Personalized Music Recommendation Based on Clustering Algorithm

Zhang Lei¹, Wang Wencong¹, Liu Hao¹, Liu Tong¹, Xu Shengxiang², Li Guangli¹, Lü Shuai^{1*}

¹College of Computer Science and Technology, Jilin University, Changchun 130012, China

²College of Software, Jilin University, Changchun 130012, China

*lus@jlu.edu.cn

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited

ARTICLE DETAILS

Article History:

Received 12 May 2017

Accepted 12 July 2017

Available online 14 September 2017

Keywords:

Recommendation system; Music recommendation; Feature design; Clustering algorithm.

ABSTRACT

Aiming at the recommendation of music field, this paper proposes a music recommendation algorithm based on attribute selection and application clustering. Firstly, the recommended progress of music conduct in-depth analysis to focus on building properties of music and the problems of interaction in the field of music recommendation. With clustering algorithm as the main method, more accurate clustering will make the recommendation more precise. Based on attribute building and clustering, the overall recommendation scheme is designed, and the music is clustered by attribute judgment. Experimental results show that music recommendation proposed algorithm has a better recommendation effect, can effectively improve the user experience.

1. INTRODUCTION

With the rapid development of network techniques, users have to spend more time to find the information needed, which so-called "information overload" phenomenon. It has the personalized recommendation system which provides the active recommendation function to the user, and it is widely used in various portal websites and electronic commerce system. It automatically recommends the object that meets the characteristics of the user [1, 2].

The concept of personalized recommendation system was presented in the 1990s. In the early representative of the AT & T laboratory in 1997, based on the collaborative filtering of personalized recommendation system PHOAKS and Referral Web [3-5].

Music recommendation is more difficult for those people who have their own needs, because people are accustomed to using music to sue their emotions and improve the quality of life. However, so many music software troubles the user. People spend a lot of time to find music which is in line with their own preferences.

This paper presents a music recommendation algorithm, and the recommended design and implementation which is based on the clustering algorithm, to achieve the function that recommending music. The algorithm uses the clustering algorithm to realize the recommendation and optimizes the recommendation process.

2. MUSIC FEATURES DESIGN

To understand the differences and preferences of each user, it is necessary to understand the user's hobby, taboo, etc. Which makes it better to meet the user's needs in the recommendation process. But how can we make the access to the user's preferences more accurate? The corresponding problem is that the protection and disclosure of individual users is particularly acute.

Use the attributes to make the recommended algorithm and users have a secret interaction. It is determined by the attribute editing of music information, and the basic information of individual user is saved at the beginning, and continuous improvement is made. The basic attributes that users can access including User ID, Age, Address, Gender, Phone Number. Do not make special requests to avoid uncomfortable experience when asking too much information.

Singer: Each singer has his own fan group or audience crowd. When users are listening to music, the system usually gives priority to choose the songs of their favorite singer.

Album: The same album in which users chose one song, with the singer the same reason, the album further narrows the range of recommended songs;

Composer: Some of the users are usually very interested in a type of songs. In order to reflect the needs of users, we need to take care of the preferences of various types of users;

Song age: For different ages have different song preferences, depending on the strength and trends of the times, users tend to prefer songs in their own age.

Language: This type of property is mainly for 80, 90, 00 and so on, with the globalization of information, the concept of no national borders in the music field gradually enjoys popular support.

Style: The style is more important and difficult to determine. Refining the genre properties of six categories instead of asymmetric binary properties. These six types of attributes can have a more perfect classification coverage over all the songs of the genre, namely pure music, rock, folk, love songs, nostalgia, fresh.

3. CLUSTERING BASED RECOMMENDATION ALGORITHM

K-Means algorithm similar with K-modes algorithm first select k samples as the centroid, use Europe distance as a similarity measure, for each sample, the remaining calculation to each centroid distance, and put it into the nearest centroid, finally re calculate the centroid [6,7]. Iteration until the centroid is no longer changed. K-Means algorithm can be divided into N data objects K minimum variance clustering. During the experiment, the attributes of N songs are selected as data samples. The K-Means algorithm can be used to divide the N song into a similar class of K properties.

The music recommendation algorithm needs to cluster the music according to the attributes of the music, and the effective extraction of each cluster can improve the system performance and accuracy. For the extraction of musical attributes mainly in the genre of the judgment, and asymmetric two-dimension attributes.

