Implementation of Item and Content based Collaborative Filtering Techniques based on Ratings Average for Recommender Systems

Rohini Nair
University of Mumbai
k.j.somaiya college of engineering

Kavita Kelkar
University of Mumbai
k.j.somaiya college of engineering

ABSTRACT
Internet provides lots of information which is useful for recommender systems. Recommender systems are mostly used where web information is available in abundance in applications like book ecommerce. There exists many approaches to achieve recommendations like basic techniques of collaborative filtering and content based approach. These approaches can be done individually or combined depending on the type of recommendations needed by individuals. Recommender systems suggests items to purchase according to the users interest. Almost all applications in e commerce are working on the concept of recommendation system. It predicts recommendations while searching through large amount of information so that users can have the easiest access to the needed that best meet their needs and preferences.
Literature survey is done on recommender systems shows that a lot of work is being carried out in this area and the project proposes a combination of various techniques of recommendation systems. All are based on the basic techniques include item based approach and content based which are the basic building blocks for a recommender systems. In this paper both algorithms are implemented and their respective results are presented and compared.

General Terms
Collaborative filtering techniques, hybrid recommender systems.

Keywords
Item based, content based, collaborative technique, recommender systems.

1. INTRODUCTION
Information on the internet is huge and so is the search for any particular data as internet is ever developing and growing at a tremendous speed [1][2][3]. It provides with many opportunities, such as sharing knowledge, information, opinion with other users. Every user who has access to the information on the World Wide Web can have access to any data being in any part of the world. All this information on the internet kept getting huge and led to the problem of information overload. Users could not have access to the exact kind of information they were looking out for. Search engines available did solve the problem but it couldn't still be specific for any user. So the solution is found in recommender systems. Recommender systems[5][6] are tools for filtering and sorting items and information by using opinions of a group of users to help individuals belonging to that specific group to more effectively identify the area of interest from a set of choices. There are various algorithms and approaches which can be used to provide personalized recommendations. Recommender System is designed to provide meaningful recommendations to a group of users with a common interest. Many algorithms and techniques are used to understand the traditional and modern approaches [7] [8]. Problems of combining different technologies and recommending approaches for better results will always exist and will be the reason of new researches.

2. RELATED WORK
Recommender system is an important mean of information filtering and a potential method to solve the information overload problem. And collaborative filtering (CF) is the most important technique of recommender system. [1] [2] [3]. The business owner provides recommendations (such as books, videos, etc) that the customers may be interested in according to what other users bought. The customers' needs and interests are often not clear or ambiguous, so if the business owner can provide recommendations that just right meet their needs, the potential customers needs are becoming the real needs and the purpose of improving turnover can be achieved. Another approach is using content based technique [4] [5]. Currently, almost all large-scale e-commence systems (such as Amazon, EBay, etc) have used CF methods to recommend products to customers. Other Internet applications such as Music and News web sites also need CF methods to do personalization. Recommender systems[6][7][8] have (i) background data, the information that the system has before the recommendation process[9][10][11] begins, (ii) input data, the information that user must communicate to the system in order to generate a recommendation, and (iii) an algorithm that combines background and input data to arrive at its suggestions.

2.1 Collaborative filtering approach [1][2]
Collaborative recommendation is probably the most familiar, most widely implemented and most mature of the technologies. Collaborative recommender systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. The greatest strength of collaborative techniques is that they are completely independent of any machine-readable representation of the objects being recommended, and work well for complex objects such as music and movies. The critical step of collaborative filtering approach lies in searching the similar preference customers with the active customer, that is, find the similar customers. After finding

similar customers, it then presents recommendation for active customer according to the preference of similar ones.

2.1.1 Item to item based collaborative filtering algorithm

This feature called as item to item collaborating feature is similar to the impulse items in a supermarket checkout line, but our impulse items are targeted to each customer. The algorithm, item-to-item collaborative filtering [3], scales to massive data sets and produces high-quality recommendations in real time. Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list. Below are the general steps of item based approach.

Step 1: Registered users create their user profile, where a login and password is created which provides md5 protection

Step 2: Calculation of similarity distance using Pearson correlation coefficient is carried out.

Step 3: Next matching of similar items is carried out which is then passed as an argument to the generate recommendations functions which considers all the existing users items purchased and predicts accordingly using inputs from step2.

