



International Journal of Engineering, Business and Enterprise Applications (IJEBA)

www.iasir.net

Adding Weights to User Rating in Recommendation Systems

Dipak V. Patil

Associate Professor

Department of Computer Engineering

Gokhale Education Society's

R.H. Sapat College of Engineering,

Management Studies and Research, Nashik, M.S., India.

Abstract: Large amount of data is available on web and information extraction takes place on this large volume of data. User submits queries to search engine in natural languages and generally contains one or two words. Search engines cannot recognize natural language and thus it becomes difficult to retrieve required information. This results in need of recommendation system. Multiple recommendation techniques are available and each technique has its own advantages and disadvantages. The algorithms proposed so far do not consider weight of user prediction based on user characteristics. Recommendation systems support various types of data sources. The data is in the form of text, images, audio, video etc. Well-organized method to handle different types of data is, to model them in the form of graph and then apply recommendation algorithm with weights on it. The weights are allocated based on user's expertise in field, qualifications and experience. Initially, proposed system gathers data from users. Consequently the system applies different algorithms for predicting user's interest. The item to item base collaborative filtering algorithm, Pearson correlation base collaborative filtering algorithm are some of the algorithms. These two algorithms explore similarities between item and users respectively. Slope one algorithm aims to find out the rating of un-rated items. In proposed method weights are added to rating given by the user based on various user parameters mentioned above. Results of these algorithms are combined for enhancing the performance of the recommendation system.

Keywords: Weights, Recommendation system, collaborative filtering, web graphs.

I. Introduction

Web mining extracts interesting pattern from the web. Web mining is categorized as content, structure and usage mining. Content mining is a text extraction; it mainly deals with unstructured data. Structure mining mines data from hyperlinks; it gets the summary of the web pages. In case of usage mining the data is received from log files. Data available on web is large in size and extracting interesting data patterns from such a large data is quite difficult task, another problem with such data is heterogeneity which makes it time consuming and complex task. To solve these problems recommendation system is needed. Usually queries are submitted to search engines. The queries are expressed in one or two words and it becomes difficult to search semantically relevant data. The results presented by these queries are built on the basis of ranking known to the pages and the result may not have data in connection to user's interest. In addition search engines may not consider the personalization features, that is, do not emphasize on the historical data. The data is the data accessed previously by that user representing user's interest. To deal with such problems various algorithms are used but there is need of generalized approach that can handle these problems. Designing such generalized approach is difficult task, as available data on the web is heterogeneous by nature.

To perform mining operation on web various recommender systems are used. Recommender is a technique that automatically predicts the interest of an active user by collecting rating information, behavior of the user from other similar users or items. But various data sources are used on the web like image, text, audio, and video and for different data sources different recommendation algorithms are used. It is required to provide a general framework which can be applied to all type of data sources. It is possible to model this data sources in the form graph and then by applying recommendation algorithm results are obtained. One of problem related to recommendation system is that some time user may not give rating to some items, processing this incomplete data leads to inefficient results. This problem can be solved by predicting users rating before applying recommendation algorithms.

Recommender system has fundamental capabilities such as it gives recommendations and tries to predict user's interest but due to huge amount of data and scattered data they sometimes it fails to predict users interest and unable to give personalize recommendation.

Different recommendation algorithms are available for extracting data from web. But for different type of data sources different algorithms are required. There is need of some generalized method which can be applied to all this type of data sources.

User some time may not give rating to some items, this incomplete information is processed, and that may lead to inefficient results.

II. Literature Survey

Recommendation systems are the information filtering system that try to predict the 'rating' or 'preferences' that users would give to an item. These systems use details of users profile, their opinions and their habits and compare the information to reference characteristics to present the recommendations.

Recommendation Techniques

Recommender systems got popular in extremely short time, and they are used in a variety of applications. There are many recommendation systems developed for applications like movies, music, news, books, research articles recommendation. This section takes an overview of different existing recommendation system.

