A User-Interests Approach to Music Recommendation

Ye-In Chang, Chen-Chang Wu and Meng-Chang Tsai

Abstract—**In recent years, music has become increasingly universal due to technological advances. All kinds of music have become more complex and a large amount around us. How recommending the music that user is interested in from a wide variety of music is the development intentions of the music recommendation system MRS (Music Recommendation System). Chen *et al.* have proposed the CB method for music recommendation. The CB method is to recommend the music objects that belong to the music groups the user is recently interested in. Each transaction is assigned a different weight, where the latest transaction has the highest weight. But in the CB method, the formula of computing music group weight pays much attention to the weight of the transaction. This will lead to the result that the group weight of music group B which appears once in the later transaction is larger than the group weight of the music group A which appears many times in the earlier transaction. Therefore, in this paper, we propose the TICI (Transaction-Interest-Count-Interest) method to improve the CB method. Considering the two situations of the music group that user is interested in, the large count of music group and the appearance in the later transaction, we put two parameters: Count-Interest and Transaction-Interest in our TICI method to let users choose which weight they want to emphasize. From our simulation results, we show that our TICI method could provide better performance than the CB method.**

Index Terms—**music recommendation system, user interest, transaction, count, weight.**

I. Introduction

TWO approaches for a recommendation system have been discussed in the literature: the content-based filtering approach and the collaborative filtering approach. In the content-based filtering approach, the representations of the data items which have been accessed in the past are used as the user profiles. Based on the user profiles, the system recommends only the data items that are highly relevant to the user profiles by computing the similarities between the data items and the user profiles. Examples of such systems are News Dude [1], Infofinder [6], and NewsWeeder [7]. In this approach, the representation of data items and the description of user preferences in profiles are key issues which dominate the effectiveness of recommendation.

Instead of computing the similarities between the data items and the user profiles, the collaborative approach computes the similarities between the user profiles. Users of similar profiles will be grouped together to share the information in their profiles. The main goal of the collaborative approach is to make recommendation among the users in the same group. Examples of such systems are Siteseer [10] and Ringo [11]. In the collaborative filtering approach, the system may have a high possibility to recommend unexpected data items by the nature of information sharing.

Some systems use both contend-based and collaborative filtering approaches. The FAB system [2] analyzes the accessed webpages to derive the user profiles and compares the user profiles to group users for collaborative recommendation. For the video data, the recommendation system is developed in [3]. The Personalized Television system [12] provides a personalized list of recommended programs.

In recent years, the music becomes more popular due to the evolution of the technology. Various kinds of music around us become more complexity and huge. In addition to searching expected music objects for users, it becomes necessary to develop a music recommendation service. The Music Recommendation System (MRS) is a website which provides the service of music recommendation based on music data grouping and user interests.

There have been many researches in the field of MRS, such as content-based music filtering system with editable user profile [5], the sensitivities of user profile information in music recommender systems [8] and the music recommendation based on music data grouping and user interests [4]. A content-based music filtering system with an editable user profile [5] is using a decision tree in a content-based music filtering system [9]. The sensitivities of user profile information is describing empirical research into the factors influencing the trade-off between the perceived benefits of personalization and the privacy 'costs' experienced by individuals [8]. Instead of textual descriptions, the music recommendation based on music data grouping and user interests considers the perceptual properties of music objects, such as pitch, duration, and loudness, which can be directly extracted from the music objects [4].

Arbee L.P. Chen *et al.* have proposed an alternative way of music recommendation based on music data grouping and user interests [4]. For users, the preferences are derived from the access histories and recorded in profiles. They have proposed the CB method to recommend the music objects that belong to the music groups the user is recently interested in. To capture the recent interests of the user, they analyze the latest transactions in the access history as follows. Each transaction is assigned a different weight, where the latest transaction has the highest weight. An example of the access history is shown in Table I. Suppose the number of music objects to be recommended is 20. The result of recommending music objects is shown in Table II.