Step 4: Average of all the recommendations with the ratings are calculated and only those above a certain threshold-which changes always- are displayed to the user.

Step 5: Final Recommendations are displayed to user based on step 4.

The recommendations are based on the items in the customer's cart calculating the similarity between a single product and all related products. Below is a simple pseudo code[4][5] for any item based collaborative filtering.

For each item in product catalog, ITEM 1

For each customer A who purchased ITEM 1

For each item ITEM 2 purchased by customer A

Record that a customer purchased ITEM 1 and ITEM 2

For each item ITEM 2

Compute the similarity between ITEM 1 and ITEM 2

Item-to-item Collaborative Filtering is a widely used technique where rather than matching similar users, they match a user's rated items to similar items. In practice, this approach leads to faster online systems, and often results in improved recommendations. In this approach similarity between pairs of items i and j are computed offline using Pearson correlation, given by the simple formula as in equation 1:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{([n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

Where n-number of samples; x and y are the two inputs in our case the 2 ratings by different users, Now, the rating for item for user a can be predicted using a simple weighted average, as in: Weighted Average = [(n1 x N1) + (n2 x N2)] / (n1 + n2) where n1,n2… are relative weights and N1,N2… are values.

2.1.2 The Item to item based collaborative filtering Analysis

Item based collaborative filtering technique is used when we compare various items purchased by existing users and the recommendations vary depending on the item purchased at a given time. Recommendations are based on items purchased along with their ratings by other existing users and the averages amongst those are calculated and a final prediction is displayed.

2.2 Content-based filtering approach [6][7]

The content-based filtering approach has its origins in information retrieval and information filtering. The item recommended by content-based filtering often indicates textual information, such as news webs and documents. And these items usually describe with keywords and its weights. Database fields are queried and used to analyze the textual feature content of items and recommend suitable content based on items characteristics and the user's preference. The challenge of this approach includes limited content analysis because of limited keywords, overspecialization problems and new user problems. Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem. In the recommendation process, the engine compares the items that were already positively rated by the user with the items he didn't rate and looks for similarities. Those items that are mostly similar to the positively rated ones, will be recommended to the user. There are different algorithms [8][9] of measuring similarities among items in data base and those in user's profile. Better described items lead to more accurate recommendations.

2.2.1 Content based algorithm

Step 1: Registered users create their user profile, where a login and password is created which provides md5 protection

Step 2: Information is stored in the database which is queried to display a list of available items to the user.

Step 3: Once the user selects the desired category, for example books, the database queries the records and returns a match for the chosen category.

Step 4: Average of all the recommendations are calculated and only those above a certain threshold-which changes always- are displayed to the user.

Step 5: Final Recommendations are displayed to user based on step 4.

In the content based recommendations algorithm proposed the example considered is for the applications of books in a library database. Initially a user profile is created by taking various information from the user and storing it in the database. Recommendations are made on the basis of keywords in the user profile and passed as a query. For example once a user selects a particular book from the menu its subcategory is retrieved and all the books belonging to that subcategory are displayed. The reason for displaying all under the subcategory is because we need to display certain results to the user. Next the average of the ratings is calculated from the displayed results by calculating from the results. There is

no fixed threshold as the recommendations then will not be accurate and useful.

2.2.2 Content based algorithm Analysis

Content based recommendations are used when the user preferences are extracted from the information entered by the user at the time of profile creation. The results of this algorithm can be based on various criteria thus improving recommendations at every stage.

3. DISCUSSION

In this section advantages and disadvantages [10][11] of item based and content based collaborative techniques used for recommendations are discussed as given in Table 1:

Table1: Differences among Item and content based recommendations

Algorithm	Advantages	Disadvantages
Item based recommendations	A user may receive items that has never been used before but may be of potential interest.	Difficult to provide recommendations for users having unusual preferences.
	Facilitates the sharing of information and experiences among users having similar views.	Difficult to classify users with ever changing preferences.
Content based recommendations	A user can receive proper recommendations without help from other users.	Certain kinds of items like multimedia are difficult to analyze.
	It is possible to look into the problems of multiple users by monitoring the ever changing and evolving user profiles.	User receives items similar to past experiences.