Collaborative information filtering

Collaborative filtering is the method which aggregates rating or preferences on items and by using this historical data recommendation is done [2],[3],[4]. This method also shares its rating information between different users which helps the other users to find the data of their interest. Consider an example user A and user B gives similar rating to item I or have similar behavior like purchasing, watching movie etc. Then they may have same area of interest, thus a system can suggest items to user A which are previously referred by user B or vice versa. There are two types of algorithms that are used for collaborative filtering first is memory based and second one is the model based algorithm. Memory based algorithms are based on total ratings given by the user on database. This is further classified into two type user based rating and item based ratings. In user based rating algorithm, the users having same interest are considered. In item based rating algorithm, it calculates similarity between two items and according to that it makes the group of it.

Content-based information filtering

Content-based filtering system considers historical data that means recommend those items that are similar to the ones that the user preferred in the past [5],[6]. Here, content means a web object. This system finds similar items by comparing their features or matching the attributes of user profile which contain information related to users preferences. For example, if in past user purchase a book that belongs to the poetry genre, then the system can learn to recommend other books from same genre.

Knowledge-based information filtering

Knowledge-based systems recommend items based on specific domain knowledge about how certain item features meet user's needs and preferences [7], [8]. Similarity function is used to find out the recommendation set which compare problem description and problem solution. Items are recommended using similarity score. Constraint-based systems are another type of knowledge-based recommendation system. Constraint-based recommenders consider predefined knowledge bases that contain explicit rules about how to relate customer's requirements with item features. These systems works better than others at the beginning of their deployment but if they are not equipped with learning components they may be surpassed by other shallow methods that can exploit the logs of the human/computer interaction. Knowledge-based information filtering is the personal logic system [2]. It helps users to make decisions on variety of products. It collects users requirements on a particular product, and consulting its knowledge base to find the items that best meet the users requirements. This system have several advantages like it not need to be initialized with a database of user preferences since its recommendations do not depend on the user ratings of items, this system also adjust their recommendation quickly as a user's interests change because its recommendations are independent of the users preferences in the past. It is important that to make good recommendations, this system must understand the product domain. It must have knowledge of important features of the product. Advantages of this system are that it doesn't have the "ramp-up" problem because recommendation given by this system is not depend on any data base of user ratings. Users are encouraged to explore and understand the information space and, by doing so, they elaborate more on their needs.

Community-based information filtering

This system considers preferences of current user and their friends. This technique follows the epigram Tell me who your friends are, and I will tell you who you are [2]. It is observe that people tend to rely more on recommendations from their friends than on recommendations from similar but anonymous individuals. This type of recommendation system models mainly use information of user and there social relations. It recommends items to user which is preferred by users' friends. It is observe that in some cases social-network data yields better recommendations. As social networking field is growing rapidly it provides connection between different domains. People having interest in same area are coming together by this social network; it is possible that they may rate same content. These dimensions can be used to detect communities among people. In this community detection is required. Community detection in social media analysis helps to understand more of users' collective

behavior. Aim of community detection system is to find subgroups among people, interaction between different subgroups [9],[10].

PageRank

This algorithm is developed by Google's developer Sergey Brin and Larry Page at Stanford University [11],[12]. Google uses this algorithm for information retrieval. Google provides the facility to view the PageRank of any website which is given on its toolbar in green color.

Hypertext Induced Topics Search (HITS)

It is developed by Jon Kleinberg. It contains hubs and authorities as the nodes [13],[14]. Authority node contains the webpage having information related to key term. Hub node contains links to the webpage which have information related to key term. Hub score and authority score is calculated for each node. Initially we have to assume some hub and authority score for each node after that these scores are updated on each iteration of the algorithm. Hub score is calculated as the sum of authority score it is pointing to and authority score is the sum of hub score which have outgoing links to this node.

Image Recommendation technique

Image Recommendation technique is also one of the interesting recommendation application on the web [15],[16]. This technique mainly focuses on recommending interesting images to users based on their preferences. This can be done by providing some images to users, and according to their interest they rate images and by analyzing this rating data images are recommended to user.

Demographic-based Recommendation

This system categorizes the user based on personal attributes like college students, teenagers, women, men, etc. Based on the this data collected from user profile's are created [17],[18]. For example teenagers prefer different products than the elders. The benefit of a demographic approach is that it may not require a history of user ratings of the type needed by collaborative and content-based techniques. Demographic data can be used to identify users' interest and create "stereotypes".

Various Algorithms Used in Recommendation System

This section discusses various approaches used in recommendation systems.

