A music recommendation system based on music data grouping and user interests

Hung-Chen Chen and Arbee L.P. Chen

Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan 300, R.O.C.
Email: alpchen@cs.nthu.edu.tw

ABSTRACT

With the growth of the World Wide Web, a large amount of music data is available on the Internet. In addition to searching expected music objects for users, it becomes necessary to develop a recommendation service. In this paper, we design the Music Recommendation System (MRS) to provide a personalized service of music recommendation. The music objects of MIDI format are first analyzed. For each polyphonic music object, the representative track is first determined, and then six features are extracted from this track. According to the features, the music objects are properly grouped. For users, the access histories are analyzed to derive user interests. The content-based, collaborative and statistics-based recommendation methods are proposed, which are based on the favorite degrees of the users to the music groups. A series of experiments are carried out to show that our approach is feasible.

Keywords

music recommendation, perceptual properties, access histories, recommendation methods, user profiles

1. Introduction

Concerning a large amount of various data available on the Internet, there exist websites which provide services for users to look for useful data. For text data in webpages, the websites providing keyword-based searching or recommendation are developed, such as the search engine of Yahoo! [Yaho] and the book recommendations of Amazon [Amaz]. For multimedia data, however, the websites providing such kinds of services are still limited.

Regarding the music recommendation, a preliminary recommendation can be accomplished by notifying users when new music objects arrive. The mechanism for the notification service is described as follows. For an incoming music object, the corresponding description is manually attached to the music object, such as the music genre, title, and composer. The users are required to specify their preferences in music. The users' preferences will be compared with the descriptions of the music objects. If matched, the system will send a notification of the matched music objects to those interested users.

In this paper, we propose an alternative way of music recommendation. Instead of textual descriptions, we 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. Three recommendation methods are proposed to approach the corresponding goals. Based on the perceptual properties of the music objects and the elaborated profiles, better recommendation can be obtained by applying our methods.

The rest of this paper is organized as follows. In Section 2, we introduce the music recommendation system in which the modules of track selector, feature extractor and classifier are detailed. In Section 3, we present the three recommendation methods which have been implemented in our system. In Section 4, we perform a series of experiments and illustrate the experiment results to show that our approach is feasible. Finally, in Section 5, we conclude this paper.

1.1 Related work

Two approaches for a recommendation system have been discussed in the literature, *i.e.*, 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 NewsWeeder [Lang95], Infofinder [Krul96], and News Dude [Bill99]. 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

CIKM'01, November 5-10, 2001, Atlanta, Georgia, USA.

recommendation among the users in the same group. Examples of such systems are Ringo [Shar95] and Siteseer [Ruck97]. 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. For example, Tapestry [Gold92] and GroupLens [Kons97] allow users to comment cm Netnews and group users by computing the similarities of their ratings of newsgroups. In addition, for the process of recommendation, users have to specify their profiles and describe the features of data items which they are interested in. For the video data, the recommendation system is developed in [Basu98]. The user interests are derived from the types, the actors, and the scenarios of videos that the user accessed in the past. The users are also required to specify the satisfactory degrees of the accessed videos. With respect to videos, users who specify similar satisfactory degrees will be grouped together for collaborative recommendation. Similarly, the Personalized Television system [Smit00] provides a personalized list of recommended programs. The FAB system [Bala97] analyzes the accessed webpages to derive the user profiles and compares the user profiles to group users for collaborative recommendation. The OTS [Wu01] employs the techniques of association rule mining to derive user interests and behaviors to be used as the user profiles. After classifying the user profiles into clusters, three kinds of recommendation methods are then provided using these clusters.

2. Music Recommendation System

The Music Recommendation System (MRS) is a website which provides the service of music recommendation based on music data grouping and user interests. The music objects in the database of MRS, as well as the incoming music objects, are candidates for music recommendation. As shown in Figure 1, the system consists of seven function blocks, namely, the track selector, the feature extractor, the classifier, the profile manager, the recommendation module, the interface, and the database. When a new music object is inserted in the database of the MRS, it goes through two function blocks, i.e., track selector and feature extractor. According to the extracted features, the incoming music object is properly assigned to certain music group by the classifier function block. These three function blocks will be described in the following subsections. The profile manager and recommendation methods will be presented in Section 3. In addition, the interface will be shown in Section 4.