The CB method recommends recently hot music to users according to the access history of users. But in the CB method, the formula of computing music group weight pays much attention to the weight of the transaction occurring time. Table III is an example of access history H1 of a user. In Table III, we focus on the group weight of groups A

Manuscript received December 14, 2010.
The authors are with the Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. (e-mail: changyi@cse.nsysu.edu.tw).

TABLE I
A SAMPLE OF THE ACCESS HISTORY

Access Time	Object ID	Music Group	Transaction
2001/4/06 AM 11:47:03	1	B	T1
2001/4/06 AM 11:47:03	23	C	T1
2001/4/12 AM 10:11:25	7	D	T2
2001/4/12 AM 10:11:25	5	C	T2
2001/4/12 AM 10:11:25	32	B	T2
2001/4/16 AM 09:51:33	16	A	T3
2001/4/16 AM 09:51:33	19	B	T3
2001/4/16 AM 09:51:33	42	A	T3
2001/4/20 AM 08:31:12	31	D	T4
2001/4/20 AM 08:31:12	63	C	T4
2001/4/20 AM 08:31:12	26	A	T4
2001/4/22 AM 10:24:49	53	B	T5
2001/4/22 AM 10:24:49	12	A	T5

TABLE II
NUMBER OF MUSIC OBJECTS TO THE RECOMMENDED IN EACH GROUP

Music Group	Number of Recommended Music Objects
A	8
B	6
C	4
D	4

and B. We find group A appearing many times in the early transactions. On the other hand, group B appears one time in the latest transaction. But the group weight of group B is larger than the group weight of group A in the CB method, the result is not conventional. Observing the result of the CB method, we can find when the music group B appears once in the later transaction, it will have larger group weight than the group weight of the music group A which appears many times in the earlier transaction. This result may be not good for some users, because the purpose of the CB method is to recommend the music object which the users are interested. When the count of music group is large in the user's access history, it means that this user is interested in this group, too.

Therefore, in this paper, we propose the TICI (Transaction-Interest-Count-Interest) method to improve the performance of the CB method. In our TICI method, for the same access history shown in Table III, we can decide the rank of the music group weight between groups A and B. And we put two parameters: Count-Interest CI and Transaction-Interest TI in our TICI method to let user choose which weight they want to emphasize. From the simulation results, we show that our TICI method could provide better performance than the CB method in terms of the weight differences. That is, our TICI method can decide the rank of the group weight precisely.

The rest of the paper is organized as follows. Section 2 gives a survey of some music recommendation systems. Section 3 presents the proposed TICI method. Section 4 makes a comparison between our TICI method and CB method. Finally, Section 5 gives the conclusions.

TABLE III
A SAMPLE OF THE ACCESS HISTORY H1

Transaction	Music Group
T1	AA
T2	AC
T3	DEF
T4	GHI
T5	JK
T6	B

TABLE IV
A SAMPLE OF THE ACCESS HISTORY

Access Time	Object ID	Music Group	Transaction
2001/4/06 AM 11:47:03	1	B	T1
2001/4/06 AM 11:47:03	23	C	T1
2001/4/12 AM 10:11:25	7	D	T2
2001/4/12 AM 10:11:25	5	C	T2
2001/4/12 AM 10:11:25	32	B	T2
2001/4/16 AM 09:51:33	16	A	T3
2001/4/16 AM 09:51:33	19	B	T3
2001/4/16 AM 09:51:33	42	A	T3
2001/4/20 AM 08:31:12	31	D	T4
2001/4/20 AM 08:31:12	63	C	T4
2001/4/20 AM 08:31:12	26	A	T4
2001/4/22 AM 10:24:49	53	B	T5
2001/4/22 AM 10:24:49	12	A	T5

II. RELATED WORK

The music objects in the database of the Music Recommendation System (MRS), as well as the incoming music objects, are candidates for music recommendation. When a new music object is inserted in the database of the MRS, it goes through the track selector and the feature extractor. According to the extracted features, the incoming music object is properly assigned to certain music group by the classifier function block. When the user accesses a music object from the list of music objects or the recommendation results, the profile manager will record the object information into the access history. An example of the access history is shown in Table IV. As shown in Table IV, the information of each accessed music object, *i.e.*, the access time, the object ID, the corresponding music group which the object belongs to, and the corresponding transaction is recorded in the access history. Note that the transaction ID is monotonically increasing.

