ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

Recommendation Tool Using Collaborative Filtering

Aditya Mandhare[1], Soniya Nemade[2], M.Kiruthika[3]
Student, Computer Engineering Department, FCRIT, Vashi, India [1]
Student, Computer Engineering Department, FCRIT, Vashi, India [2]
Co-ordinator, Computer Engineering Department, FCRIT, Vashi, India [3]

Abstract: With an increasing need to improve users' experience on the website, recommendation tools have become an indispensable resource for any e-commerce or social networking website. These tools analyze users' behaviour on website and their usage pattern, and thus try to predict a user's likes and dislikes. This is done by employing recommendation algorithms on the available usage data. The results obtained are then compared at different operational thresholds. The recommendations obtained as an output of these tools are then provided to the users, which helps them take better decisions or make better choices

Keywords: Database, collaborative filtering, recommendation system, Item based filtering, person based filtering, Data mining, Knowledge mining.

I. INTRODUCTION

There has been an explosive growth in the information available over the internet. World Wide Web has become a platform to store, propagate and retrieve as well as for mining utile knowledge. Web data is huge, diverse, dynamic and unstructured in nature. Due to this web data research has encountered a lot of challenges, such as scalability, multimedia and temporal issues etc. When a user interacts with the Web, there is a wide diversity of user's navigational preference, which results in needing different contents and presentations of information. Recent times have seen a rise in the interest of analyzing users' behaviour. This is due to the realization that quality service can be provided to the users not merely by providing huge amount of data, but by providing relevant data at right time and in suitable form which is easy to access and use.

To improve the Internet service quality and increase the user click rate on a specific website, thus, it is necessary for a Web developer or designer to know what the user really wants to do, predict which pages the user is potentially interested in, and present the customized web pages to the user by learning user navigational pattern knowledge.

Recommendation systems have been a popular topic of research ever since the ubiquity of the web made it clear that people of hugely varying backgrounds would be able to access and query the same underlying data. The initial human–computer interaction challenge has been made even more challenging by the observation that customized services require sophisticated data structures and well thought-out architectures to be able to scale up to thousands of users and beyond. In recent years, recommendation agents are extensively adopted by both research and e-commerce recommendation systems in order to provide an intelligent mechanism to filter out the excess of information available and to provide customers with the prospect to effortlessly find out items that they will probably like according to their logged history of prior transactions.

Recommendation Tool needs to employ efficient prediction algorithms so as to provide accurate recommendations to users. If a prediction is defined as a value that expresses the predicted likelihood that a user will ''like'' an item, then a recommendation is defined as the list of n items with respect to the top-n predictions from the set of items available. Improved prediction algorithms indicate better recommendations. This explains the essentiality of exploring and understanding the broad characteristics and potentials of prediction algorithms and the reason why this work concentrates on this research direction.

A. Methods used by recommendation algorithms

There are generally two methods to formulate recommendations both depending on the type of items to be recommended, as well as, on the way that user models are constructed. The two different approaches are content-based and collaborative filtering, while additional hybrid techniques have been proposed as well

1. Content based recommendation algorithms:

They are mainly used to recommend documents, web pages, news etc. to the users. The tool maintains information about user preferences either by initial input about user's interests during the registration process or by rating documents. The content of documents is taken into account and recommendations are formed by filtering in the ones that better match the user's preferences and logged profile.

2. Collaborative filtering based recommendation algorithms:

Collaborative-filtering algorithms aim to identify users that have relevant interests and preferences by calculating similarities and dissimilarities between user profiles. The

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014
ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

idea behind this method is that, it may be of benefit to one's search for information to consult the behaviour of other users who share the same or relevant interests and whose opinion can be trusted.

B. Challenges of recommendation algorithms

The challenges for recommendation algorithms expand to three key dimensions, identified as sparsity, scalability and cold-start.

1. Sparsity: Even users that are very active, result in rating just a few of the total number of items available in a database. As the majority of the recommendation algorithms are based on similarity measures computed over the co-rated set of items, large levels of sparsity can be detrimental to recommendation agents.

2. Scalability: Recommendation algorithms seem to be efficient in filtering in items that are interesting to users. However, they require computations that are very expensive and grow non-linearly with the number of users and items in a database. Therefore, in order to bring recommendation algorithms successfully on the web, and succeed in providing recommendations with acceptable delay, sophisticated data structures and advanced, scalable architectures are required.

3. Cold-start: An item cannot be recommended unless it has been rated by a substantial number of users. This problem applies to new and obscure items and is particularly detrimental to users with eclectic taste. Likewise, a new user has to rate a sufficient number of items before the recommendation algorithm be able to provide reliable and accurate recommendations.

