

Influence Analysis in Music: Artists and Genres

Naizheng Jia¹, Dian Zuo², Baojun Jia³

^{1,2}College Information and Science, Northeastern University, Shenyang, China

³Fushun Petrochemical Corporation, China

¹jianaizheng@stumail.neu.edu.cn, ²2811303969@qq.com, ³jiabj@petrochina.com.cn

Abstract: Music is an important part of human civilization, which is accompanied by the emergence and development of human civilization. While music is developing, different genres of music or musicians' styles are all influencing each other. Therefore, it is significant to understand and measure the influence of music influencers and their followers in different genres of music with social development. Address to identify the influence network of American music, firstly, a PageRank model was proposed to calculate the music influence of every artist among the network. According to the results, we built the subnetworks to illustrate the influence of different sets. Then, we did research on distinguishing between genres, which utilized the random forest classifier. Furthermore, Gini impurity is used to demonstrate the feature's importance. Our result shows that Pop is the best influential genre, and acousticness is the most distinguishing feature.

Keywords: Graph network, PageRank, Random forest.

1. Introduction

With the continuous development of human society, music has been an important topic from the beginning. It has a long history, with many works of literature spread, directly or indirectly affected the change of society and civilization, is an important part of the history of human spiritual civilization. People have never stopped working hard on how to classify and understand music. When watching the history of music, one can see how the different graphic shapes have slowly developed with a strict reliance both on innovation headways and changes in music hones[1]. Nowadays, music is made up of multiple genres, labelled with more convenient identification, and recommended to people with specific musical tastes using deep learning algorithms in music software. Throughout the field of music research, music genres, evolution and influence have always been indispensable elements for analyzing music. This can be due to a sequence of small changes, a cooperative effort of artists, a series of influential artists, or a shift within society.

Recently, there will be mutual influences between different music genres, making music diversified. (i.e., PJ Harvey in Pop was influenced by Tricky in Electronic.) Therefore, analyzing the source of music development and the factors that promote development will have a certain guiding role in the improvement of social science and in the future. The combination of information technology and traditional field on the Internet has even changed the music industry's development trend greatly[1-3].

With the popularity of big data analysis in recent years, we can also use better mathematical models to analyze the influence of music and its development more accurately. Computer music production software is used to process music compositions to solve the current problems of low efficiency in music processing, audio adaptation, digital recording and other aspects of the traditional composition mode[4]. used transfer learning to predict the genres of the music[5]. utilizes CE-GAN to make AI music composition[1]. highlights how multimodal algorithms can help Music Computing applications.

As for the music classification, there are lots of mathematics models and AI algorithms applied in it[6]. reveals Cifras can be used in Latin Music Mood[7]. proposes an active transfer music genre classification method for musical genre classification.

However, nowadays, it is still difficult to identify the influence of a musician or a genre among the music field, and the difference between the genre is so confused that can not be distinguished clearly intuitively because the growing complexity of genres and the rapped combination of multiple genres continuously challenge the accuracy of classical algorithms. For solving the problems, we propose Page Rank calculate the influence of musicians and the genres. Secondly, Random Forest is used to making classification of different genres and get the importance of the genres feature. It can be found that Pop\Rock and Billie Holiday are the best influential genre and individual musicians in music history, respectively. For the difference of the genres, the feature of 'acousticness' obtains the highest score of importance. Our result can be useful for the singer and musician to develop their music career and make more attribute in music education.

2. Data Description

Our data is downed from the American dataset, and the dataset represents musical influencers and followers These data contain influencers and followers for 5,854 artists in the last 90 years. The influence dataset and its attributes are shown in Table 1.