In the data collection phase, 1-10 style class attribute points score, the rules for different positions belong to different songs from the list in the data, different positions represent different scores, in the ranking position closer to the former, the higher the score is. And the same song can appear in multiple lists.

After the completion of the above steps, you can begin to build a music recommendation algorithm based on clustering:

Step 1. The clustering algorithm is used to cluster the attributes of the music, and the music of the data set is clustered to form different clusters, the similarity of the songs in the same cluster is high, and the songs of different clusters are large;

Step 2. Understand user preferences and learn user preferences by interacting with the user. Select a certain number of songs in each cluster to recommend to the user, according to the recommendation of the user's songs, such as pull black, listen, like, skip and other operations.

Step 3. Begin to recommend songs for users. First in each cluster for users to recommend a few songs to generally understand the user's preferences. In the initialization of the song, each song will have 10 minutes of the initial points. Users like all the songs in the cluster plus if not like all the songs in the cluster score minus 5. After the initial recommendation algorithm. According to the score of the song, the higher the score, the higher the probability of being recommended. A maximum of 20 points and a minimum of 0 points is not recommended.

Step 4. Music recommendation to the user, the proposed algorithm in the implementation of each recommendation will be based on the user's behavior and then fine tune the score of all songs to achieve the ultimate goal of accurate recommendation.

4. EXPERIMENT ANALYSIS

4.1. Selection of Clustering Attributes

Cluster analysis was performed using a 10x4 property attribute matrix. The matrix attribute is deleted one column and two columns to get 10x3 and 10x2 items attribute matrix, then it is used to simulate the situation of excessive reduction or number of attributes is too small. The last column of the matrix is copied and rerouted out of two columns to get 10x5 and 10x6 item attribute matrices to simulate the situation of redundant attributes. All matrix data is shown in Figure 1.

81	5	81	5	90	81	5	90	5	81	5	90	5	5	81	5	90	5	5	5
94	15	94	15	90	94	15	90	5	94	15	90	5	5	94	15	90	5	5	5
95	5	95	5	25	95	5	25	25	95	5	25	25	25	95	5	25	25	25	25
88	5	88	5	5	88	5	5	45	88	5	5	45	45	88	5	5	45	45	45
99	10	99	10	80	99	10	80	15	99	10	80	15	15	99	10	80	15	15	15
75	5	75	5	85	75	5	85	45	75	5	85	45	45	75	5	85	45	45	45
95	80	95	80	25	95	80	25	5	95	80	25	5	5	95	80	25	5	5	5
87	5	87	5	45	87	5	45	35	87	5	45	35	35	87	5	45	35	35	35
93	80	93	80	25	93	80	25	5	93	80	25	5	5	93	80	25	5	5	5
99	95	99	95	95	99	95	95	5	99	95	95	5	5	99	95	95	5	5	5

Figure 1: Item attribute matrix

Using the K-Means algorithm, the number k is set to 4, and the difference between the clustering results in each case and the clustering results in the standard case is compared. The clustering results in each case are shown in Table 1. The data in the table is the cluster number of the cluster to which the item belongs, and the last row is the result of comparison with the standard cluster in each case.

It can be concluded from the data in the table that when the number of attributes in the cluster is too small, the clustering results cannot show the characteristics of the individual well. When the attributes in the cluster are redundant, the clustering results change little. Therefore, attributes should be selected between the smaller redundancy of the property.

In the original experimental data, there are song labels and the score values of the tags. The two-sample t-test is used to verify the redundancy matrix in the above data [8]. And remove the redundant attributes to get the standard attribute. Use this method to verify the above simulation data (10x6 item attribute matrix) In the t test, the smaller the t value is, the greater the similarity between the eigenvalues. The oblique diagonal elements should be removed from the above matrix. The similarity between attribute 4 and attribute 5,6 is big, so they should be removed too. Finally, the attribute 1,2,3,4 is left.

4.2. Clustering Process

The clustering algorithm uses the K-Means algorithm [9]. The test data set is still tested using the 10x4 item attribute dataset described above. The choice of K value is very important. A single point cluster begins when K = 4, and K = 3 should be selected to avoid the generation of single point clusters. The same method is used to test all the data. After analyzing the total data, a single point cluster is generated at K=7. Therefore, the data set should take the K value of 6, and the resulting clustering center performs the projection transformation of the first two attributes, the clustering results are as follows:

Figure 2: Data set cluster results

4.3. Recommended Process

After the clustering algorithm above is completed, the data will be divided into K clusters. According to the classification of the data classification of music, the recommended priority decides the recommended order of the song. The specific algorithms and processes are as follows:

(1) The first step is evaluating the clusters. When $K = 3$, three different clusters of music will be recommended to the user in turn, according to the user's score to determine the priority to the cluster. For example, the user 1 makes the evaluation on the cluster for like, do not like, do not like.