2.1 Track Selector

In the MRS, the music objects are of MIDI format. There are two kinds of music objects, i.e., monophonic music objects and polyphonic music objects. Usually, a polyphonic music object consists of several tracks, one for melody and the others for accompaniment. We observe that the track for melody contains much more distinct notes with different pitches than the tracks for accompaniment. In [Uitd98], the method used to extract a melody from a MIDI tile is developed, which considers all tracks and chooses the notes with the highest pitch for the melody. This method may result in an extracted melody containing the notes which belong to the tracks of accompaniment.

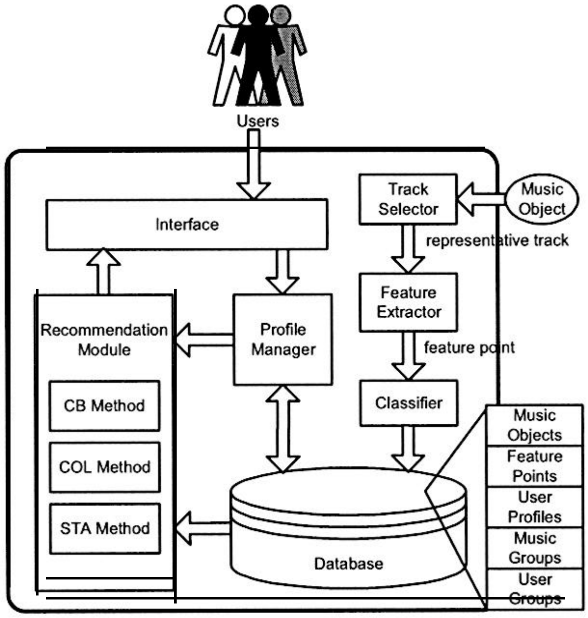

Figure 1. The system architecture of the MRS.

Different from the method used in [Uitd98], we use a measure of *pitch density* to select a ***representative track***. The pitch density of a track is defined as follows:

$$pitch\ density = \frac{NP}{AP} \tag{1}$$

where NP is the number of distinct pitches in the track

AP is the number of all distinct pitches in MIDI standard, i.e., 128.

Tbe track with the highest density is selected as the representative track of a polyphonic music object.

2.2 Feature Extractor

The purpose of the feature extractor is to extract features from the perceptual properties of the representative track. The six features are described as follows:

- **Mean (MP) and standard deviation (SP) of the pitch values**

  From the representative track, we compute the mean and standard deviation of the pitches.

- Pitch **density (PD)**

  The definition of pitch density has been given in equation (1).

- **Pitch entropy (PE)**

  The *pitch entropy PE*, derived from [Sayo00], is defined as follows:

$$PE = -\sum_{j=1}^{NP} P_j \log P_j \tag{2}$$

where P_j is defined as follows.

$$P_j = \frac{N_j}{T} \tag{3}$$

where Nj is the total number of notes with the corresponding pitch in the representative track
T is the total number of notes in the representative track

In equation (2), *PE* has a maximum value when each P_j *is* the same.

. Tempo degree (TD)

The *tempo degree is* defined as a ratio of the number of *fast measures* to the number of measures in the representative track. A measure is a fast measure if the average note duration in the measure is shorter than one.

. Loudness **(LD)**

The feature of *loudness* is defined as the average value of the *note velocities* which can be derived from MIDI tiles.

2.3 Classifier

After extracting features, each music object in the MRS is represented as a 6-tuple, (MP, SP, PD, PE, TD, LD). These features will be used to classify music objects into music groups. A music group is represented by a centroid and six thresholds, each for a feature. These thresholds are used to restrict the number of music groups, which can be decided by a method to be illustrated in Subsection 4.2.2. For easier illustration, we show an example of the classification “sing two features, PD and PE. For each incoming music object, there are two situations to consider when performing the classification. In the first situation, no music group exists in the database. Therefore, the new feature point of the incoming music object will be the centroid of a new group. In the second situation, some music groups exist. The distances between the new feature point and each group centroid are computed. The group with the minimum distance is selected for consideration. There are two cases to consider. I” case 1, if the new feature point falls into the rectangular area of the selected group formed by the centroid and the two thresholds δ_{PD} and δ_{PE} (for features PD and PE, respectively), the new feature point will be assigned to the selected group, as shown in Figure 2. The centroid of the group is recomputed accordingly. In case 2, if the new feature point does not fall into the area of any group, it will become the centroid of a new group, as show” in Figure 3. Note that the classifier “sing multiple features may produce too many groups. We “se a maximum number of groups K to limit the total number of music groups. If more than K groups are created, we enlarge the thresholds and re-classify the music objects such that the number of groups become less than or equal to K.