Table 1: The attributes of the dataset

Attribute	Description
influencer id	Musician ID in the dataset
influencer name	The name of the influencer
influencer main genre	Main genres of influencers among musicians
influencer active start	Date the influencer started his music career
follower id	Musician ID in the dataset
follower name	The name of the follower
follower main genre	Main genres of followers among musicians
follower active start	Date the follower started his music career

As Table 1 shows, according to the existing data, we can build a directed network, where an artist represents a node, and the

relationship between an artist and another artist is represented as a directed edge, from follower to influencer. The influencers and the followers can be connected to a graph, as Figure 1 shown, A is influenced by B who is the influencer of A. And the relationship is the connecting line. Traditionally, the size of the node is the number of lines. However, it is not useful and interpretable for the node. In our influence datasets, musicians will be influenced by others, and they will also be influenced by others at the same time, which is very similar to the aforementioned web recommendation systems. Therefore, we propose the Page Rank model, which can quantify the influence of musicians in history.

The second dataset is the artist music component dataset, which includes 14 features, 3 identity variables (musician id, musician name, and musician genres). The full dataset description is shown in Table 2.

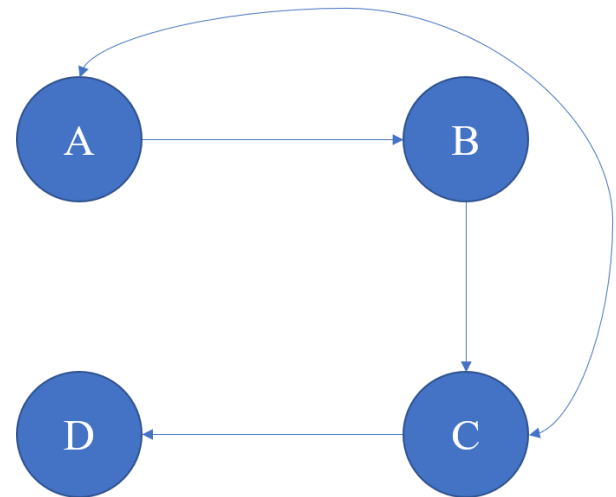


Figure 1: The influence internet graph

Table 2: Dataset description

Feature	number	mean	std	min	25%	50%	75%	max
danceability	5602	0.55	0.13	0.0804	0.5	0.5	0.64	0.96
energy	5602	0.59	0.21	0.00198	0.4	0.6	0.75	1.00
valence	5602	0.55	0.20	0.02785	0.4	0.6	0.69	0.98
tempo	5602	120.48	19.08	30.946	109.6	119.3	129.84	206.68
loudness	5602	-9.91	4.12	-40.147	-12.2	-9.5	-6.85	1.34
mode	5602	0.82	0.38	0	1.0	1.0	1.00	1.00
key	5602	5.53	3.52	0	2.0	6.0	9.00	11.00
acousticness	5602	0.34	0.29	1.01E-06	0.1	0.3	0.57	1.00
instrumentalness	5602	0.13	0.23	0	0.0	0.0	0.14	0.97
liveness	5602	0.20	0.12	0.0116	0.1	0.2	0.23	0.97
speechiness	5602	0.07	0.08	0.0232	0.0	0.0	0.07	0.96
Duration ms	5602	246150.34	82703.52	45707	197507.0	234984.5	275699.73	1640000.00
popularity	5602	39.67	13.12	0	32.4	40.3	48.50	81.00
count	5602	34.69	77.62	1	4.0	10.0	32.00	1369.00

3. Music Influence Network Based on PageRank Algorithm

3.1 Page Rank

The decision of the ranking of the web page is very important in web, as its growing and changing very rapidly. Ranking of the results in a search engine for a query plays a crucial role for a huge database like the Web, where one query can have millions of results[8]. Inspired by Google's early page ranking algorithm, here the same PageRank algorithm can be used to calculate the influence of musicians. Page Rank is the link analysis algorithm proposed by Google founders Larry

Page and Sergey Brin in 1997 when they built the early search system prototype. Since Google has achieved unprecedented commercial success, this algorithm has become a computational model that is of great concern to search engines and academia. Page Rank shows that the result of a node's influence on the system is that the nodes connected to it also have a certain influence.

