ISSN 2278 – 0211 (Online)
ISSN 2278 – 7631 (Print)

User Based Collaborative Filtering for Music Recommendation System

M. Sunitha Reddy
Department Of CSE, Vasavi College Of Engineering, Ibrahimbagh, Hyderabad, A.P., India
Dr. T. Adilakshmi
Department Of CSE, Vasavi College Of Engineering, Ibrahimbagh, Hyderabad, A.P., India
M. Akhila
Department Of CSE, Vasavi College Of Engineering, Ibrahimbagh, Hyderabad, A.P., India

Abstract:
Recommender systems have been proven to be valuable means for web online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce. The recommendations provided are aimed at supporting their users in various decision making process, such as what items to buy. In this paper we recommend items to users based on their logs. First we use collaborative filtering method to identify the users who are similar based on their listening history. Then recommend the items to new users based on the user clusters formed. At last we have evaluated the performance of the algorithm and propose the ideas that improve the recommendations.

Key words: *collaborative filtering, recommender system, clustering*

1. Introduction

With the expansion of the Internet and the advent of smart phones, people now are able to get easy access to and extend their activities on the Web. The domestic Internet usage rate reached 78%, and the number of Smartphone users totaled 30 million in 2012 [1]. Under these environmental changes, today's users have an enormous amount of information within their reach and can see it increase exponentially. In the flood of such information, they want to have a means to search for desired information easily and quickly. In attempts to achieve this, portals have created and posted a list of search words that have been most sought, while shopping malls have pro- vided a lineup of hot-selling items. Users want to search for their required information in the flood of information easily and quickly. To cope with these demands, portals have created and posted a list of search words that have been most sought, and shops have provided information on most sought-after products in their catalogues. However, these methods do not seem to offer help with the user's decision-making due to their inability to reflect the characteristics of individual users. Personalized services mean providing consumers with products and services most suitable to individual tastes based on their personal information, but without explicit questions about their desired products and services [2]. Among personalized services, recommender systems make recommendations of services or products that target customers might like. A wide range of recommender systems have been developed and also widely adopted by such Internet shopping malls as Amazon and CD Now [3]. Among these, collaborative filtering systems have been known to be the most successful method, and have found a variety of applications in Web pages, movies, thesis, and newspaper articles [4-6].

Collaborative filtering, which is the most widely used technique in music recommendation systems, is a method of making automatic recommendations of certain items by creating profiles based on diverse kinds of information collected from multiple users, and subsequently making predictions based on these profiles, about the interests of a user who has preferences similar to other like-minded users. In a collaborative filtering system, users give preference ratings to items based on their taste. After that, the system calculates preference similarities among users from such ratings, and makes predictions about a user's rating for a product which the user has not rated yet. A collaborative filtering system is designed to analyze a user' ratings given on the same product and predicts the user's rating for a product which the user has not rated yet. Accordingly, many users are required to make ratings on many items to come up with better recommendations. However, users are normally unable to assess all the items in the system, which always presents a fundamental problem named 'rating sparsity' to the collaborative filtering systems [7].

Music recommender systems are decision support tools that help tame the information overload by recommending only the items that are estimated as relevant to the user, based on the user's music preferences [8]. For example, Last.fm [1] a popular Internet radio and

recommender system that allows a user to mark songs or artists as favorites, and based on this information can identify and recommend music content that is likely to be of interest to the user.
This paper uses user listening history for collaborative filtering system based on user clusters in music recommendation systems. The rest of the paper is organized as follows. In Section 2 we explained fundamentals of the recommendation systems. Section 3 describes about the proposed algorithm for recommendation and evaluation measures used in the proposed system. Experimental setup, Data set used in experiment and results are discussed in Section 4. Conclusion and future scope is explained in Section 5.

2. Related Works

This section describes collaborative filtering system, such details as recommender systems, personalization techniques, and methods of selecting collaborative filtering systems that are required for the recommendation of music.

2.1. Recommender System

A recommender system makes recommendation of products that are suitable for a customer's demands, based on the analysis of such information as products that many customers are interested in, demographic data, and past purchasing activity [8]. Personalized services tailored to individual tastes have been emphasized in e-commerce transactions. Personalization means the process of quickly responding on the Internet to a customer's needs that are unique and specific. Web personalization is defined as activity made on the Internet by an individual in response to his/her interests or tastes [9]. The reasons why personalized service is important are that customers can reduce their attempts to search for products, and companies not only increase customers' loyalty to their e-commerce sites through the recommendation of proper products but also build attachment between them and their customers [10].

2.2. Personalization Techniques

The personalization techniques for recommender systems include:

2.2.1. Content-Based Recommender System

This system analyzes item information and recommends certain items to users. It is suitable for recommendation of such items as texts, documents, news, and web pages whose contents are abundant and easy to analyze [14].