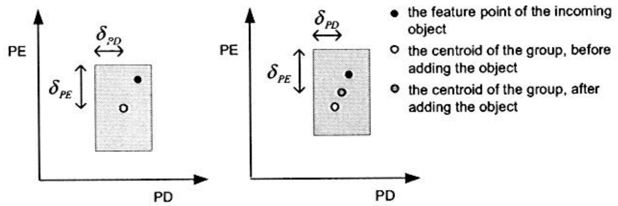

Figure 2. The new feature point falls into the area of a group (case 1).

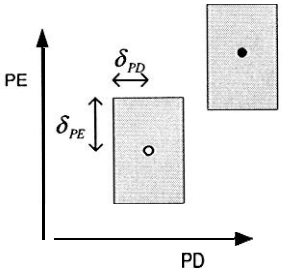

Figure 3. The new feature point does not fall into the **area** of any group (case 2).

3. Recommendation Mechanisms

In this section, the access history of the user in the MRS is tint introduced. We also present the three recommendation methods.

3.1 The Profile Manager

In the MRS, the profile manager is implemented for the purpose of updating the access histories of the users. 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. A” example of the access history is shown in Table 1.

Table 1. 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

As shown in Table 1, 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. The transaction is defined as the set of music objects accessed at the same time. Note that the transaction id is monotonically increasing.

3.2 The CB Method

Based on the content-based filtering approach, the purpose of the CB method is 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, we analyze the latest transactions in the access history as follows. In the following example, we only “se the latest five transactions for simplicity.

Each transaction is assigned a different weight, where the latest transaction has the highest weight. Moreover, the music group containing more accessed music objects in a transaction has a higher weight than other groups in the same transaction. The weight GW_i of music group G_i is computed as follows:

$$GW_i = \sum_{j=1}^{n} TW_j \ x \ MO_{j,i} \qquad (4)$$

where TWj *is the weight of transaction Tj*
n is the number of latest transactions used for analysis
MOj,i is the number of music objects which belong to music group Gi in transaction Tj

These weights will be recorded in a *preference table* for the user. After calculating the weight for each music group, the MRS ranks all the music groups. The music group with a greater weight takes a higher priority of recommendation. To avoid recommending a large number of music objects to users, the MRS limits the number of music objects for recommendation. According to the GW_i, different numbers of music objects from the music groups will be recommended. The number of music objects R_i from each music group is decided as follows:

$$R_i = \left\lceil N \times \frac{GW_i}{\sum_{k=1}^{M} GW_k} \right\rceil \qquad (5)$$

where N is the number of music objects in the recommendation list
G Wi is the weight of the target group
M is the total number of music groups in MRS

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 weight in the decreasing order. In the same music group, the latest music object will be first recommended.

Example 1. Take the user's access history shown in Table 1 as an example. Assign the weights 0.4096, 0.512, 0.64, 0.8, and 1 to T1, T2, T3, T4, and T5, respectively. Using equation (4), the weight for each music group is calculated, as shown in Table 2.

Table 2. The preference table for the user.

Music Group	Weight
A	3.08
B	2.5616
C	1.7216
D	1.312

According to Table 2, the total weight of all music groups is 8.6752. Suppose the number of music objects to be recommended is 20. By applying equation (5), the result is shown in Table 3.

Table 3. Number of music objects to he recommended in each group.

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

Take music group A for example. The latest eight music objects in music group A will be selected for recommendation. Note that $\sum_{i=1}^{M} G_i$ may be more than N (i.e. N=20 in our example), and we select the first N music objects for recommendation.