4. EXPERIMENTAL RESULTS

This section represents the implementation results of Item to Item based collaborative filtering and content based recommendations. All experiments are performed on Intel Core i3-380M processor with 4 GB memory using MySQL server as a backend for database and php for front end programming. A huge database of approximately 1000 records of books information is used. User login screens are created and item based and content based approaches are implemented. Time required for execution of the two algorithms is as given in Table 2:

Table 2: Execution time for the two algorithms

Algorithm	Elapsed Time (sec)
Item based approach	21.00012 approx
Content based approach	18.23000 approx

Graphical output

Graphical output in Figure 1 shows the elapsed time presented in terms of bar chart for item based and content based approach.

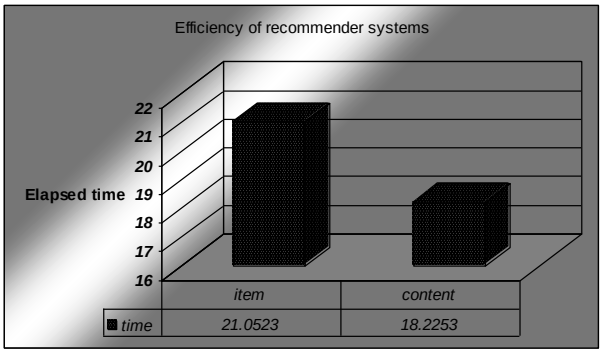

Figure 1 : Efficiency of recommender algorithms based on time elapsed.

Figure 2 shows password with md5 encryption of users.

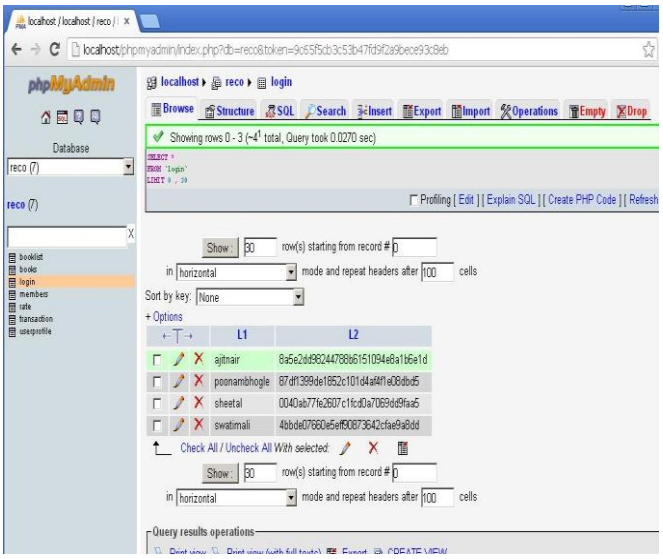

Figure 2 Screenshot of md5 protection for password of user

Implementation results of Item to item based recommendations are given in Figure 3

Welcome 'ajitnair'

Books suggested	Rating
Data Structure	5
Circuits & networks	5
Introduction to Languages & Theory of Computation	5
Best Practices in Engineering Education	5
Essential of Business Communication	5
TCP / IP and Distributed System	5
Oralce Developer 2000	5
Networking concept	5
Robotics & Control	5
Applied cryptography	4.3
NT Windows 95	4.3
Proceedings of Advanced Computing and Communication Technologies (ICACCT-2008)	4.1
ASP.NET web prog. With VB2008	4.1
Engineering Circuit Analyasis	4.1
Computer System Architecture	4.1
Object-Oriented Programming in TURBO C++	4.1
Electronic Communication Systems	4
TCP / IP Protocol Suit	4
Oracle SQL and PL/SQL	4
Management information systems	4
Electronic Device and Circuits	4
Discrete Mathematics For Computer Scientists & Mathematicians	4
E-commerce & Mobile Commerce Tecnology	4

Figure 3 Item based recommendations

Implementation results of Content based recommendations (subject wise) are given in Figure 4

Welcome 'sheetal'

Theory of Computer Sciecne | View similar

All the books under - tcs- are displayed

Title	Ratings
Introduction to Automata Theory, Languages & Computation	9
Introduction to Languages & the Theory of computation	9
Introduction to Languages & Theory of Computation	8
Theory of Computer Sciecne	6
Introduction to Languages & the Theory of computation	6
Theory of Computer Science	4
Theoretical Computer Science	2

Books recommended for you on the subject -tcs- which has ratings as below :

Title	Ratings
Introduction to Automata Theory, Languages & Computation	9
Introduction to Languages & the Theory of computation	9
Introduction to Languages & Theory of Computation	8
Theory of Computer Sciecne	6
Introduction to Languages & the Theory of computation	6