2.2.2. Rule-Based Filtering

This technique specializes in the acquisition of users' information profiles by means of questions to users about their interests and preferences. Users' profiles can be obtained by asking questions about the users' tastes and preferences on particulars and collecting and analyzing their answers. This filtering system recommends to users or provides them with information about products that are considered to be suitable given a user's psychological and preference information based on such profiles [10].

2.2.3. Demographic Filtering

This system makes recommendations using users' information such as age, sex, and education level [12]. Demographic attributes have an advantage of making an easy analysis of users' preferences regarding various kinds of items and item categories.

2.2.4. Collaborative Filtering

This system makes recommendations by utilizing each user's assessment information [13]. As a collaborative filtering system makes use of rating information, it has an advantage of performing recommendations without the information on a user or on a specific item.

2.2.5. Learning Agent-Based Filtering System

This personalization technique utilizes learning agents that are designed to trace users' at- tributes, habits, and personal preferences through the analysis of log files including records of visit to websites and their frequency, access location, and time [11].
A recommender system is a program which makes predictions about relations among customers or among items, and searches for items that a user may be expectedly to desire. The purpose of many studies on recommender systems have mainly focused on their capability of how likely they are able to recommend products that a customer is satisfied with. Collaborative filtering is the method most frequently used to categorize similarities among items.

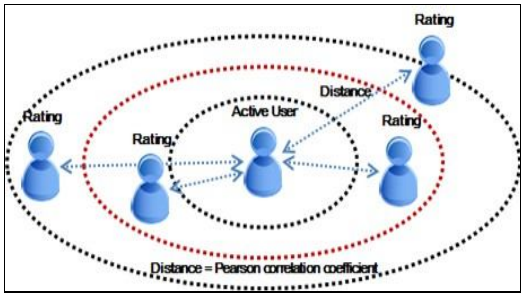

Figure 1: Collaborative Filtering Algorithm

Fig. 1 shows a neighborhood-based algorithm that has generally been used in collaborative filtering systems [7]. The active user calculates distances to other users and selects as its neighbors the number of users who are located at nearest distances. The distance between users can be calculated using the Pearson correlation coefficient, the mean-square difference, or vector similarity. In [13], the Pearson correlation coefficient produced a better result than the vector similarity, and [19] showed that the Pearson correlation coefficient brought about a better outcome though a selection of either a too small or too large number of neighbors might lead to a reduction in its prediction capability. When distances to other users have been calculated, a predicted score for an item can be computed by summing other users' rated scores in proportion to their distance weights, using the following equation [7].

$$P_{a,i} = r_a + \frac{\sum_{u=1}^{n}\left(r_{u,i} - \bar{r}_u\right) * w_{a,u}}{\sum_{n=1}^{n} w_{a,u}}$$

Equation is introduced to calculate a distance through the Pearson correlation coefficient. '$P_{a,i}$' indicates an active user a's prediction about an item 'i'. 'n' is the number of the neighboring users, '$r_{u,i}$' means the rating of a user 'u' on an item 'i', and '$w_{a,u}$' is defined as the weighted similarity between the active user 'a' and its neighbor 'u' [5].

The Pearson correlation coefficient has a number close to '1' when a user A rates a movie high that a user B has also rated high and user A also gives a low rating to a movie that the user B has given a low rating; and, it is close to '0' when the vice versa holds true.

2.3. Approaches To Collaborative Filtering

There are two kinds of collaborative filtering: user-based collaborative filtering and item- based collaborative filtering.

2.3.1. User-Based Collaborative Filtering

This approach is to calculate distances to quantify how closely two users match each other in respect with a certain common item. For example, if user1 and user2 put in same ratings in the same item, the distance will be 0. On the other hand, assuming they give different ratings, the distance will be farther depending on the difference.

2.3.2. Item-Based Collaborative Filtering

Most recommender systems utilize an item-based collaborative filtering technique rather than a user-based one. For instance, when users who like item1 also like item2, the distance between two items is regarded as being close.

3. Proposed Approach for Recommendations

This section describes about the similarity measures used, forming the clusters of similar users, recommendation of items to new users and evaluation measures.

3.1. Similarity Measures

3.1.1. Cosine Similarity Measure

Similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a Cosine similarity of 1, two vectors at 90° have a similarity of 0, and tow vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0,1].

The technique is also used to compare documents in text mining. In addition, it is used to measure cohesion within clusters in the field of mining. One of the reasons for the popularity of Cosine similarity is that it is very efficient to evaluate, especially for sparse vectors, as only the non-zero dimensions need to be considered.