3.3 The COL Method

As described above, the recommendation of the CB method depends on the users' interests and the interests are derived from the users' access histories. Therefore, the users will never get a recommendation of the music objects belonging to the music groups they never accessed before. That is, the CB method tends to provide expected and interesting music objects for users. Based on the collaborative approach, the purpose of the COL method is to provide unexpected findings due to the information sharing between *relevant users.*

To refer to the information from other users, we group the users first. In the COL method, we applyt the technique proposed in [Wu01] for user grouping. The idea of the technique is to derive the profiles of user interests and behaviors from transactions in the access histories. Users with similar profiles of interests and behaviors will be identified as relevant users. In [Wu01], the *large-l itemsets* derived from transactions in the access history are used for user interests and the *large-2 itemsets* are used for user behaviors. Two examples are shown as follows. Example 2 shows the process of capturing user interests and Example 3 shows the process of capturing user behaviors.

Example 2. Suppose there are five transactions in the access history as shown in Table 4.

Table 4. The access history of a user.

Transaction	Music Groups in Transaction
T1	A, C, E
T2	B, C, E, F
T3	C, D, E, F
T4	B, C, D, F
T5	A, G

We construct the *interest table* from the access history of the corresponding user.

Table 5. The interest table.

Music Group	Count	First Transaction (FT)	Last Transaction (LT)
A	2	T1	T5
B	2	T2	T4

C	4	T1	T4
D	2	T3	T4
E	3	T1	T3
F	3	T2	T4
G	1	T5	T5

In this method, the *support* of a music group is calculated by the following formula:

$$support = \frac{count}{T_c - FT + 1} \quad (6)$$

where T_c *denotes the current transaction number*

Therefore, we compute the support for each group by equation (6). Suppose the T_c is 5. The support for each music group is shown below:

Table 6. The supports of the music groups.

Music Group	Support
A	0.4
B	0.5
C	0.8
D	0.67
F	0.75
G	1

If the minimum support is 75%, there will be three large-1 itemsets, *i.e.*, music groups C, F and G, which form the *interest profile* for the user.

Example 3. Take the access history shown in Table 4 for example. We construct the behavior table and compute the support of each music group pair as follows:

Table 7. The behavior table with the corresponding support.

Music Group Pair	Count	FT	LT	Support
AC	1	T1	T1	0.2
AE	1	T1	T1	0.2
AG	1	T5	T5	1
BC	2	T2	T4	0.5
BD	1	T4	T4	0.5
BE	1	T2	T2	0.25
BF	2	T2	T4	0.5
CD	2	T3	T4	0.67
CE	3	T1	T3	0.6
CF	3	T2	T4	0.75
DE	1	T3	T3	0.33
DF	2	T3	T4	0.67
EF	2	T2	T3	0.5

If the minimum support is 0.65, there will be four large-2 itemsets, *i.e.*, pairs AG, CD, CF and DF. Therefore, the *behavior profile* {AG, CD, CF, DF} is derived for the user. After deriving the interest profile and the behavior profile of the user, we construct an I-B matrix and transform it into an I-B vector. For example, if a user has interests {C, F, G} and behaviors {AG, CD, CF, DF}, the I-B matrix of the user is shown in Table 8.

Table 8. The I-B matrix.

	A	B	C	D	E	F	G
A	0	0	0	0	0	0	1
B		0	0	0	0	0	0
C			1	1	0	1	0
D				0	0	1	0
E					0	0	0
F						1	0
G							1

Then, we transform the I-B matrix to the I-B vector (0000001 000000 11010 0010 000 10 1). Therefore, each user has a corresponding I-B vector. According to the I-B vector, we compute the Euclidean distance between two users. Then, we apply the clustering algorithm to group users.

In the COL method, we capture user interests and behaviors from transactions in the user's access history by applying the technique proposed in [Wu01]. The users are then grouped based on their interests and behaviors. To make a recommendation for a user, the weights of each music group associated with the relevant users in the same group will be averaged. These average weights will be recorded in a *reference table* for the user. When the user chooses the COL method for recommendation, the recommendation module will compute the difference of the weights for each music group in the associated preference table and the reference table. According to the weight differences, the COL method recommends music objects to the user in a way similar to the CB method. Example 4 shows the process to construct a reference table and to make recommendation using the COL method. As in the CB method, the latest five transactions in each access history are considered.

Example 4. Suppose there are three persons UA, UB, and UC in user group U. Table 9 shows the partial access histories of UA, UB and UC. We omit access time and object id for clearer illustration.

