	0	0	3	1
24880	8	0	0	0	0	0	0	0	**1**	0	1	0	0	1	3	1
89193	12	1	0	0	1	0	1	1	0	0	0	1	**1**	0	6	1
76648	10	0	0	1	0	0	0	1	0	0	**1**	0	1	1	5	1
53339	9	0	0	0	0	0	1	0	0	**0**	1	1	0	0	3	0
67943	9	0	0	0	0	0	0	0	0	**1**	0	0	0	0	1	1

Note: *In table 2, criterion 1 is the number of music genres in the prediction bucket. Criterion 2 indicates whether the prediction bucket contains the genre that the sample really belongs to.*

Finally, we calculated the average of criteria 1 (indicating prediction range) and the average of criteria 2 (indicating prediction accuracy) of all final testing samples.

The above processes are repeated with the other five machine learning models (Random Forest, Knn, Decision Tree, SVM, and Adaboost). The results are shown in table 3. As can be seen from the table, with binary XGBoost classifiers, the average number of music genres in the bucket is 2.87, which is the smallest among all 6 machine learning classifiers. Moreover, using binary XGBoost classifiers, 89.39% of buckets include the real genre that the final testing samples belong to, which is the highest among six machine learning models. The results show that compared with the other five machine learning models, our proposed music genre classification method based on XGBoost could not only provide the smallest prediction range of music genres, but also the most accurate predictions. Therefore, the music genre classification method based on XGBoost has the best classification ability.

Table 3: The performance of the music genre classification method based on XGBoost and 5 other models

	XGBoost	RF	Knn	Decision Tree	SVM	Adaboost
Avg. criteria 1	**2.87**	3.22	5.71	3.63	5.54	3.03
bucket precision rate	**89.39%**	85.53%	68.73%	86.21%	60.66%	87.60%

4.3 Feature discovery method based on XGBoost

In 4.2, we solved the problem of music genre classification. Now in this section, we further proposed a feature discovery method based on XGBoost to find the most important features that distinguish a genre from the other genres.

Every music genre has a specific set of distinguishing features, which is a unique pattern or combination of characteristics as tempo, energy, and valance. We use the binary XGBoost classifiers mentioned in 4.2 to rank the importance of features in the classification process. To be more specific, after inputting the training set data into the binary XGBoost classifiers, we can get the set T including K trees f, ($f_i \in T, i = 1,2, \dots \dots, K$). To get the importance of different features in the classification process, we count the total number of occurrences each feature appears in all trees' split nodes. Finally, the top N most important features are selected. The feature discovery algorithm based on XGBoost is as follows:

Feature discovery algorithm based on XGBoost

XGBoost_FD algorithm

Input: The dataset containing M features
Output: A set containing N features
1. based on XGBoost, the set T including K trees is obtained
2. define list [M]
3. for every f in T:
4. for every node in f:
5. add 1 to the corresponding position of the list according to the corresponding features of the node
6. sort list
7. return the N biggest features in list

We used the XGBoost Feature discovery algorithm to find the most distinguishing features for every single music genre. Take the binary XGBoost classifier for differentiating International music as an example. The feature importance scores are shown in Figure 1. As can seen from the figure, duration_ms is the most important feature that distinguishes International music from other categories, followed by acousticness and year, while explicit, mode and key have the least contribution.

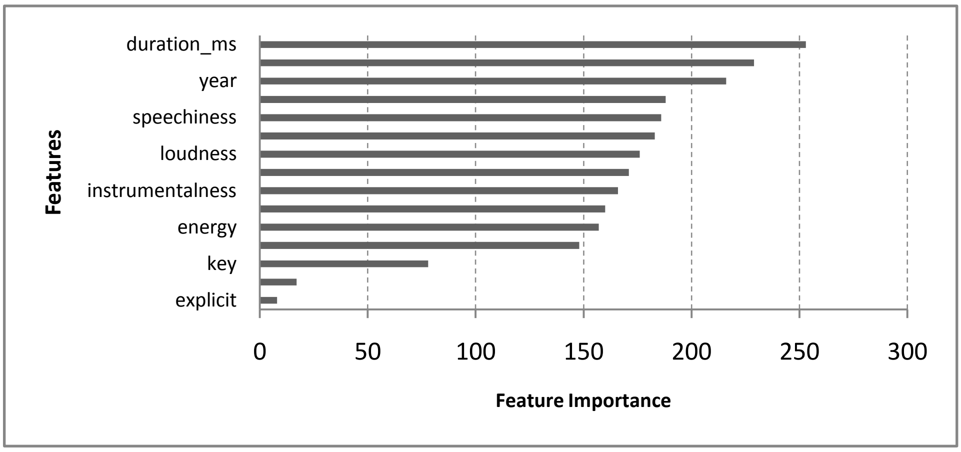

Figure 1: Feature importance of binary XGBoost classifier for classifying International music

