

A Survey on Recommendation Techniques in Numerous Domains

Gourav Jain,
School of Information
Technology
R.G.P.V Bhopal, M.P.,
India, 462036

Nishchol Mishra,
School of Information
Technology
R.G.P.V Bhopal, M.P.,
India, 462036

Sanjeev Sharma
School of Information
Technology
R.G.P.V Bhopal, M.P.,
India, 462036

ABSTRACT

With the increase in the amount of information available on the internet, there is a challenge of providing relevant and useful information to the interesting users on the basis of their interest although when user wants to search data of their interest, they have to search in whole databases, which is very tedious and time consuming too. So a system is needed which provide useful information based on user interest named Recommendation System. A Recommendation System is a sturdy and valuable tool used for decision making and provides a ranking of the most popular items based on user preference. Various algorithms were proposed by different researchers for recommendation of web pages, items, movie, video etc. This paper gives us a snapshot of latest work accomplished in the field of recommendation.

Keywords— Collaborative Filtering, LDA, Naive Bayes, ERPM, TyCo.

1. INTRODUCTION

Recommendation systems found their application in the field of e-commerce and internet where items suggest to a group of user on the basis of their requirement based on their interest. A recommendation system is a type of information filtering system that built a model from the characteristic of an item according to the rating or prediction, given by a user to an item. The main motivation behind using the recommendation system is that it is based on real activity, reduced overloading and requires less organizational maintenance.

Any recommendation system consists of two basic entities user and items and based on ranking phenomena of items which help in decision making. A ranking is defined as the relationship between a set of items such that, for any two items, the first is either getting higher or lower or equal ranking than second, however the higher ranked item get's more preference as compare to lower ranked items. The output of recommendation system is depending upon the type of input given to a system wherein input may be either rating or data in recommendation system. The rating is a suggestion given by user on items and data can be age, gender and education of users. Based on input given to system output of recommendation system may be a prediction or recommendation. Prediction is defined as the opinion of users given to any item, which gives lower error rates and recommendation is the one which is most liked by users. The technique which used in recommendation for analyzing the data and finding patterns in a set of data is data mining. Data mining is the process of analyzing data from different perspectives and succinct it into useful information. The

impetus behind this paper is to analyze the various algorithms used for the recommendation of the Web Pages, Items, movie, venue, and video etc. The rest of the paper is organized as follows: Section 2 to 6 discusses different methods based on the Recommendation of Items, Recommendation of Web Pages, Recommendation of movies, Recommendation of tags and Recommendation of others. Finally Section 7 concluded the papers.

2. RECOMMENDATION FOR ITEMS

The recommendation of items can be done using several methods which are discussed below as follows:

2.1 Item Based Collaborative Filtering

Collaborative filtering is one of the most important and powerful technique used for recommendation in which every user interacts with other user and provides an option to set of products for establishing the quality of product. The traditional collaborative filtering is able to recommend the data effectively to the interested users with small amount of information, but the amount of information is increasing, with time, the traditional collaborative filtering system fails in achieving scalability and quality of the recommendation. To overcome this problem a new recommendation algorithm is needed which produced high quality recommendation even for large scale data .In addressing these issues Sarwar. et. al [1] proposed a new algorithm named item based collaborating filtering recommendation algorithm. In the proposed algorithm firstly computing the item-item similarity for identifying different relationship, and based on this relationship author compute the recommendation for users. The author used the different technique for calculating the item similarity and recommendation such as cosine similarity, regression model etc. The experimental result reveals the fact that an item based algorithm gives better result that user based collaborative recommendation algorithm.