Arbee L.P. Chen *et al.* have proposed the CB method to recommend the music objects that belong to the music groups the user is recently interested in [4]. Instead of textual descriptions, they consider the perceptual properties of music objects, such as pitch, duration, and loudness, which can be directly extracted from the music objects. For users, the preferences are derived from the access histories and recorded in profiles.

To capture the recent interests of the user, they analyze the latest transactions in the access history as follows. Each transaction is assigned a different weight, where the latest transaction has the highest weight. The weight GW_i of music group G_i is computed as follows:

$$GW_i = \sum_{j=1}^{n} TW_j * MO_{j,i} \tag{1}$$

where TW_j is the weight of transaction T_j, n is the number of latest transactions used for analysis, $M0_{j,i}$ is the number of music objects which belong to music group G_i in transaction T_j. These weights will be recorded in a preference table for the user. The MRS ranks all the music groups.

To avoid recommending a large number of music objects to users, the MRS limits the number of music objects for

TABLE V
THE PREFERENCE TABLE FOR THE USER

Music Group	Weight
A	3.08
B	2.5616
C	1.7216
D	1.312

TABLE VI
NUMBER OF MUSIC OBJECTS TO THE RECOMMENDED IN EACH GROUP

Music Group	Number of Recommended Music Objects
A	8
B	6
C	4
D	4

recommendation. The number of music objects R_j from each music group is decided as follows:

$$R_i = \lceil N * \frac{GW_i}{\sum_{k=1}^{M} GW_k} \rceil \tag{2}$$

where N is the number of music objects in the recommendation list, GW_i is the weight of the target group, M is the total number of music groups in MRS. In the same music group, the latest music object will be first recommended.

Example. Take the user's access history shown in Table IV as an example. Assign the weights 0.4096, 0.512, 0.64, 0.8, and 1 to Tl, T2, T3, T4, and T5, respectively. The weight for each music group is calculated, as shown in Table V.

According to Table V, the total weight of all music groups is 8.6752. Suppose the number of music objects to be recommended is 20. The result of recommending music objects is shown in Table VI.

III. The Proposed Method

First, we give initial conditions of music recommendation system [4]. When a user accesses a music object from the list of music objects or the recommendation results, the profile manager will record the object information into the access history. An example of the access history H1 is shown in Table VII. As shown in Table VII, the information of each accessed music object, *i.e.*, the access time, the object ID, the corresponding music group which the object belongs to, and the corresponding transaction are recorded in the access history. Note that the transaction ID is monotonically

TABLE VII
A SAMPLE OF THE ACCESS HISTORY H1

Access Time	Object ID	Music Group	Transaction
2001/4/06 AM 11:47:03	1	B	T1
2001/4/06 AM 11:47:03	23	C	T1
2001/4/12 AM 10:11:25	7	D	T2
2001/4/12 AM 10:11:25	5	C	T2
2001/4/12 AM 10:11:25	32	B	T2
2001/4/16 AM 09:51:33	16	A	T3
2001/4/16 AM 09:51:33	19	B	T3
2001/4/16 AM 09:51:33	42	A	T3
2001/4/20 AM 08:31:12	31	D	T4
2001/4/20 AM 08:31:12	63	C	T4
2001/4/20 AM 08:31:12	26	A	T4
2001/4/22 AM 10:24:49	53	B	T5
2001/4/22 AM 10:24:49	12	A	T5

TABLE VIII
DESCRIPTION OF PARAMETERS

Parameter	**Description**
TID	The target transaction ID
$FTID$	The first transaction ID
Nt	The number of transactions
TW_j	The weight of transaction T_j
GW_i	The weight of music group G_i
CI	The interest of the count
TI	The interest of the transaction
$GA_{j,i}$	The number of appearances of music group i in transaction j

increasing. Each transaction is assigned a different weight, where the latest transaction has the largest weight. Moreover, the music group containing more accessed music objects in a transaction has a larger weight than other groups in the same transaction. According to the weights of music groups, different numbers of music objects from the music groups will be recommended. For music group G_i, we select the latest R_i music objects which have not been accessed by the user. In the recommendation list, the music objects will be sorted by the corresponding group.