Item-based Collaborative Filtering

Item-based collaborative filtering [1] determines similarity between two items. It indicates proportion of similarity between different items. The similarity value is used so as to recommend similar type of products to users according to their interest in that type of items. Vector based similarity measure is used to measure the similarity value in terms of direction of two vectors in n dimensional space. The input is rating specified by different users to item. Record for every item I_i and item I_{i+1} in item catalogue is preserved for each user U who procured item I_i as well as item I_{i+1} . Estimate similarity between item I_i and item I_{i+1} . The operation results in finding similarity amongst two items.

Cosine based similarity Measurements: In case of vector-based similarity, intended for the two items, the ratings assigned to the items are considered as vectors and angle between these vectors characterizes the similarity between items. When these vectors point in identical directions conclusion is they are more similar to each other, otherwise vectors point in opposite directions indicating they are less similar to each other. The consequential similarity value is recognized as cosine base similarity. Its formulation is shown in equation (1)

Pearson correlation coefficient based collaborative filtering

Pearson correlation coefficient based collaborative filtering is used to compute similarity between users. Contributing active user's preferences are compared. Rating specified by users to different items is used as input in computation. Data for every user U_i and U_{i+1} is kept in user catalogue; it records data for each product P purchased by U_i as well as U_{i+1} . Compute similarity between U_i and U_{i+1} .

In Pearson correlation coefficient based similarity measurement, The Pearson correlation coefficient is the basic correlation formula is applied on samples for rating information. It measures how much two users diverge from their normal votes, that is, the direction/magnitude of every vote as compared to their average vote. If they vary in the same manner on the items they have rated in common, it is a positive correlation; otherwise, a negative correlation. Let x and y be a rating given by different users, Let n be number of common item rated by that users. The correlation is computed by following equation 2.

Slope One Rating Prediction

This approach is rating based collaborative filtering approach [12]. This method computes rating of un-rated items by using the rating assigned to those items by other users.

Table 1: Example presenting rating of items along unrated item.

	Item I1	Item I2
User1	X11=3	X21=4
User2	X12= 4	X22=?

$$X_{22} = X_{12} + (X_{21} - X_{11}) = 5.$$

This method generally considers two items and users at time to compute the rating of unrated items, in case if numbers of users are greater than two numbers, then the method computes rating prediction value using data of all other user who have rated both items, considers average of all rating values to compute final rating prediction value. d by U_2) is calculated as $X_{22} = X_{12} + (X_{21} - X_{11}) = 5$.

III. Proposed System

In proposed approach initially click-through data records are acquired and various operations are performed on this data. Thus proposed system works in multiple stages. Fig. 2 presents working of working in proposed system. The stages include data extraction, pre-processing of data, applying various algorithms, applying hybrid approach and recommendation of items.

Data extraction: The first stage in proposed system is data extraction. Click-through data registers the activities of web users; the facts related to interest of user are acquired, the semantic relationships amongst users, queries and clicked internet documents. A dataset contents queries as data and metadata related to queries. Each line of click-through data comprises: user ID, query delivered by the user, URL on which the user clicked, the ranking of that URL, and the timing at which the query was submitted.

Applying Algorithms: Various algorithms are applied to compute top n recommendations. The algorithms are Item based collaborative filtering algorithm, Pearson correlation base collaborative filtering algorithm are used to compute similarities between item and users respectively. Slope one algorithm computes the rating of un-rated items according to user's weight.

Similarities concerning different items are computed using item to item collaborative filtering algorithm. The items with similarity value greater than some threshold are considered for recommendation. Slope one algorithm computes the rating score of unrated item, thus more items are available for recommendation. As two methods in combination are used the efficiency of recommendation system gets enhanced. As next phase of this method, similarities between users are calculated by using person correlation filtering method. The similarity value characterizes the similar interests of users. Users with similar interest are computed and the items designated for recommendation in previous phase of hybridization are recommended to users in which both users have similar interests.

IV. Method of Experimentation and Results

This experimentation is extension to the implementation [22]. For experimentation evaluation a sample data set is formed by considering 5 user volunteers (U_1 to U_5) to rate 4 items I_1 to I_4 . The rating scale is scale 1 to 5 and 0 as unrated. The proposed approach is implemented using Java. The computed results are as follows. These results are obtained in stages after applying Pearson correlation coefficient algorithm and item to item collaborative filtering algorithm. Sample data set is presented in Table 1.