In our model, an artist's musical influence is mainly measured by two aspects.

i).The greater the entry-degree of an artist, the greater its musical influence.

ii).The higher the quality of the entry source of an artist node, the greater its musical influence.

Firstly, if musician T is influenced by a musician T_n , we define the Page Rank score of musician T as:

$$PR(T_i) = \frac{PR(T_1)}{L(T_1)} + \frac{PR(T_2)}{L(T_2)} + \dots + \frac{PR(T_N)}{L(T_N)}, \quad (1)$$

where $L(T_n)$ is the musician linking to a musician T_i . However, since there are some nodes whose affected by no musician, in other words, those nodes that do not link to any other musicians. PageRank formula needs to introduce the damping coefficient q to be revised, which is shown as:

$$PR(T_i) = q\left(\frac{PR(T_1)}{L(T_1)} + \frac{PR(T_2)}{L(T_2)} + \dots + \frac{PR(T_N)}{L(T_N)}\right) + \frac{1-q}{N}, \quad (2)$$

where N is the number of musicians in this net.

After that, the Page Rank vector R is:

$$R = \begin{bmatrix} PR(T_1) \\ PR(T_2) \\ \vdots \\ PR(T_3) \end{bmatrix} \quad (3)$$

However, R does not consider the influence of musicians on each other. Firstly, we propose each $PR(T_n)=1$. To solve final R , the adjacent matrix A is defined as:

$$A = \begin{bmatrix} \ell(T_1, T_1) & \ell(T_1, T_2) & \dots & \ell(T_1, T_N) \\ \ell(T_2, T_1) & \ddots & & \\ \vdots & & \ell(T_i, T_j) & \\ \ell(T_N, T_1) & & & \ell(T_N, T_N) \end{bmatrix}, \quad (4)$$

where $\ell(T_i, T_j)$ is the probability of T_i is influenced by T_j . And then, R could be calculated as:

$$R = AR \quad (5)$$

At last, through iteration,

$$R = \lim_{n \rightarrow +\infty} A^n R \quad (6)$$

However, equation (6) will waste the computing memory. Equation (6) is a Markov process. A Markov process in which time and state are both discrete is called a Markov chain. The final R will be:

$$R = RA \quad (7)$$

Here, R will be obtained.

3.2 Result and Discussion

After Page Rank, we get every individual musician Page Rank, as Figure 2 shown, Day26 is the most influential musician, which obtains 0.04 Page Rank.

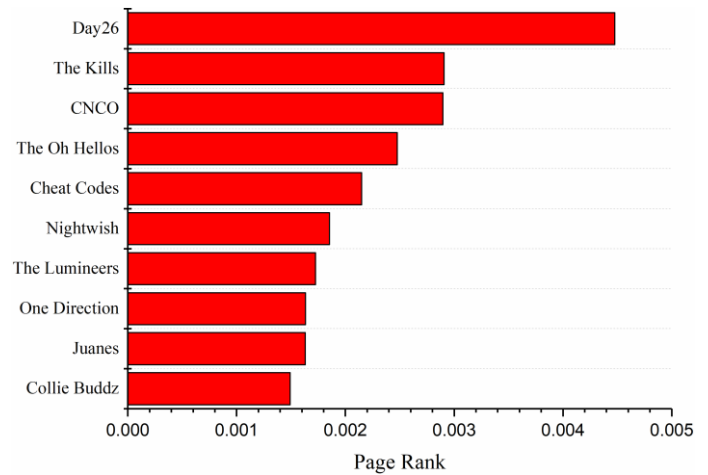


Figure 2: Top 10 Page Rank of musicians

Then, we also calculate the Page Rank of the genres. Figure 3 shows the Page Rank of each genre. Pop/Rock and Electronic obtain the most significant influence.