Although the CB method can find the recently hot music group according to the user's access history [4], the result is not fair. They pay much attention to the weight of time. Therefore, we propose a fair formula which emphasizes both the weight of time and the weight of count of music group.

To simplify the comparison, first, we assume that two same groups will not appear in one transaction. We use the example that is an access history H1 which has six transactions to compare the results between our formula and the CB method. Table VIII shows the parameters used in our proposed method.

The formula for the group weight of group i (GW_i) of CB method is

$$GW_i = \sum_{j=1}^{n} TW_j * GA_{j,i} \tag{3}$$

where TW_j is the weight of transaction j and $GA_{j,i}$ is the number of music objects which belong to music group G_i in transaction T_j. In this formula, the equation of TW_j is not given in [4]. Therefore, we give a new equation of TW_j:

$$TW_j = \frac{TID_j - FTID + 1}{Nt} \tag{4}$$

where TID_j is the target transaction ID, $FTID$ is the first transaction ID and Nt is the number of transactions. Note that the following TW_j is calculated by using this equation.

According to Formula 3, when the music group A appears once in the later transaction, it will have larger weight than the weight of the music group B appears many times in the earlier transaction. This result is not good for some users, because the purpose of the CB method is to recommend the music object which the users are interested. When the count of music group is large in the user's access history, it means that this user is interested in this group, too. Therefore, we propose a new formula to avoid this problem:

$$GW_i = \sum_{j=1}^{n} TW_j * TI + GA_{j,i} * CI \tag{5}$$

TABLE IX
PARAMETERS USED IN THE EXPERIMENT

Parameters	**Meaning**
N	The number of transactions in the access history
M	The number of music groups
MinT	The minimum length of the transaction
MaxT	The maximum length of the transaction

TABLE X
THE ACCESS HISTORY OF USER U1

Transaction	Music Group
T1	A,B,C,D,E
T2	A,C,E
T3	A,B,C,D,E
T4	C,E,K
T5	A,D,E,M

where Count-Interest CI ($0 \leq CI \leq 1$) and Transaction-Interest TI ($0 \leq TI \leq 1$) are assigned by users, TI = 1 - CI. According to each user's preferences, our formula adds two parameters CI and TI to let users decide the importance of the time and count.

IV. PERFORMANCE

In this section, we study the performance of the proposed TICI. We also make a comparison with the CB method. The simulation was performed on an Intel Pentium Core2 1.86G Hz CPU computer with 1GB of RAM, and the operation system is Microsoft Windows XP service pack 3.

A. Generation of Synthetic Data

We generated synthetic access histories to evaluate the performance of the methods. The parameters used in the generation of the synthetic data are shown in Table IX. The length of a transaction is chosen randomly between *MinT* and *MaxT*. For the TICI method, the *MinT* and *MaxT* is 2 and 5, respectively. In the comparison between the CB method and the TICI method, the music group will appear one time in a transaction or appear more than one time in a transaction. Therefore, for the music group appears more than one time in a transaction, we choose the music group in the set of music group randomly. For the music group appears one time in a transaction, we use the flag to record the appearance of the music group so that the music group will not appear again in a transaction. This way can achieve the goal that the music group appears one time in a transaction. For generating synthetic data, we assign an *occur rate Orate*, and we generate a random real number which is between 0 and 1. If the random number is larger than *Orate*, the generation runs normally. If the random number is smaller than or equal to *Orate*, we let the music group which never appears in the earlier transactions appear in the last transaction. The larger *Orate* is, the larger repeatability of the music group is. We call this synthetic data *DataType1*.