Table 2: Data presenting Users Rating for Different Items I_i represents Items and U_j represents Users

	Weight	I_1	I_2	I_3	I_4
U_1	1.0	3	4	0	1
U_2	0.75	3	4	3	5
U_3	1.0	0	2	3	5
U_4	1.0	1	4	4	5
U_5	0.80	5	5	0	3

Slope one algorithm predicts the rating of unrated items. It considers two items and two users to predict the rating of unrated items, if number of users are greater than two, then it searches for rating prediction value for all other user those have rated both item. Get average of those values to compute final rating prediction value. Results after applying slope one algorithm are represented in Table 2.

Table 3: Predicted Rating of Values after Applying Slope One Algorithm

User	I_1	I_2	I_3	I_4
U_1	3	4	0	1
U_2	2	3	2	4
U_3	0	2	3	5
U_4	1	4	4	5
U_5	5	4	0	2

Pearson Correlation Coefficient: In this approach two users are compared with each other to compute the similarity between two users. Pearson correlation coefficient formula is applied to explore the relationship between two users. Correlation factor values resolve in how these users are similar to each other. Threshold value fixed is 0.60.

Users bearing similarity value greater than 0.60 are deliberated as similar users. In Table 5 user U2 and user U3 have similarity value 0.87 thus they are abundant similar. Thus, items those are recommended to user U2 conferring his/her interest is also commended to user U3.

Table 4: Predicted Rating of Values after Applying Slope One Algorithm

	I1	I2	I3	I4
U1	3	4	5	1
U2	2	3	2	4
U3	3	2	3	5
U4	1	4	4	5
U5	5	4	5	2

Table 5: Similarity between Different Users is determined by Pearson Correlation Cofactor

User_id	U1	U2	U3	U4	U5
U1	0.00	0.27	0.42	0.42	0.91
U2	0.27	0.00	0.87	0.87	0.54
U3	0.21	0.87	0.00	0.60	0.40
U4	0.42	0.87	0.60	0.00	0.32
U5	0.91	0.54	0.40	0.32	0.00

Cosine based Similarity: In a recommendation system similar items need to be recommended. Consequently items are compared with each other to find out similarity between them. Association among these items is computed using cosine based similarity formula. The threshold value set is 0.95. The items having cosine similarity value exceeding threshold value are similar to each other. The results showing similarity among different users (computed using cosine similarity) are presented in Table 4.

Combining Algorithms: The proposed recommendation approach initially applies slope one algorithm and computes rating of unrated items. Item based collaborative filtering process computes association amongst items. Similarly based content filtering algorithm computes the association amongst users. The proposed recommendation system combines outcomes of these algorithms and computes the interest of users.

Cosine similarity based method computes similar data items for example items I2 and I3 are computed to be similar item. Similar users are computed by means of Pearson correlation factor. The similar users computed are {U1, U5}, {U2, U3} and, {U2, U4}. Item I2 is rated as 4 by user U1 thus I2, I3 are Recommended to similar user U5. Similar cases are presented in Table 7.

The proposed hybrid method calculates similarity between different users using Pearson Correlation cofactor; in addition it uses cosine similarity measure to find out similarity between different items.

Table 6: Similarity between Different Items after applying Cosine Similarity algorithm

Item_id	I1	I2	I3	I4
I1	1.00	0.88	0.92	0.76
I2	0.88	1.00	0.97	0.92
I3	0.92	0.97	1.00	0.94
I4	0.76	0.92	0.94	1.00

Table 7: Results after Applying Hybrid Recommendation Algorithm. Items Recommended to different User

Item_id	Rated By	Recommended To	Recommended Item
I2	U1	U5	I2,I3
I3	U5	U1	I3,I2
I2	U4	U2	I2,I3

Thus hybrid method is more accurate as compared to cosine similarity measure. It is observed that number of recommendations using hybrid method are less as compared to cosine similarity method, but as discussed earlier though number of recommendation are less accuracy of recommendation is high using proposed hybrid approach.