2.2 Ranking by Naive Bayes

Sometimes the problem in suggesting the true popular item to the users, can give more importance to some items, but may affect the ranking of other items, those rankings is dependent on the resulting items. So, Gouthami. et. al [2] proposed a model based recommender algorithm called Naive Bayes which provide fast and highly scalable recommendation. The Naive Bayes algorithm uses the Bayes theory concept based on conditional probabilities for ranking and author work on every attribute contained in the data, and treats as each is equally important and independent of each others. The algorithm constructs a decision tree, which helps in ranking

the popular items. The Naive Bayes algorithm constructs a suggested set which has a less or equal (if the total no of items is less) number of items those were checked in suggestion set. No problem occurred if the item is present in suggestion set , but in case if the item is not presented with the suggestion set than requested item from the entire set is replaced with suggestion set items who have a less probability. The naive Bayes algorithm is the best suitable method for ranking and suggesting the popular items based on true popularity of items.

2.3 Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques

Recommender systems are becoming important to individual users and businesses for providing personalized recommendations. Techniques used for recommendation mainly emphasized on improving the accuracy while other aspects such as diversity of recommendations, have not been focused. So, Gediminas. et. al [3] proposed an item ranking technique that produced a recommendation with high diversity along with accuracy. The traditional recommender system had high accuracy because system ranks the relevant items in descending order of their predicted rating for each user and recommended top N items. In this method author considers additional factors, such as item popularity which increases the diversity along with accuracy, when ranking the item for the recommendation. The benefit of using ranking techniques is that they offer flexibility to system designers, in terms of parametric and can be used in conjunction with different rating prediction algorithms. The proposed method not only shows improvement in diversity about 20-25 percent with 0.1 percent of accuracy loss, but also has several other advantageous characteristics like these techniques are extremely efficient, parameterizable and flexible along with advantage of diversity improvement.

2.4 Method of Collaborative Filtering based on the Uncertain User Interest Cluster

The recommender system is a mainly used in environments where used want to purchase items based on their interest such as electronic commerce. With the help of recommendation system user take the decision easily about what music is to listen, what news to read, what items to be bought. The collaborative filtering recommendation system suffers from the problem of data scalability so Xiang. et. al [4] proposed a clustering based collaborative filtering recommendation method where clustering is done on the basis of user uncertain features. The uncertain interest of users is considered because computer logs take down data that have uncertain features and these features are solved with the help of clustering algorithms. Weighted factor is used to measure the quality of clustering result, which further helps in making a new improve method of collaborative filtering recommender system. The experimental results show that the proposed method outperforms in comparison to traditional collaborative filtering methods.

3. RECOMMENDATION FOR WEB PAGES

As similar to item recommendation there were various algorithms has been proposed in recommendation of web pages. Different methods were proposed by a number of authors for the recommendation of web pages which is discussed below as follows:

3.1 Weighted Page Rank

Wenpu. et. al [12] proposed a weighted page rank algorithm for computing the page rank of web pages. In the proposed algorithm it assumes that the larger rank value is assigned to more popular pages and popularity of the page is depend upon the number of incoming and outgoing links. Instead of dividing the rank value based on the number of outgoing links on a page rank algorithm, Weight page rank algorithm computes the weight of a node according to their popularity. The important parameter of this algorithm is back links and forward links. Algorithm classified the pages into four categories based on the relevancy to a given query. These are very relevant pages (VR), relevant pages (R), weak relevant pages (WR), irrelevant pages (IR). The weighted page rank algorithm applies the relevant rule to find out the relevancy score of each page in the list of pages, which differentiate the WPR (Weight Page Rank) and PR (Page Rank). The experimental results concluded that weighted page rank algorithm gives more relevancy score and it has less Complexity ($< O(\log N)$) comparing to page rank algorithm ($O(\log N)$).

3.2 Flexi Rank algorithm

Hits algorithm [18] used for ranking the web page has one drawback, that they do not consider textual content, So Debajyoti. et. al [13] proposed a Flexi Rank algorithm, which focus on the syntactic classification of web pages. The Classification used in FlexiRank is not dependent on the semantics of page content and takes a proper class of pages for a given query based on the user requirement. Flexi rank classification of web pages includes index page, home page, article, advertisement pages etc. In place of working on a number of pages, FlexiRank give a way to rank fewest numbers of pages where changes in user query gives the changes in weighted, for example article type page authority weight is high, while for advertisement type pages hub weight is high. The main purpose of FlexiRank algorithm is to provide the flexibility to the user and provides flexibility in two areas namely property selection and finding the weight of selected properties. Flexi Rank provides accurate results along with flexibility because of the ease of change in weight of the parameter which is used for ranking like relevance weight, hub and authority weight and link analysis of pages.