B. Simulation Results of Synthetic Data

In this subsection, we make a comparison of our TICI method with the CB method by using the synthetic data *DataType1*. We study the impact of five parameters on Table IX. We let *M* be 50 because the numbers of music groups are not more than 50 in the current environment and we let *MinT* = 2 and *MaxT* = 5. Moreover, we make the comparison between the TICI method and the CB method under the two cases. One case is that the music group can appear more than one time in a transaction, another case is that the music group can only appear one time in a transaction. A comparison of the music group weight difference in the TICI method and CB method is shown in Figure 1 and Figure 2.

Note that our TICI method has three cases: CI = 0.3, CI = 0.5 and CI = 0.7 to compare the change of the result that emphasizes weight of transaction and the weight of count. We use the group weight difference to be our performance measure. The group weight difference is the difference between the group weights of the group weight rank which are decided by the methods and we add all group weight differences to be the results of comparison between our TICI method and the CB method. When the group weight difference is larger, it means that the method can decide the rank of the group weight clearly.

TABLE XI
THE ACCESS HISTORY OF USER U2

Transaction	Music Group
T1	A,B,C,D,E
T2	B,C,E
T3	A,B,C,D,E
T4	C,E,K
T5	B,D,E,M

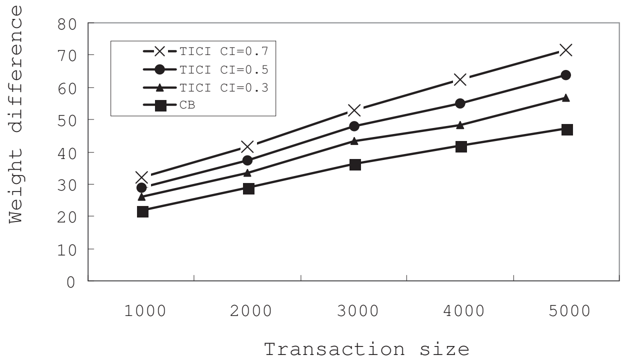

Fig. 1. A comparison of the group weight difference under the case that the music group appears more than one time in a transaction

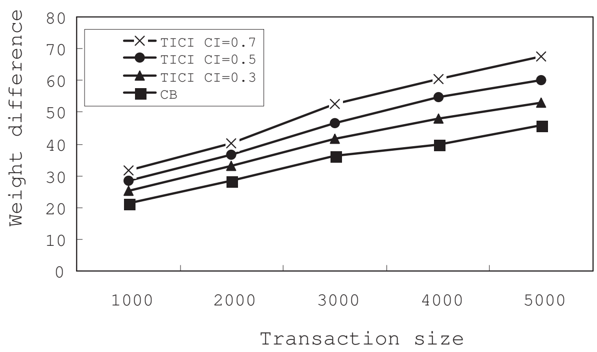

Fig. 2. A comparison of the group weight difference under the case that the music group only appears one time in a transaction

In Figure 1, the range of N is set to 1000, 2000, 3000, 4000 and 5000, while the other parameters are kept as their base values. Under changing the value of N, we can find that the group weight differences of our TICI method are larger than that of the CB method. Because the CB method only emphasizes the weight of transaction, the impact of the count of music group decreases when the transaction size increases. For example, when the transaction size is 1000, the transaction weight of Transaction 5 is $\frac{5}{1000}$ and the transaction weight of Transaction 900 is $\frac{900}{1000}$. The group weight of group A that appears five times in Transaction 5 is still smaller than the group weight of group B which appears one time in Transaction 900. Therefore, the group weights are usually the same; that is, the CB method usually can not decide the rank of the group weight so that the group weight difference of the CB method is small.