We summarized the five most important features of every binary XGBoost classifier and counted how many times each feature occurs in the top five feature importance ranking. The result is shown in Figure 2, indicating that duration_ms, year and dancebility play the most paramount role in generally distinguishing a music genre, followed by acousticness and speechiness.

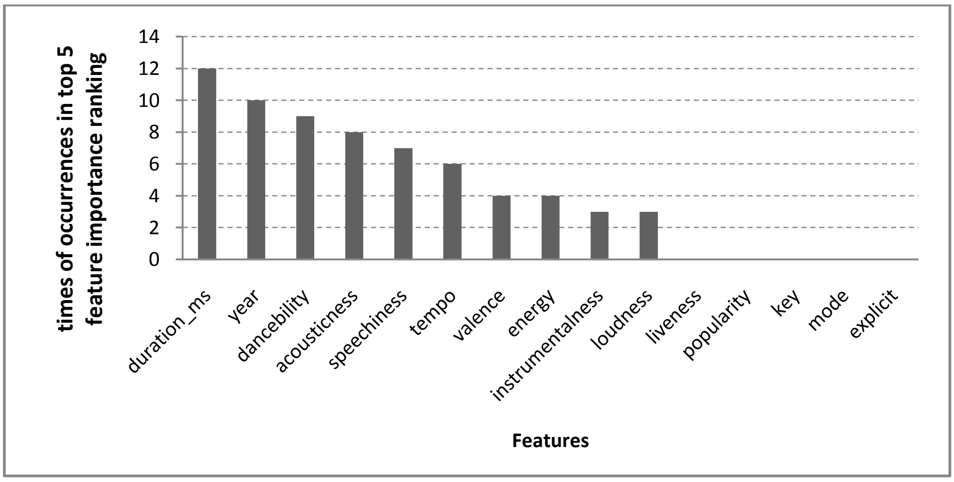

Figure 2: The number of occurrences each feature appears in the top five feature importance ranking

5. Summary

The study of music similarity shows that songs within the same genre are more similar than songs between genres, laying a solid foundation for further classifying music genres based on the 14 features of songs. In order to tackle the problem of music genre classification, we proposed a method based on XGBoost. By comparing the performance of our model with 5 other machine learning models, we concluded that XGBoost outperforms all other models in terms of both prediction range and prediction accuracy. We further used the XGBoost feature discovery algorithm to find the most distinguishing features for music genres. As music genre classification is acquiring its value in an increasing number of scenarios, we will further improve our work and explore the real applications of our work.

References

1. Hu JK, Wu L, Gao Y. MP3 music classification method based on LCS. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2007(04):417-421.
2. Juliano Henrique Foleis, Tiago Fernandes Tavares. Texture selection for automatic music genre classification [J]. Applied Soft Computing Journal,2020(89): 106127
3. DU Wi, LIN H, SUN JWw, et al. A New Music Genre Classification Method based on Hierarchical Support Vector Machines [J]. Journal of Chinese Mini-Micro Computer Systems, 2018, 39(05):888-892.
4. YUE Qi, XU Zhongliang, GUO Jifeng. Sparse Feature Extraction Method for Mixed Instruments Music Analysis [J/OL]. Computer Engineering and Applications: 1-7 [2021-03-20].
5. Jiang W, Zhang XY, Liu JY, et al. Research on the Associations between Timbre Perception Feature of Musical Sound and Texture Features of Image [J]. Journal of Fudan University (Natural Science), 2020, 59(03): 314-321.
6. Jiang W, Liu J Y, Zhang X Y, et al. Analysis and modeling of timbre perception characteristics of Chinese musical instruments [J]. Applied Sciences, 2020, 10(3):789.
7. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System [C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco, California. New York: ACM, 2016:785-794.