II. TYPES OF COLLABORATIVE FILTERING ALGORITHMS

A . Person-to-Item Collaborative Filtering

Collaborative Filtering (CF) is the most successful recommendation technique to date. The basic idea of CF-based algorithms is to provide item recommendations or predictions based on the opinions of other like-minded users. The opinions of users can be obtained explicitly from the users or by using some implicit measures. CF algorithms represent the entire m x n user-item data as a ratings matrix, A. Each entry ai,j in A represents the preference score (ratings) of the ith user on the jth item. Each individual ratings is within a numerical scale and it can as well be 0 indicating that the user has not yet rated that item.CF approaches assume that those who agreed in the past tend to agree again in the future. For example, a collaborative filtering or recommendation system for music tastes could make predictions about which music a user should like given a partial list of that user's tastes (likes or dislikes). CF methods have two important steps, (1) CF collects taste information from many users, and this is collaborating phase. (2) Using information gleaned from many users' predictions and recommendation of users interest were automatically generated, and this is filtering phase. Assume we have m persons and n items which can be arranged in a M*N matrix where each row is a person, each column is an item. Thus if a person i likes item j then M(i,j)=1,otherwise M(i,j)=0. Clustering methods like k-means are used. Finally the missing values are taken from the cluster representation.

B. Item-to-Item Collaborative Filtering

Item-based Collaborative Filtering Algorithms the item-based approach looks into the set of items the target user has rated and computes how similar they are to the target item i and then selects k most similar items {i1,i2…,ik}.. At the same time their corresponding similarities {si1,si2...,sik}are also computed. Once the most similar items are found, the prediction is then computed by taking a weighted average of the target user's ratings on these similar items. One critical step in the item-based collaborative filtering algorithm is to compute the similarity between items and then to select the most similar items. The basic idea in similarity computation between two items i and j is to first isolate the users who have rated both of these items and then to apply a similarity computation technique to determine the similarity sij. In item-based collaborative filtering, an item- item matrix is maintained. A list of items is maintained on one side of the matrix and a list of same items is maintained on the other side. These entries are also updated as per the transactions.

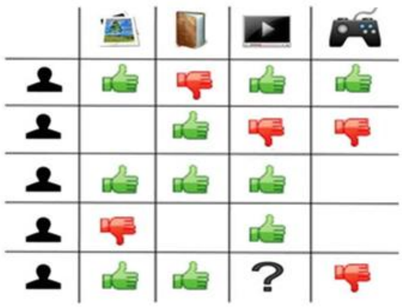

Fig.1. Person-to-Item Matrix

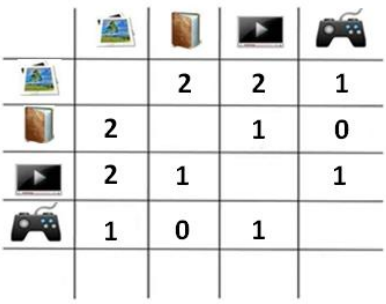

Fig.2. Item-to-Item Matrix

III. PROPOSED SYSTEM

There are some existing tools which provide personalized recommendations based on either person based collaborative filtering or item based collaborative filtering. The proposed system makes use of both these techniques. It tries to combine the best of both these collaborative filtering techniques. It tries to provide user with recommendations which people with similar tastes as his prefer as well as the items which are closely related to the

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

ones in which he has already shown interest are recommended to him.

The recommendation tool will scan both these matrices periodically. On the basis of this scan, the recommendations will be determined. The results of both the schemes are combined to get improvised recommendations. Performance of item-based prediction algorithms is of superior quality than user-based prediction algorithms.

Given below are two examples which explain how the two recommendation schemes can be used to strengthen the overall recommendation system. These examples show how items which are recommended by one algorithm but not by other are treated by the recommendation algorithm.

Consider an example,

Consider 5 items, A, B, C, D, E. There are 5 transactions in which different people and items are involved,

[T1] P1 → A, B, D
[T2] P2 → B, E, D
[T3] P3 → A, B, F, D
[T4] P4 → A, B, C
[T5] P5 → B, C, E

Here T1, T2, T3, T4, T5 are the transactions and P1, P2, P3, P4, P5 are the id's corresponding to people involved in the respective transactions. The → indicates the Items purchased in the respective transactions. Based on the above transactions we can plot a Person-to Item matrix.