(2) The algorithm adjusts the priority of the cluster according to the user's further rating. When the user enters the "official recommendation" interface, there are "listening", "skip" and other operations, according to the user different operations, there will be different scores, thus affecting the priority of the cluster. If the user 1 listened to song 1, skip song 2, then the relevant priority list after the operation is:

Table 2: Recommended list after the general operation

Song number	Priority	Belongs to cluster
1	3	1
2	0	2
3	3	1
4	3	1
5	3	1
6	1	3
7	0	2
8	0	2
9	1	3
10	1	3

(3) User has a particular preference for a particular music. When the user collects songs, even if the music is not in the priority of the cluster, the music will still be recommended. When the user pulls songs into the blacklist, even if the music is in a high priority cluster, the music will still be rejected. If the user pulls 5 songs into the blacklist and collects 7 songs, then the recommendation priority list changes to:

Table 3: Recommended list after specified operation

Song number	Priority	Mark	Belongs to cluster
1	3	General	1
2	0	General	2
3	3	General	1
4	3	General	1
5	0	Blacklist	1
6	1	General	3
7	5	Collect	2
8	0	General	2
9	1	General	3
10	1	General	3

5. CONCLUSIONS

In this paper, the K-Means clustering algorithm is analyzed based on the related principles of recommendation technology, and the music recommendation algorithm based on K-Means is proposed. At the same time, the selection of clustering attribute is discussed. Through the structure of the music features, we can recommend music which are more in line with the user's preferences and improve the user's experience.

In future research, we will focus on how to extract the data that meets the requirements in a large number of unstructured text and how to model the data structure more accurately.

ACKNOWLEDGEMENT

The work was supported by the National Nature Science Foundation of China under Grant Nos.61300049, 61502197, 61503044, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No.20120061120059, the China Postdoctoral Science Foundation under Grant No.2011M500612, the Key Program for Science and Technology Development of Jilin Province of China under Grant No.20130206052GX, the Natural Science Research Foundation of Jilin Province of China under Grant Nos.20140520069JH, 20150520058JH, 20150101054JC.

REFERENCES

- [1] Kanetkar, S., Nayak, A., Swamy, S., Bhatia, G. 2015. Web-based personalized hybrid book recommendation system, International Conference on

Advances in Engineering and Technology Research. IEEE, 1-5.

- [2] Deng, S., Wang, D., Li, Y., Cao, B., Yin, J., Wu, Z., Zhou, M. 2016. A Recommendation System to Facilitate Business Process Modeling. *IEEE Transactions on Cybernetics*, 47 (6), 1380 - 1394.
- [3] Hsu, W.C., Chen, L.C. 2017. A Novel Recommendation System for Dental Services Based on Online Word-of-Mouth. *Information Resources Management Journal*, 30 (1), 30-47.
- [4] Ren, Y., Harper, F.M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., Kraut, R.E. 2012. Building member attachment in online communities: applying theories of group identity and interpersonal bonds. *Mis Quarterly*, 36 (3), 841-864.
- [5] Xu, L., Li, L., Nagarajan, V., Huang, D., Tsai, W.T. 2013. Secure Web Referral Services for Mobile Cloud Computing IEEE. *Service Oriented System Engineering (SOSE)*, 2013 IEEE 7th International Symposium on, 584-593.
- [6] Shahrivari, S., Jalili S. 2016. Single-pass and linear-time k-means clustering based on MapReduce. *Information Systems*, 60 (C), 1-12.
- [7] Chen, X., Yang, F., Zhang, C., Zhou, J., Guo, L.J. 2016. Large-Area High Aspect Ratio Plasmonic Interference Lithography Utilizing a Single High-k Mode. *American Chemical Society Nano*, 10 (4), 4039–4045.
- [8] Allefeld, C., Görgen, K., Haynes, J.D. 2016. Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *Neuroimage*, 141, 378-392.
- [9] Capó, M., Pérez, A., Lozano, J.A. 2017. An efficient approximation to the K -means clustering for massive data. *Knowledge-Based Systems*, 117, 56-69.