Assign the weights 1, 0.8, 0.64, 0.512, and 0.4096 to the latest five transactions, respectively. We apply the equation (4) in the CB method to each access history. The result is shown in Table 10.

To make a recommendation for UA, the reference table for UA is constructed as shown in Table 11.

Table 11. The reference table for user UA.

Music Group	Weight
A	2.7568
B	2.332
C	1.9248
E	0.7048
F	0.6048

Table 9. The latest five transactions in the access histories of users UA, UB, and UC.

Partial access history of UA.

Music Group	Transaction
B	T8
C	T8
D	T9
C	T9
B	T9
A	T10
B	T10
A	T10
D	T11
C	T11
A	T11
B	T12
A	T12

Partial access history of UB.

Music Group	Transaction
E	T13
F	T13
A	T14
A	T14
B	T15
C	T15
A	T16
D	T16
B	T17
A	T17
B	T17
E	T17

Partial access history of UC.

Music Group	Transaction
A	T11
C	T11
B	T12
B	T12
A	T13
A	T13
D	T14
C	T14
F	T14
A	T15
C	T15
B	T15
D	T15
C	T15

Table 10. The preference tables for users UA, UB and UC.

Preference table of UA.

Music Group	Weight
A	3.08
B	2.5616
C	1.7216
D	1.312

Preference table of UB.

Music Group	Weight
A	2.824
B	2.64
C	0.64
D	0.8
E	1.4096
F	0.4096

Preference table of UC.

Music Group	Weight
A	2.6896
B	2.024
C	3.2096
D	1.8
F	0.8

The weight for each music group in the reference table is subtracted from that in the preference table, and the result is shown in Table 12.

Table 12. The table of weight differences.

Music Group	Weight Difference
A	-0.3232
B	-0.2296
C	0.2032
D	-0.012
E	0.7048
F	0.6048

In the COL method, the music group with zero or negative weight difference will not be recommended to the user. Therefore, we recommend music groups C, E, and F to UA. The numbers of music objects from music groups C, E, and F for recommendation are decided by equation (5), as shown in Table 13. Note that the M in equation (5) is set to 3 in this case.

Table 13. Number of music objects to be recommended in each group for user UA.

Music Group	Number of Recommended Music Objects
E	10
F	8

The order of the music objects to be recommended is decided by the same way as the CB method.

3.4 The STA Method

The third recommendation method is based on the statistics. We define ***the long-term hot music group*** as the music group containing the most music objects in the access histories of all users. Similarity, ***we*** define the ***short-term hot music group*** as the music group containing the most music objects in the latest five transactions in the access histories of all users. When the user chooses this recommendation method, the MRS recommends the latest N music objects (which have not been accessed by the user), half from the long-term hot music group and the other half from the short-term hot music group to the user.

4. Experiments

The implementation of the MRS is shown in Subsection 4.1. Moreover, the results of a series of experiments are shown and explained in Subsection 4.2.

4.1 Implementation

Figure 4 shows the user's access history. Figure 5 shows the recommendation result by applying the CB method. Figure 6 shows all operators in the MRS.

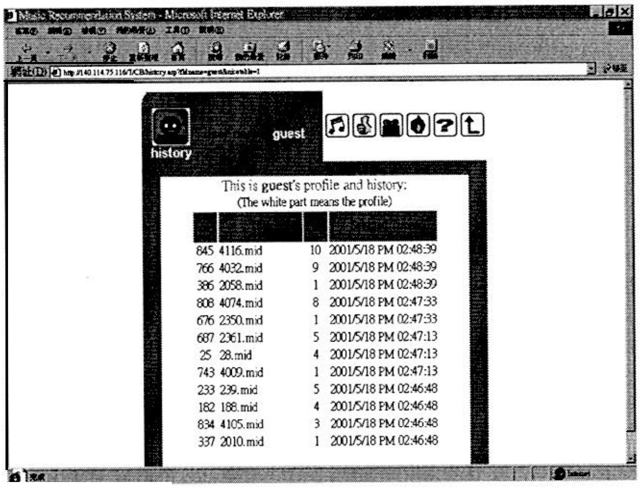

Figure 4. The access history.