V. Conclusion

A hybrid recommendation system is proposed here. All existing recommender systems use one or more basic techniques: content-based, collaborative, demographic, utility-based and knowledge-based. A survey of these technique shows that they have complementary advantages and disadvantages. A hybrid recommender system is developed for improving performance. The proposed design is hybrid recommendation technique to get more personalized recommendation by combining multiple recommendation algorithms. This also provides solution for

problem of unrated item. User some time does not rate to some items, this incomplete information is processed by the system that may lead to inefficient results, and proposed system gives solution to this problem. Initially Slope one algorithm rates un-rated items considering users weight. Item based collaborative filtering algorithm determines the relationship between items and content based filtering algorithm determines the relationship between users. In proposed hybrid method results of these algorithms are combined. Results of these algorithms are used for determining the interest of users and according to that items are recommended to user. It is observed that the proposed hybrid method works efficiently. It is observed that number of recommendations using hybrid method are less as compared to cosine similarity method, but as discussed earlier though number of recommendation are less accuracy of recommendation is high using proposed hybrid approach.

References

- [1] Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B. Kantor, "Recommender Systems Handbook", Springer Science Business Media, LLC, 2011.
- [2] Jerzy Surma and Mariusz Lapczynski, "Selecting Data Mining Model for Web Advertising in Virtual Communities", The First International Conference on Advances in Information Mining and Management, 2011 ISBN: 978-1-61208-162-5.
- [3] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering", In Proc. of WWW, Banff, Alberta, Canada, 2007.
- [4] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative Filtering recommender systems", ACM Trans. Information system, 2004.
- [7] R. Burke, "Integrating Knowledge-based and Collaborative-filtering Recommender Systems", Proceedings of the Workshop on AI and Electronic Commerce, AAAI 99, Orlando, Florida.
- [8] R. Burke, K. Hammond, and E. Cooper, "Knowledge-based navigation of complex information spaces" In Proceedings of the 13th National Conference on Artificial Intelligence, pages 462-468, 1996.
- [9] Youngchul Cha Bin Bi Chu-Cheng Hsieh Junghoo Cho, "Incorporating Popularity in Topic Models for Social Network Analysis", SIGIR13, July 28-August 1, 2013, Dublin, Ireland.
- [10] Y. Cha and J. Cho, "Social-network analysis using topic models", In SIGIR, ACM 978-1-4503-1472, 2012.
- [11] Danil Nemirovsky, "Web Graph and PageRank algorithm", Department of Technology of Programming, Faculty of Applied Mathematics and Control Processes, St. Petersburg State University, Russia, 2009.
- [12] Matteo Pasquinelli, "Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect", In Konrad Becker and Felix Stalder (eds), Deep Search: The Politics of Search Beyond Google, London, Transaction Publishers: 2009.
- [13] Jayman Farahat, Thomas Lofaro, Joel C. Miller, Georg Rae, and Lesley A. Ward, "Authority Rankings From Hits, PageRank, and Salsa: Existence, Uniqueness, and Effect Of Initialization", ACM Sigir, 2001.
- [14] Jon M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment", Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, and as IBM Research Report RJ 10076, May 1997.
- [15] G. Pass, A. Chowdhury, and C. Torgeson, "A picture of search", In The First International Conference on Scalable Information Systems Kong, Hong Kong, June 2006.
- [16] Aixin Sun, Sourav S. Bhowmick and Jun-An Chong, "Social Image Tag Recommendation by Concept Matching", MM11, November 28-December 1, 2011, Scottsdale, Arizona, USA.
- [17] Manisha Hiralall, "Recommender systems for e-shops", Business Mathematics and Informatics, Vrije Universiteit, Amsterdam, 2011.
- [18] Robin Burke, "Hybrid Recommender Systems: Survey and Experiments", User Modeling and User-Adapted Interaction, Volume 12, Issue 4, pp 331-370, November 2002.
- [19] R. I. Kondor and J. D. Lafferty, "Diffusion kernels on graphs and other discrete input spaces", In Proc. of ICML, 2002.
- [20] Hao Ma, Irwin King and Michael R. Lyu, "Mining Web Graphs for Recommendations", IEEE transaction on knowledge and data engineering, 2011.
- [21] X. Wang and C. Zhai, "Learn from web search logs to organize search results", In Proc. of SIGIR, The Netherlands, 2007.
- [22] Priyanka U Chavan, P M Yawalkar and D. V. Patil, "A Hybrid Approach for Recommendation System in Web Graph Mining", International Journal of Computer Applications, 95(24):23-27, June 2014.