The cosine of two vectors can be derived by using the Euclidean dot product formula:

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\theta$$

Given two vectors of attributes, A and B, the cosine similarity, cos(θ), is represented using a dot product and magnitude as

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity. For text matching, the attribute vectors A and B are usually the term frequency vectors of the documents. The cosine similarity can be seen as a method of normalizing document length during comparison.
In the case of information retrieval, the cosine similarity of two documents will range from 0 to 1, since the term frequencies (tf-idf weights) cannot be negative. The angle between two term frequency vectors cannot be greater than 90°.

3.2. Formation Of Clusters By Using K-Means Algorithm
Once the similarity between the users is found by using cosine similarity measure, the next step is to form the user clusters based on this similarity measure. We used a threshold value and based on the value users are clustered into different clusters. The following is the Algorithm used to form user clusters.

Algorithm Threshold_Kmeans()
Begin
Initialize the threshold value to th_cutoff
For each user in u_1, u_2 ……………u_n
Put u_1 into C_1 cluster and find the similarity with u_2
Put u_2 into C_1 if the similarity is within the similarity threshold th_cutoff
Otherwise create a new cluster C_2
Repeat this for all users and all Clusters
Return the clusters C_1, C_2 ………… C_k
End

3.3. Recommendation
After getting the user clusters, we used these clusters to recommend items to new users. Use the following Algorithm for recommendations
Algorithm Recommendation ()
Begin
For each new user
Find the similarity with each cluster mean
Find the cluster with highest similarity
Then recommend the items preferred by the users in the cluster
End

3.4. Evaluation Measures
Many methods have been proposed for assessing the accuracy of collaborative filtering methods. We have used mean Average Precision (mAP) as the measure.
The mAP metric emphasizes the top recommendations, and is commonly used throughout the information retrieval literature. For any k, the precision-at-k (Pk) is the proportion of correct recommendations within the top-k of the predicted ranking:

$$P_k(u,y) = 1/k \sum_{j=1}^{k} M_{u,y}(j)$$

for each user, we now take the average precision at each recall point:
r

$$AP(u,y)= 1/n_u \sum_{k=1} P_k(u,y) * M_{u,y}(k)$$

where n_u is the number of positively associated songs for user u. Finally, averaging over all m users, we have the mean average precision:

$$mAP = 1/m \sum_{u} AP(u, y_u)$$

where y_u is the ranking predicted for user u.

4. Experiment and Results

This section describes about the Dataset used for experiment, experimental set up and results.

4.1. Data set

Million Song Dataset (MSD) a freely-available collection of audio features and meta-data for a million con- temporary popular music tracks [7]. Comprising several complementary datasets that are linked to the same set of songs, the MSD contains extensive meta-data, audio features, tags on the artist- and song-level, lyrics, cover songs, similar artists, and similar songs. It consists of four datasets namely Last.fm, Second hand data set, Musixmatch and Taste profile data set. We used taste profile Data set for our experiment

4.1.1. Taste Profiles

The collection of data we use is known as the Taste Profile [15] Subset. It consists of more than 48 million triplets (*user*, *song*, *count*) gathered from user listening histories. The data was provided by an undisclosed set of applications, where each user could select the song they wanted to listen to. The data consists of approximately 1.2 million users, and covers more than 380,000 songs in MSD. A raw sample of the data is shown in Fig 2.

User ID	**Song ID**	**Play Count**
b80344d063b5ccb3..	SOYHEPA12A8C13097	8
b80344d063b5ccb3..	SOYYWMD12A58A7BCC	1
b80344d063b5ccb3..	SOZGCUB12A8C13399	1
b80344d063b5ccb3..	SOZOBWN12A8C130999	1
b80344d063b5ccb3..	SOZZHXI12A8C13BF7D	1
85c1f87fea955d09...	SOACWYB12AF729E581	2
85c1f87fea955d09...	SOAUSXX12A8C136188	1
85c1f87fea955d09...	SOBVAHM12A8C13C4CB	1
85c1f87fea955d09...	SODJTHN12AF72A8FCD	2

Figure 2: A few lines of the raw data of the Taste Profile Subset http://labrosa.ee.columbia.edu/millionsong/tasteprofile
The three columns are user ID, song ID and play count. The user ID's have been truncated for visualization purposes.

4.2. Experimental Setup

We have taken 10000 records from Taste profile data set for experiment. It consists of 198 unique users and 7453 unique items. We have taken only those users who listened at least 30 songs and those songs which are listened by at least two users. With these constraints we got 98 unique users and 254 unique songs as shown in Fig 3.. We formed clusters by taking 60 users as training data and 38 users as test data.