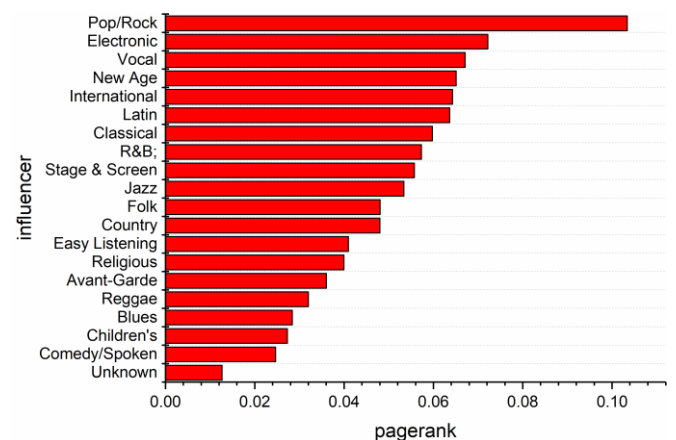


Figure 3: The Page Rank of genres

According to the results obtained by PageRank, Pop/Rock and Jazz is selected to form the influential net. Figure 4 reveals the influential net of Pop/Rock and Figure 5 is the Jazz individually.

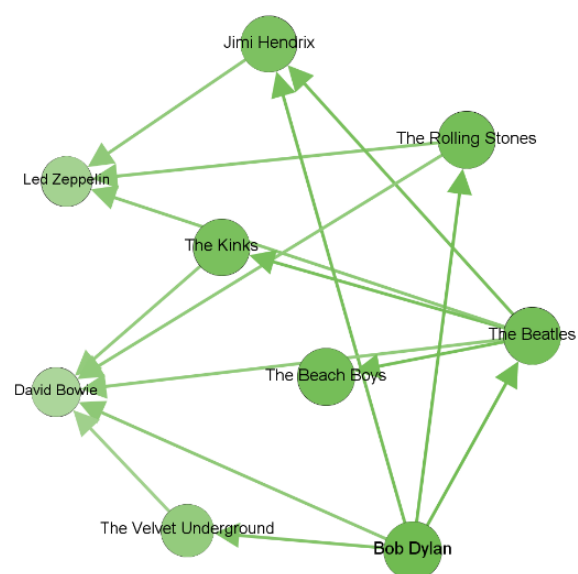


Figure 4: The influential net of Pop/Rock

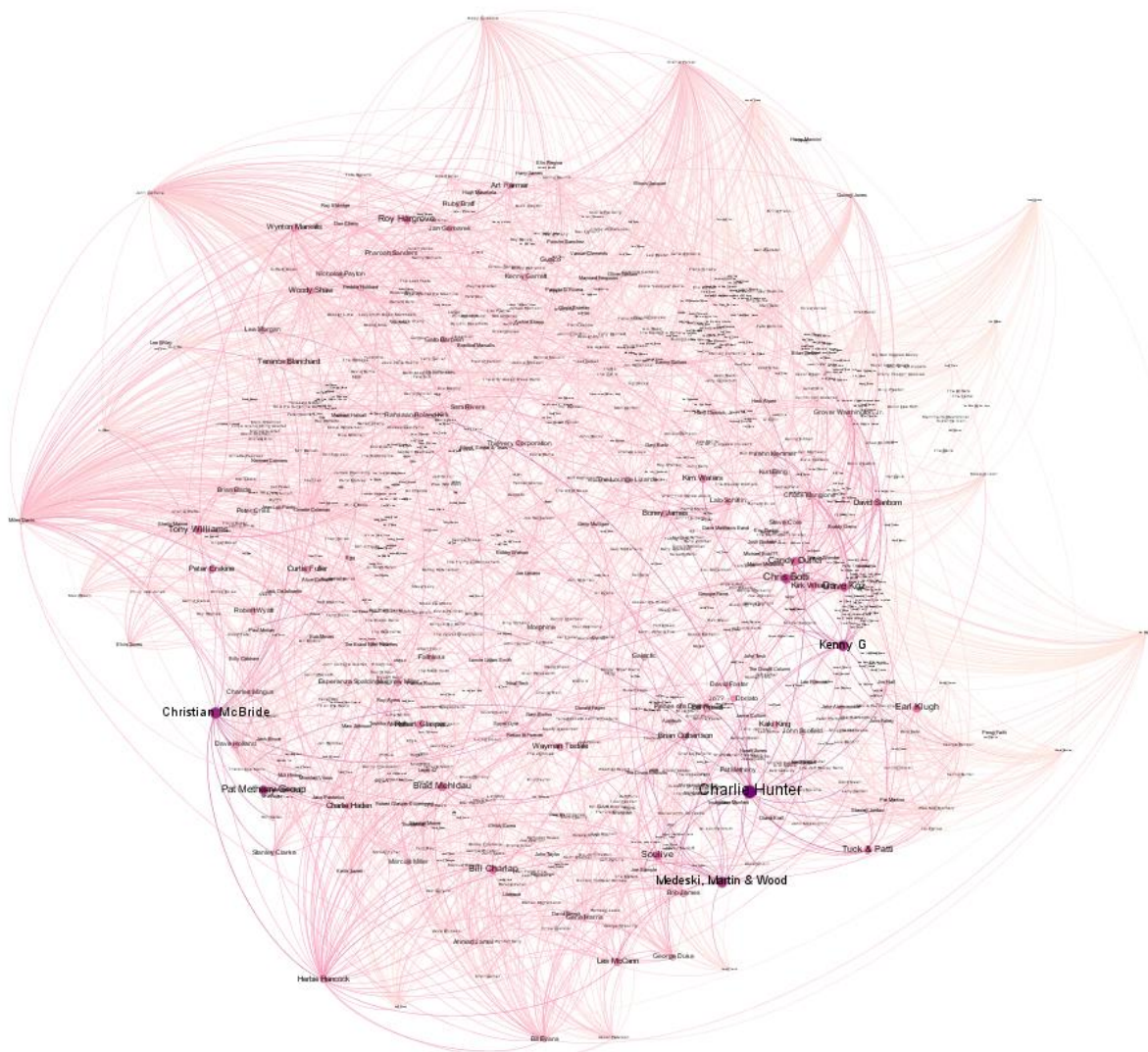


Figure 5: The influential net of Jazz

4. Genre Feature Analysis

4.1 Random Forest

A random forest is a meta estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

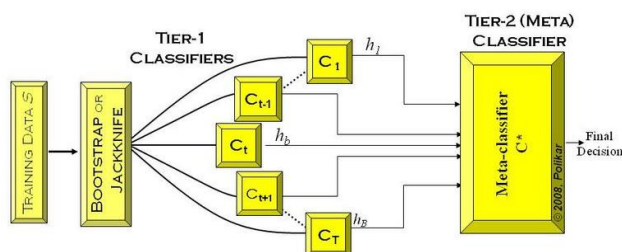


Figure 6: The diagram of Random Forest

When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. This OOB (out-of-bag) data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance.

Random forest mainly includes decision tree growth and voting models. The decision tree growth model utilizes bootstrap aggregating to form a forest.

The random forest voting formula is shown as follow, where n_c is the total number of the total categories, firstly, we will get:

$$C_p = \operatorname{argmax}_c \left(\frac{1}{n_{\text{tree}}} \sum_{i=1}^{n_{\text{tree}}} I \left(\frac{n_{n_i,c}}{n_{h_i}} \right) \right), \quad (8)$$

where $I(X = x_i) = -\log_2 p(x_i)$. The features of Random Forest are the features in Table2, the target is the music genres.

Then cross-validation is used to identify the performance, where kfold is 5. And the accuracy of classification is 64.5%.