This matrix will have persons along each row and an item as each column. Thus along the rows we have people , P1, P2, P3, P4, P5, P6. Also along the columns we have items, A, B, C , D, E, F. The symbol of tick mark in the column indicates that a person has purchased corresponding item. The person-item matrix and item-item matrix are shown below, with which, item B has a strong association. Consider the item-item matrix in fig(4). From the fig (4), we can infer that there is a strong association between B & D and A & B. Items B and D have been together in 3 transactions. Also Items A and B have been together in 3 transactions.

Therefore,

Item → Person ↓	A	B	C	D	E	F
P1	✓	✓		✓		
P2		✓		✓	✓	
P3	✓	✓		✓		✓
P4	✓	✓	✓			
P5		✓	✓		✓	
P6	$	✓	✓	?	$	

Fig.3. Person-to-Item Matrix of above Transactions

Item → Item ↓	A	B	C	D	E	F
A	-	3	1	2	0	1
B	3	-	2	3	2	1
C	1	2	-	0	1	0
D	2	3	0	-	1	1
E	0	2	1	1	-	0
F	1	1	0	1	0	-

Fig.4. Item-to-Item Matrix of above Transactions

Consider a transaction which includes B & C in fig.(3).

[T6] P6 → B, C

In this case person-based collaborative filtering will recommend A and E by considering the transactions [T4] & [T5]. In transaction [T4] the person purchased item A along with B & C. Also in transaction [T5] the person has purchased item E along with B & C. Hence this scheme recommends items A and E to the person P6 who has purchased B and C in transaction [T6]. The person-item matrix shows items purchased by different people in different transactions and also indicates the items to be recommended to the person P6 based on the person-based collaborative filtering. $ symbol indicates recommended items. However this scheme fails to recommend item D

B↔D
A↔B

Here is where item-based collaborative filtering will assist our tool. Since item-based filtering recommends A & D, the item D will appear in the recommendations for the user as item-based filtering has higher weightage as compared to person-based. 'A' will receive highest rating. 'D' will receive lower rating than 'A' but will definitely be placed above 'E'.

Consider another example,

A store has items for sale. It has bread, sugar, butter, sugar-free, etc as items for sale. It is observed that most people who purchase bread also purchase butter. With this observation, it can be observed that there is a strong association between bread and butter. Hence the item-based collaborative filtering will recommend butter to a person purchasing bread.

Bread↔Butter , is a strong relationship developed by item based collaborative filtering. However in case of an obese person, his doctor has advised him against having butter and sugar. Hence he purchases bread and sugar-free. However he is not interested in purchasing butter or sugar. However the item-based collaborative filtering will recommend him butter, which he is not supposed to have.

Here is where person-based collaborative filtering will assist our recommendation scheme. The person-based

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

collaborative filtering will find that those people who purchase bread and sugar-free do not purchase butter. Due to this it will not recommend butter to that person.
Now our recommendation tool takes into consideration the results of both the item-based and person-based collaborative filtering schemes. Since, butter is recommended by one scheme and not by the other, it will get a lower rating and will not appear in the shortlisted items. Thus this tool has improvised accuracy. Thus we find that the two collaborative filtering algorithms assist each other in providing improvised recommendations to the user.

IV. ALGORITHM FOR RATING RECOMMENDATIONS

1. If a particular item is recommended by both, item-based and person-based filtering algorithms, then it will be definitely recommended to the user and will get highest rating (Top, 'T').
2. If an item is recommended by an item-based filtering technique but not by person-based, it will get a lower rating as compared to the first case (Medium, 'M').

3. Finally if an item is not recommended by both the filtering techniques, then it will not be recommended, to the user.

V. DESIGN MODULES

The block diagram of the Collaborative Filtering Tool is given below. Let us consider each unit of this tool,

A. Database / Data Warehouse

Contains the logged data which includes 'transaction logs' in case of an e-commerce website and 'followed' information in case of social networking website. It also stores the final recommendations to be shown to the user. Refer fig(5) for this Database unit.

B. Transaction Analysis

All the 'followed' information is periodically analyzed. The new transactions are identified and the changes are then propagated to the next stage. Refer fig(5) for this Transaction Analysis unit.

C. Matrix Update Unit

Based on the analysis the matrices are periodically updated. These matrices can be updated periodically or after 'n' number of new transactions. There are two matrices, person-item and item-item matrices. For every new transaction (in case of e-commerce website) an entry gets added to the person-item matrix. Similarly for each new transaction, the association strengths between items (involved in the transaction) are updated. Refer fig(5) for this Matrix Update unit.