In the three cases of our TICI method, we can find that the rank of the group weight difference is CI = 0.7, CI = 0.5 and CI = 0.3. According to the result, we can find when we emphasizes the weight of transaction, *i.e.*, CI = 0.3 and TI = 0.7, the group weight difference is smaller than other cases. When we emphasizes the weight of count, *i.e.*, CI = 0.7 and TI = 0.3, the group weight difference is larger than other cases. The reason is that the impact of the count is larger than the impact of the transaction weight. For example, the transaction size is 1000 and the transaction weight is between $\frac{1}{1000}$ and $\frac{1000}{1000}$. On the other hand, the music group at least appears one time, the count is larger than or equal to the transaction weight. Therefore, the impact of the count is always larger than the impact of the transaction weight. When the transaction size increases, the impact difference between the count and the transaction weight increases. Therefore, the larger the transaction size is, the larger the group weight difference is.

From Figure 1 and Figure 2, we can find the group weight difference in Figure 1 is larger than the group weight difference in Figure 2. Because the data in Figure 1 allows that the music group can appear more than one time in a transaction, and the data in Figure 2 only allows that the music group appear one time in a transaction. The count of the music group in Figure 1 is larger than the count of the music group in Figure 2. Therefore, the group weight difference in Figure 1 is larger than the group weight difference in Figure 2.

V. Conclusion

In this paper, we have proposed the TICI (Transaction-Interest-Count-Interest) method for the music recommendation in music databases. The TICI method can improve the performance of the CB method by change the formula which calculates the weight of music group. We also have studied the performance of the TICI method and the CB method. The simulation results have shown that the performance of the TICI method is better than that of the CB method in terms of the weight difference. This is because the TICI method consider the count of the music group and the time of appearance of the music group so that the TICI method can decide the rank of the group weight more precisely than the CB method.

Acknowledgment

This research was supported in part by the National Science Council of Republic of China under Grant No. NSC-97-2221-E-110-058-MY3 and National Sun Yat-Sen University.

References

[1] D. B. And, D. Billsus and M. J. Pazzani, "A Hybrid User Model for News Story Classification," in *Proc. of the 7th Int. Conf. on User Modeling*, Banff, Canada, 1999, pp. 99-108.

[2] M. Balabanovic and Y. Shoham, "Content-based, Collaborative Recommendation," *Commun. ACM*, vol. 40, no. 3, pp. 66-72, 1997.

[3] C. Basu, H. Hirsh and W. W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," in *Proc. of National Conf. on Artificial Intelligence*, Madison, Wisconsin, 1998, pp. 714-720.

[4] H. C. Chen and A. L. P. Chen, "A Music Recommendation System Based on Music Data Grouping and User Interests," in *Proc. of the 10th Int. Conf. on Information and Knowledge Management*, Atlanta, Georgia, 2001, pp. 231-238.

[5] Y. Hijikata, K. Iwahama and S. Nishida, "Content-based Music Filtering System with Editable User Profile," in *Proc. of ACM Symp. on Applied Computing*, Kennesaw, Georgia, 2006, pp. 1050-1057.

[6] B. Krulwich and C. Burkey, "Learning User Information Interests Through Extraction of Semantically Significant Phrases," in *Proc. of the AAAI Spring Symposium on Machine Learning in Information Access*, Technical Papers, Stanford, 1996.

[7] K. Lang, "NewsWeeder: Learning to Filter Netnews," in *Proc. of the 12th Int. Machine Learning Conf.*, Tahoe City, California, 1995, pp. 331-339.

[8] E. Perik and B. de Ruyter, P. Markopoulos and B. Eggen, "The Sensitivities of User Profile Information in Music Recommender Systems," in *Proc. of the 2nd Annual Conf. on Privacy, Security and Trust*, New Brunswick, Canada, 2004, pp. 137-141.

[9] J. R. Quinlan, *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*, 1993.

[10] J. Rucker, and M. J. Polanco, "Siteseer: Personalized Navigation for the Web," *Commun. ACM*, vol. 40, no. 3, pp. 73-76, 1997.

[11] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating "Word of Mouth"," in *Proc. of ACM Conf. on Human Factors in Computing Systems*, Denver, Colorado, 1995, pp. 210-217.

[12] B. Smyth and P. Cotter, "A Personalized Television Listings Service," *Communications of the ACM*, vol. 43, no. 8, pp. 107-111, 2000.