A music recommender based on audio features

Conference Paper · January 2004

DOI: 10.1145/1008992.1009106 · Source: DBLP

Qing Li
Southwestern University of Finance and Economics

Byeong Man Kim
Kumoh National Institute of Technology

A Music Recommender Based on Audio Features

Qing Li, Byeong Man Kim, Dong Hai Guan, Duk whan Oh
Kumoh National Institute of Technology
188 Shinpyung-Dong, Kumi, Kyungpook, 730-701, South Korea
{liqing, bmkim, eastsea, dhoh}@se.kumoh.ac.kr

ABSTRACT
Many collaborative music recommender systems (*CMRS*) have succeeded in capturing the similarity among users or items based on ratings, however they have rarely considered about the available information from the multimedia such as genres, let alone audio features from the media stream. Such information is valuable and can be used to solve several problems in RS. In this paper, we design a *CMRS* based on audio features of the multimedia stream. In the *CMRS*, we provide recommendation service by our proposed method where a clustering technique is used to integrate the audio features of music into the collaborative filtering (*CF*) framework in hopes of achieving better performance. Experiments are carried out to demonstrate that our approach is feasible.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval –*information filtering.*

General Terms
Algorithms, Experimentation.

Keywords
Music recommender, information filtering, collaborative filtering

1. INTRODUCTION
Most popular recommendation systems are based on the *CF* technique, where target users apply peer opinions to predict their interests. However, they do not consider the fact that the content information of items can help provide good services. Therefore content-based collaborative recommenders have been developed to use the content information of items. However, most of them extract information from textual objects, such as textual contents of webpages. As for multimedia, they usually use the textual descriptors of multimedia annotated by producers instead of content information, due to inconvenience of extracting accurate information from multimedia. In this paper, we design a *CMRS* based on audio features of the multimedia stream in order to solve three challenges in CF, that is non-transitive association, user bias and cold start.

As for the CF, there are two ways to calculate the similarity for clique recommendation, one is item-based way and the other is user-based way. Badrul [1] has proved that the item-based CF is better than the user-based CF on precision and computation complexity. Thus we adopt former method in our *CMRS* to solve the three challenges.

The first challenge is non-transitive association for items. That is, if two similar items have never been rated or selected by the same user. The relationship between them is lost. Therefore those two items can not be classified into the same community by a pure item-based CF. Similar items in different groups will definitely affect the quality of recommendation service negatively.

SIGIR'04, July 25-29, 2004, Sheffield, South Yorkshire, UK.
ACM 1-58113-881-4/04/0007

The second challenge is user bias from historical ratings. For instance, the item-based recommender cannot distinguish two items which have the same historical ratings of users but different content attributes. As Table 1 shows, music 3 and 4 have the same historical ratings as music 1 and 2 made by user 2 and 3. According to the item-based *CF*, music 3 and 4 have the same opportunity to be recommended to user 1 by the system. However, if music 1, 2 and 3 belong to rock music and music 4 belongs to country music, it is obvious that music 3 should have the privilege to be delivered to user 1, because user 1 prefers rock music which can be inferred from historical ratings.

The third one is the cold start problem. It is hard for a pure CF to recommend new items. Because no user made any ratings on the new, the new can not find their own communities where items have the similar ratings from users. The same to new users, however, it can be partially solved by the gauge set.

Table 1. Rating information

Genre / Music	Rock	Country		User / Music	1	2	3
1	98%	2%		1	5	4	3
2	90%	10%		2	4	4	3
3	98%	2%		3		4	3
4	2%	98%		4		4	3
5	98%	2%		5			

2. FEATURE EXTRACTION
The audio features extracted should express some aspects of audio media contents, such as genre and rhythm, which can be used to solve the above problems and improve the recommendation service. Also it reduces the burden of manual genre annotation, and further complements the genre information by providing rhythm information, which can not be obtained if only the textual descriptions of music are used. Two feature sets for representing timbral texture and rhythmic content are proposed, and the performance and relative importance of the proposed features are investigated by our system using real-world MPEG audio corpus.

A. Timbral Texture Features

The features used to represent timbral texture are based on the standard features proposed for music-speech discrimination [2].

1) Spectral Centroid: It is the balancing point of the subband energy distribution and determines the frequency area around which most of the signal energy concentrates.

2) Spectral Flux: It determines changes of spectral energy distribution of two successive windows.

3) Spectral rolloff point: It is used to distinguish voice speech from unvoiced music, which has a higher rolloff point because their power is better distributed over the subband range.

4) Sum of scale factor (SSF): the loudness distribution for whole duration.

5) MFCC (Mel Frequency Cepstral Coefficients): They are set of perceptually motivated features and provide a compact representation of the spectral envelope such that most of the signal energy is concentrated in the first coefficients.

The feature vector for describing timbral texture consists of the following features: means and variances of spectral centroid, rolloff, flux, SSF and first five MFCC coefficients over the texture window.

B. Rhythmic Content Features

The feature set for representing rhythm structure is based on detecting the most salient periodicities of the signal. Gorge's method [2] is applied to construct the beat histogram, from which six features are calculated to represent rhythmic content.