D. Collaborative Filtering Unit

The two tools, Item-based filtering tool and Person-based filtering tool work on the matrices and compute recommendations along with their strengths. Refer fig(5) for this Collaborative Filtering unit.

E. Rating Unit

Recommendations provided by both the tools are classified as top recommendations. Those present only in the Item-based and not in the Person-based are classified as medium recommendations. Refer fig(5) for this Rating unit.

F. Sorting Unit

Recommendations belonging to a particular class are sorted according to their strengths. These are then updated in the database. Refer fig(5) for this Sorting unit.

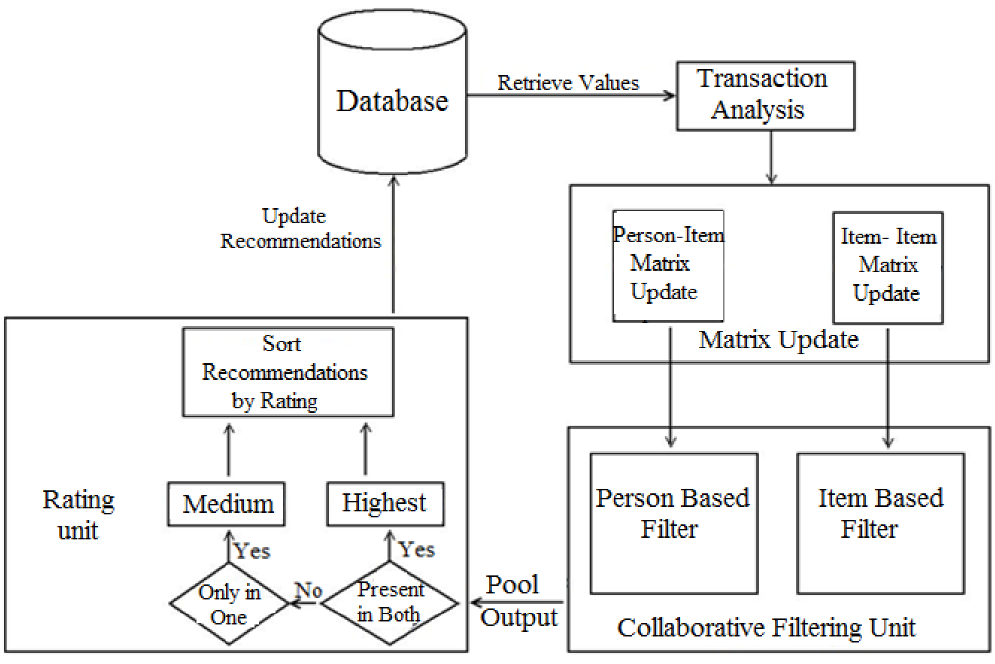

Fig.5. Block Diagram of Collaborative Filtering Tool

VI. APPLICATIONS OF TOOL

Due to the improvised recommendations provided by this tool it can be widely used in different types of websites like,

A. E-commerce website

Here the tool can be used to recommend Items to the user based on the past transactions made by the user. The Item-based collaborative filtering algorithm discovers strong associations between items made in all the transactions on the website, whereas the Person-based collaborative filtering algorithm will try to find out similar transactions in the transaction logs, and look for additional items to be recommended to the user.

B. Social networking websites

In these websites this algorithm can recommend ‘friends’ or ‘people whom you can follow’, by taking into account your existing ‘friend circles’ or ‘followed list’. Item-based collaborative filtering discovers which two ‘people’ appear together in different ‘friend lists’ or ‘followed lists’. Thus finding strong associations between persons. Person-based collaborative filtering will try to find out similar ‘friend lists’ or ‘followed lists’ and then look for friends to be recommended.

VII. CONCLUSION

The recommendation tool as proposed provides improvised recommendations by taking into account, not only the strong associations between two items but also, the user’s likes and dislikes and searching for people with similar views. Combining these two, the recommendation tool provides better, improvised recommendations to the user. It can prove to be an important Business Information Tool and can be widely used in different fields.

ACKNOWLEDGMENT

We are grateful to our principal and to our Head of the Department for giving us time and resources for our project. We would also like to thank the entire Computer Department for their support and guidance. We acknowledge all the researchers who have done pioneering work in the field of Web Personalization and Recommendation Tools and all the authors whose papers have been taken as a reference and as a guidance material for our report.

REFERENCES

[1] Amazon.com Recommendations, Item-to-Item Collaborative Filtering, Greg Linden, Brent Smith and Jeremy York.
[2] Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents, Manos Papagelis, Dimitris Plexousakis.