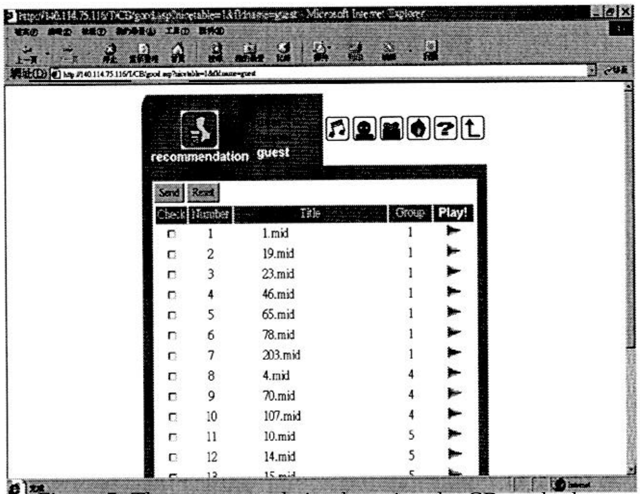

Figure 5. The recommendation by using the CB method.

4.2 Experiment results

In this subsection, we show the experiment results, including the effectiveness of the track selector, the effectiveness of the feature selection, and the quality of recommendations.

4.2.1 Effectiveness of the track selector

To evaluate the effectiveness of the track selection method, we ask an expert to select a representative track from each MIDI tile. Then, we apply our method on the same testing data set of 100 MIDI tiles. An 83% correctness rate is achieved by our method.

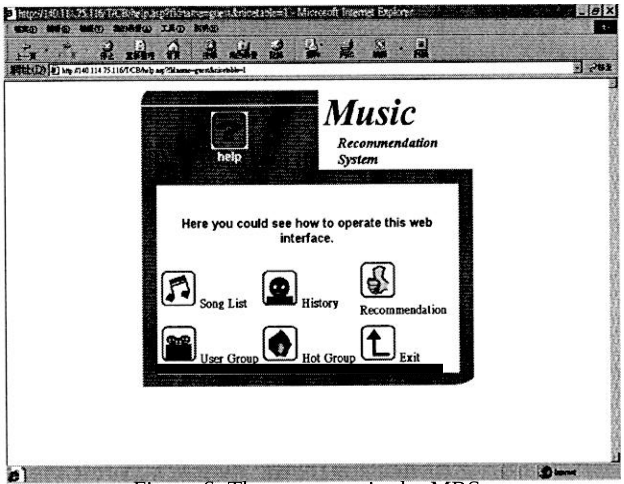

Figure 6. The operators in the MRS.

4.2.2 Effectiveness of the feature selection

The data set of 100 MIDI tiles is first classified by the expert into five groups, i.e., lyric music, jazz music, rock music, country music, and classical music. Then, we apply the K-means algorithm [Jain88] to classify the same data set on the six features, respectively. By comparing the results, the error rate is computed as follows:

$$\text{error rate} = \frac{\sum_{i=1}^{5} E_i}{100} \quad (7)$$

where Ei is the number of music objects mistakenly classified into group i

The error rate with respect to each feature is shown in Table 14

Table 14. The error rates by features.

Feature	MP	SP	PD	PE	TD	LD
Error Rate	65%	65%	60%	56%	66%	62%

According to the error rates shown in Table 14, we select the feature PE, which has the lowest error rate to be the representative feature from the perceptual property of pitch. To consider the influence of other perceptual properties on classification, we choose the features TD and LD. To compare with the error rates by single features, we apply the K-means algorithm to classify the same data set based on the four feature sets (PE, TD), (PE, LD), (TD, LD), and (PE, TD, LD). By using equation (7), the result of error rates is shown in Table IS.

Table 15. The error rates by feature sets.

Feature Sets	(PE, TD)	(PE, LD)	(TD, LD)	(PE, TD, LD)
Error Rate	43%	47%	39%	44%

The result shown in Table 15 indicates that it is better far the system to use multiple perceptual properties to represent a music object. According to the classification by using the K-means algorithm based on each feature, the mean of the distances between the feature points and the centroids of the associated groups is used as the threshold of the corresponding feature. This threshold is then used in the classifier of the MRS.

4.2.3 *Quality of recommendations*

We invite 10 students to perform the experiments. In each recommendation method, the MRS lists 20 music objects for users. The quality of the recommendation is measured by *precision* defined as follows:

$$precision = \frac{N_A}{N} \tag{8}$$

where N is the number of music objects in the recommendation list

N_A is the number of music objects which the user accesses in the recommendation list

The results are shown in Table 16 and Table 17. Note that we select 20 as N in the experiments.