A0, A1: relative amplitude of the first and second histogram peak;
RA: ratio of the amplitude of the second peak divided by the amplitude of the first peak;
P1, P2: period of the first, second peak in bpm;
SUM: overall sum of the histogram (indication of beat strength).

Table 2. Item rating matrix & group rating matrix

Item rating matrix					Group rating matrix		
Music \ User	1	2	3		Music \ Group	G 1	G 2
1	5		1		1	98%	4%
2		4			2	95%	5%
3					3	15%	96%

3. RECOMMENDATION MECHANISM

To solve the three challenges addressed in the introduction part, we have designed a new recommendation mechanism called ICHM (item-based clustering Hybrid method) to integrate the audio features for recommendation services. The details of our approach can be referred to the literature [2].

Let us see how our mechanism solves the three challenges. As for item non-transitive association, in Table 2, we can not make predictions for music 1 or 2 using a pure *CF*. Whereas, our *CMRS* puts up a reasonable solution by considering the group-rating matrix which can provide the relationship among the music based on the audio features. For instance, we can not get the similarity between item 1 and 2 in Table 2 if we only consider the item-rating matrix. However, with the help of group-rating matrix, we can easily know that the similarity between item 1 and 2 is about 1. Therefore we can get the predictive rating of Oliver on item 1 by using a weighted average of deviations from the neighbor's mean [2]. According to the calculation, we can recommend item 1 to Oliver, because the predictive value is 4, which means Oliver might interest on it.

The second case illustrates how our *CMRS* deals with the user bias from the historical ratings. As we know, the predictions of user 1 on music 3 and 4 are the same in Table 1 by a pure CF. With the help of group-rating matrix in our *CMRS*, the prediction of user 1 on item 3 is calculated as 4.6 and on item 4 is 4.0, thus music 3 has a privilege to be delivered to user 1.

In addition, if music 5 is added into the database as Table 1 shows, with the help of group-rating matrix, we can find that user 1 shows more interest on this new piece of music than others. So to speak, the new item problem is also done.

4. EXPERIMENTAL EVALUATION

Experiments are carried out on a real-word music corpus, which has 240 pieces of music and 433 users with 1,150 ratings. The music is stored as 22050Hz, 16-bit, mono MP3 audio files and full music is used to extract the audio features. We have used an *Allbut1* protocol to evaluate the obtained prediction accuracies. This means we randomly leave out exactly one rating for every user who at least posses two ratings. MAE (Mean Absolute Error) [1] is used as metrics.

In order to observe the performance and relative importance of the proposed features, we construct the group rating matrix by combining the audio features differently. The first combination (C1) is first four aspects of timbral texture (8 dimensions), the second combination (C2) is all features of timbral texture (18 dimensions), the third combination (C3) is all features of rhythmic content (6 dimensions) and the last combination (C4) is all features of both timbral texture and rhythmic content. From Figure 1, it can be observed that there is a little difference between C1 and C2. With the help of rhythmic features (C3), the performance is also improved but not as good as timbral texture features do. However, using the rhythmic features and timbral texture features together can further improve the performance as C4 shows in Figure 1.

Comparing with the item-based collaborative recommender (IC), as Figure 1 shows, our *CMRS* can provide more ideal recommendation service. Moreover, from Figure 1, we can observe that audio features extraction is efficient, and the performance of audio features is no less than the content information provided by the textural descriptor of music genre.

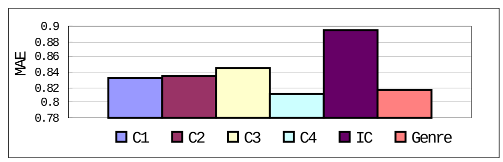

Figure 1. Comparison

As for cold start problem, we randomly select the number of music objects from 5 to 25 with the step of 5, delete all the ratings of those music objects and treat them as new items. We can observe from Table 3 that our *CMRS* deals reasonably with the new music objects.

Table 3. MAE of new music objects

MAE	5	15	20	25
New Music objects	0.885	0.891	0.892	0.891
Music objects excludes new ones	0.834	0.823	0.831	0.83

5. CONCLUSION

In this paper, we design a *CMRS*, which applies clustering technique to integrate the audio features to make recommendation. Our work indicates that the correct application of audio features can provide a good music recommendation service.

6. REFERENCES

[1] Sarwar, B. M., Karypis, G., Konstan, J. A. Item-based Collaborative Filtering Recommendation Algorithms. *In Proc. Tenth Int. WWW Conf.* pp. 285-295, 2001.

[2] Q. Li, B.M. Kim.Clustering Approach for Hybrid Recommender System, In *Proc. of IEEE/WI, Canada*. pp. 31-37, 2003.

[3] George Tzanetakis and Perry Cook. Music Genre Classification of Audio Signals. *IEEE TSAP, Vol.10 No.5,* 2002.

[4] J. S. Breese, D. Heckerman, and C. Kardie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *In Proc. Of UAI '98,* pp.43-52, 1998.