4.2 Feature Importance

Here, we introduce GINI impurity to identify the importance of genres features. VIM is defined to represent importance. The value of Gini is GI . We suppose there are m features named $X_1, X_2 \dots X_m$. The GI will be calculated as:

$$GI_m = 1 - \sum_{k=1}^{|K|} p_{mk}^2, \quad (9)$$

where k is the count of classes, p_{mk} is the proportion of the class k in node m .

The feature importance X_j which is also the variety of GI is shown as:

$$VIM_{jm}^{gini} = GI_m - GI_l - GI_r, \quad (10)$$

where GI_l and GI_r is the new Gini index.

Finally, all the value is normalized to get the importance:

$$VIM_j = \frac{VIM_j}{\sum_{i=1}^c VIM_i} \quad (11)$$

The feature importance is shown in Figure 7.

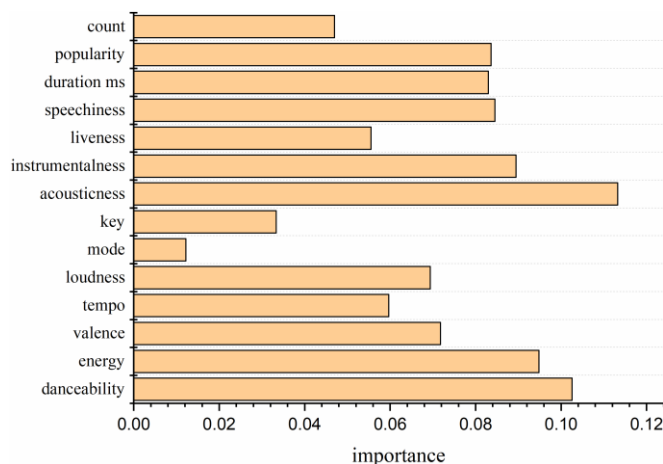


Figure 7: The importance of the features.

Figure 7 illustrates the importance of features. “Acousticness” is the most important feature of the different genres. The mode of the music is the least distinguishable.

5. Conclusion

In this paper, we make research on the artist influence and the difference between the genres. Firstly, we propose Page Rank calculate the influence of the musicians and the genres. Our result is that Day26 is the most influential musician and Pop/Rock is the most influential genre. Secondly, we utilize Random Forest Classifier to make a classification of genres between 14 features, and Gini impurity is used to identify the importance of the features. The found shows that acousticness is the most important feature. Our research is valuable for musicians to identify the genres and the orientation to develop.

References

- [1] Simonetta, F., S. Ntalampiras, and F. Avanzini. Multimodal Music Information Processing and Retrieval: Survey and Future Challenges. in 2019 International

Workshop on Multilayer Music Representation and Processing (MMRP). 2019.

- [2] Li, H. and F. Ren. Computer music and traditional music under the background of “Internet + Music” — Take the film “Song of the Phoenix” for example. in 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). 2020.
- [3] Avanzini, F. and L.A. Ludovico. Multilayer Music Representation and Processing: Key Advances and Emerging Trends. in 2019 International Workshop on Multilayer Music Representation and Processing (MMRP). 2019.
- [4] Liang, B. and M. Gu. Music Genre Classification Using Transfer Learning. in 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). 2020.
- [5] Huang, C.F. and C.Y. Huang. Emotion-based AI Music Generation System with CVAE-GAN. in 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE). 2020.
- [6] Przybysz, A.L., et al. Latin Music Mood Classification Using Cifras. in 2015 IEEE International Conference on Systems, Man, and Cybernetics. 2015.
- [7] Chen, C. and X. Steven. Combined Transfer and Active Learning for High Accuracy Music Genre Classification Method. in 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). 2021.
- [8] Kale, M. and P.S. Thilagam. DYNA-RANK: Efficient Calculation and Updation of PageRank. in 2008 International Conference on Computer Science and Information Technology. 2008.