User Id/ Song Id	**S_1**	**S_2**	**S_3**	**S_4**	**S_5**	**S_6**	**S_7**	**S_8**
U_1	2	5	0	5	11	0	0	0
U_2	0	1	0	0	4	3	0	2
U_3	0	0	0	0	2	0	0	0
U_4	0	2	5	0	7	2	4	4
U_5	1	3	0	0	0	0	0	0
U_6	5	0	3	0	0	3	0	0

Figure 3: Part of User- Item Matrix with Rows as Users, Columns as Songs

4.3. Results

We have done the experiment with various values of thresholds such 0.2, 0.25 and so on till 0.9.

We plotted the graph for threshold vs mAP and threshold vs no. of clusters. We can conclude from that as the threshold value increases the mAP also increases and number of clusters increases.

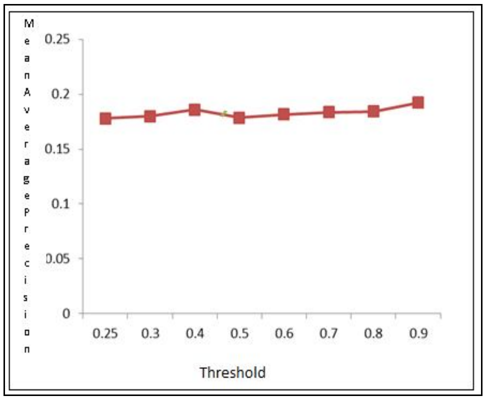

Figure 4

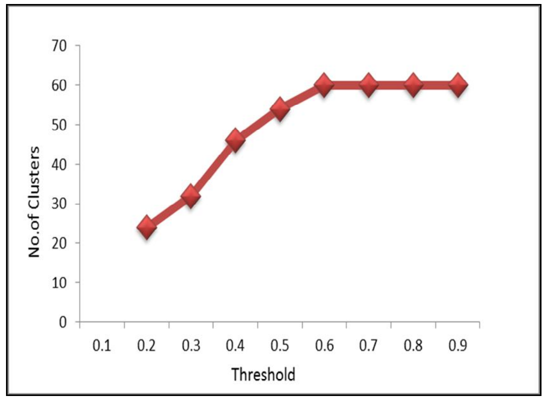

Figure 5

5. Conclusion and Future Scope

We have discussed about the user based collaborative filtering method for music recommendation system. This system is taking the user interest into consideration without taking the user feedback explicitly. We also evaluated our system on benchmark dataset. This work can be extended for recommendations by taking the time at whish user listens a particular item also into consideration.

6. References

1. 2011 Internet Use Survey Summary Report, KISA, Nov. 2011 pp.2-27.
2. M. D. Mulvenna, S. S. Anand, A. G. Büchner, "Personalization on the Net using Web mining," Communications of the ACM, Vol.43 Issue 8, Aug. 2000, pp.122-125.
3. J. Lee, S. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering," Journal of Intelligence Information Systems, Vol.13, No.4, 2007, pp.65-78.
4. W. Hill, L. Stead, M. Rosenstein and G. Furnas, "Recommending and evaluating choices in a virtual community of use," Proceeding of the SIGCHI conference on Human factors in computing systems, 1995, pp.194-201.
5. P. Resnick, N. lacovou, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of Netnews," Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill, NC, 2001, pp.175-186.
6. U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, pp.210-217.
7. H. Lee, H. Kim, "Improving Collaborative Filtering with Rating Prediction Based on Taste Space," Journal of Korean Institute of Information Scientists and Engineers, Vol.34, No.5, 2007, pp.389-395.
8. B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algo- rithms," Proceeding of the 10Th international conference on World Wide Web, 2001, pp.285-295.
9. B. Mobasher, H. Dai, T. Luo, and M.Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization," Data Mining and Knowledge Discovery, Vol.6, No.1, 2002, pp.61- 82.
10. T. Hong, H. Lee, B.Su, "A Web Personalized Recommender Systems Using Clustering - based CBR", Journal of Intelligence Information Systems, Vol.11, No.1, 2005, pp.107-121.
11. R. J. Mooney and L. Roy, "Content-based Book Recommending Using Learning for Text Categoriza- tion," In Proceedings of ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evalua- tion, 1999.
12. J. S. Breese, D. Heckeman and C. Kadie, "Empirical analysis of predictive algorithms for collabora- tive filtering," Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelli- gence, 1998, pp.43-52.
13. J. L. Herlocker, J. A. Konstan, A. Borchers, and J, Riedl, "An algorithmic framework for performing collaborative filtering," Proceedings of the 22nd annual international ACM SIGIR conference on Re- search and development in information retrieval, 1999, pp.230-237.
14. J. Heo, D. Park, Y. Jeong, "A Personalized Movie Recommendation System Based on the Personal Propensity and Collaborative Filtering," Proceedings of Korea Multi Media Society Spring Annual Conference, Vol.12, No.2, 2009, pp.325-326.
15. http://labrosa.ee.columbia.edu/millionsong/tasteprofile.