Figure 4 Content based recommendations

Implementation results of Content based recommendations (publisher wise) are given in Figure 5

172.17.15.40/publication.php

All the books for the publication - Nandu - are displayed

title	Ratings
Digital Signal Processing	8
Basic Electricity And Electronics-II.	5
Electrical Networks	4
Basic Electricity And Electronics.-I	4
Electrical Networks	4
Fundamental of Robotics	4
Electrical Networks	3
Basic Electricity And Electronics-I.	2
Basic Electricity And Electronics-II.	1
Circuits & networks	1

Books recommended for you on the publisher - Nandu - recommended :

title	Ratings
Digital Signal Processing	8
Basic Electricity And Electronics-II.	5
Basic Electricity And Electronics.-I	4
Fundamental of Robotics	4
Electrical Networks	4
Electrical Networks	4

Figure 5 Content based recommendations

5. CONCLUSION

In this paper two algorithms on Item based and Content based collaborative filtering techniques were successfully implemented on mysql/php. The algorithms were tested for approximately 700 records considered as a dataset. It was observed that both the techniques resulted in predictions which were useful to a naïve user who can just look upon the results and approve of the recommendations. Item based considers other users aspects also into its predictions while content based is confined to its own available information. On implementing item based approach proved to be more efficient as compared to the content based approach, however depending on a user's need either of the recommendations prove to be useful. The output of this implementation can be further used for hybrid recommendations which are a combination of the above results.

6. REFERENCES

[1] Recommender Systems,Prem Melville and Vikas Sindhwani,IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

[2] Algorithms and Methods in Recommender Systems,Daniar Asanov,Berlin Institute of Technology,Berlin, Germany.

[3] Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions(vol. 17 no. 6) pp. 734-749,Gediminas Adomavicius, IEEE,Alexander Tuzhilin, IEEE., June 2005.

[4] A Survey of E-Commerce Recommender Systems,Kangning Wei1, Jinghua Huang2,Shaohong

Fu3,1 2School of Economics and Management, Tsinghua University, Beijing,7/07/$20.00 IEEE ©2007.

[5] Hybrid Recommender Systems:Survey and Experiments,Robin Burke,California State,University, Fullerton,Department of Information Systems and Decision Sciences. Volume 12Issue4,Pages331–370,November2002.

[6] Recommender Systems in E-Commerce J. Ben Schafer,GroupLens Research Project,Department of Computer,Science and Engineering,University of Minnesota,Minneapolis,Joseph Konstan,GroupLens Research Project,Department of Computer,Science and Engineering,University of Minnesota,John Riedl,GroupLens Research Project,Department of Computer,Science and Engineering,University of Minnesota.

[7] Empirical Analysis of Predictive Algorithms for Collaborative Filtering,John S. Breese, David Heckerman Carl Kadie,Microsoft Research.

[8] Implementing a Rating-Based Item-to-Item Recommender System in PHP/SQL, Daniel Lemire,Universit´e du Qu´ebec,Sean McGrath,University of New Brunswick, November 4, 2005.

[9] Research and Design of an Efficient Collaborative Filtering Predication Algorithm, Qilin Li, Mingtian Zhou,College of Computer Science and Engineering,University of Electronic Science and Technology of China.

[10] Recommender Systems Handbook, Ricci, F.; Rokach, L.; Shapira, B.; Kantor, P.B. (Eds.)2011.

[11] Personalized Recommender Systems in e-Commerce and m-Commerce: A Comparative Study, *Azene Zenebe, Ant Ozok and Anthony F. Norcio* Department of Information Systems University of Maryland Baltimore County (UMBC) Baltimore, MD 21250 USA.

[12] Collaborative Filtering Recommender Systems ,J. Ben Schafer1,, Dan Frankowski2, Jon Herlocker3,Shilad Sen2,2007.

[13] Recommender Systems in e-Commerce: Methodologies and Applications of Data Mining, Dr. Bharat Bhasker , K Srikumar,July 29, 2010.

[14] Personalized Recommender Systems in e-Commerce and m-Commerce: A Comparative Study Azene Zenebe, Ant Ozok and Anthony F. Norcio Department of Information Systems University of Maryland Baltimore County (UMBC)Baltimore, MD 21250 USA.