Table 16. The precision of recommendation based on the classification by a single feature.

Single Feature	PE	TD	LD
The CB Method	37%	39%	35%
The COL Method	19%	23%	18%
The STA Method	24%	27%	21%

Table 17. The precision of recommendation based on the classification by feature sets.

Feature Sets	(PE, TD)	(PE, LD)	(TD, LD)
The CB Method	59%	51%	62%
The COL Method	17%	23%	22%
The STA Method	29%	31%	26%

For the CB method, the recommendation based on feature sets is better than the recommendation based on a single feature. The result indicates that it is better to use multiple features to represent a music object, which coincides with the result in Subsection 4.2.2. For the three recommendation methods, the precision of the CB method is better than the COL and the STA methods. The reason is that the CB method considers only private information of the user. On the contrary, the COL method tends to provide unexpected music objects for users, which may be interesting. In addition, the STA method provides hot music groups derived from all access histories. Therefore, the recommendation result of the STA method is better than that of the COL method.

5. Conclusion

In this paper, we present a music recommendation system to provide a personalized service of music recommendation. The music objects of MIDI format are analyzed. The method of track selection is proposed to determine the representative track. Based on the perceptual properties of music objects, six features are extracted from the representative track for the classification of music objects. Users are also classified into groups by the access histories. Three recommendation methods are proposed. We perform a series of experiments to show the effectiveness of the recommendation results.

It is time-consuming for users to browse all recommended music objects. In the future, we will research into the summarization of music objects. Moreover, other features and recommendation methods will be also investigated.

References

[Amaz] http://www.amazon.com.

[Bala97] Balabanovic, M. and Y. Shoham, "Fab: Content-based Collaborative Filtering Recommendation," *Communications of the* ACM, 40(3): 66-72, March 1997.

[Basu98] Basu, C., H. Haym, and W. W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," *Proceedings of National Conference on Artificial Intelligence, 1998.*

[Bill99] Billsus, D. and M. Pazzani, "A Hybrid User Model for News Story Classification," *Proceedings of International Conference on User Modeling,* June 1999.

[Gold92] Goldberg, D., D. Nichols, B. M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM,* 35(12): 61-70, 1992.

[Jain88] Jain, A. K. and R. C. Dubes, Algorithms for Clustering Data, 1 st Edition, Prentice Hall, 1988

[Krul96] Krulwich, B. and C. Burkey, "Learning User Information Interests through Extraction of Semantically Significant Phrases," *Proceedings of AAAI Spring Symposium on Machine Learning in Information Access,* 1996.

[Lang95] Lang K., "Newsweeder: Learning to Filter Netnews," *Proceedings of International Conference on Machine Learning,* 1995.

[Kons97] Konstan, J. A., B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Grouplens: Applying Collaborative Filtering to Usenet *News," Communications of the ACM,* 40(3): *77-87,* March 1997.

[Ruck97] Rucker J. and M. J. Polanco, "Personalized Navigation for the Web," *Communications of the ACM,* 40(3): *73-75,* March 1997.

[Sayo00] Sayood, K., Introduction to Data Compression, 2nd Edition, Morgan Kaufmann, 2000.

[Shar95] Sharadanand, U and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of mouth'," *Proceedings of CHI'95 Conference on Human Factors in Computing Systems,* 1995.

[Smit00] B. Smith and P. Cotter, "A Personalized Television Listings Service," *Communications of the ACM,* 43(8): 107-111, August 2000.

[Uitd98] Uitdenbogerd, A. L. and Justin Zobel, "Manipulation of Music For Melody Matching," *Proceedings of the sixth ACM international conference on Multimedia,* 1998, Pages 235 240

[Wu01] Wu, Y. H., Y. C. Chen, and A. L. P. Chen, "Enabling Personalized Recommendation on the Web Based on User Interests and Behaviors," *Proceedings of IEEE International Workshop on Research Issues in Data Engineering (RIDE),* 2001.

[Yaho] http://www.yahoo.com.

Acknowledgement

This work was partially supported by Ministry of Education, Republic of China under Contract Number MOE 89-E-FA04-1-4 and InterVideo, Inc.