3.3 Page Ranking Algorithm

Larry Page and Sergey Brin (Co Founder of Google) proposed a page rank algorithm [14] for computing the ranking of web pages. They proposed a page rank algorithm for their search engine and focused on the link structure of the web. The importance of any page is dependent on the appropriate rank score based on the number of inlinks of a page. As compared to the traditional method, the page rank also depends upon the back links along with inlinks. Inlinks from good pages carry more weight because, as the number of sites on a page increases, the importance of a page is also increased and if the sum of the rank of back links on any page is more, than that page have more page rank, compare to others. The page Rank

method can be best demonstrated by the figure below:

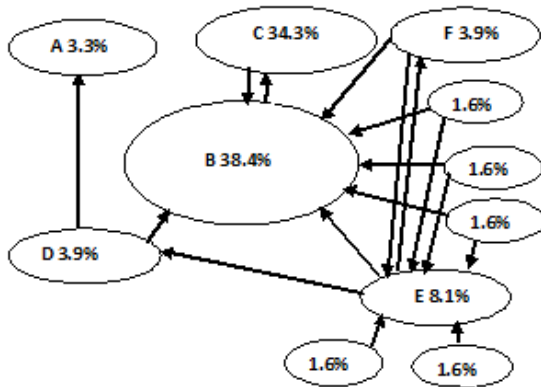


Fig1: Demonstration of Page Rank Algorithm.

As seen from the figure there are fewer links to C still Page C has a higher Page Rank than Page E because one link to C comes from an important page (B) and hence is of high value.

One of the most popular techniques which help in page rank computation is the breadth first search, which is used to figure out the structure of the network. Other than breadth first search technique, sparse matrix generally used for less memory storage without affecting the ranking of the matrix and compressed row vector which helpful to accelerate the process of multiplying the matrix are used. However by using compressed row vector only L addition and L multiplication is required in place of N addition and 2N multiplication, where L and N are the size of a row vector.

Power iteration and Monto Carlo are the two basic methods which is used for computing the page rank value. Power iteration [14] computes the page rank value in an iterative manner and one of the most popular methods. For a small set of pages there is no problem in solving the equation system, but in case of a more number of web pages, which contains billions of documents, not easy to compute the page rank score by an inspection method, in such situation power iteration method is proving to be more promising one for calculating page rank.

Another method for calculating the page rank is proposed by N. Osipova. et. al [15] named as the Monte Carlo method. The proposed algorithm considers the information on all pages during the simulation runs along with information of last visited pages. The Monte Carlo has several advantages beyond the power iteration method, it shows better performance after first iteration with higher accuracy for relatively important pages, and in addition continuous page rank update and parallel implementation are the two other advantages of Monte Carlo. There are several variants of Monte Carlo method proposed in the paper such as Monte Carlo end point with random start, Monte Carlo end point to cyclic start, Monte Carlo complete path, a Monte Carlo complete path stopping at dangling nodes and Monte Carlo complete path with random start. The following observation was observed by the author 1. Monte Carlo with cyclic start shows better results than the Monte Carlo with random start 2. Monte Carlo complete path stopping at dangling nodes outperforms Monte Carlo end point 3. The Monte Carlo complete path is more efficient than Monte Carlo end point. Although the Monte Carlo method is more efficient than power iteration method, still Google use the power iteration method for the page rank computation because as only after

the first iteration Monte Carlo gives the 1% relative error. Another factor is that, In Monte Carlo algorithm relative error will reduced on average by the factor $1/\sqrt{m}$ while power iteration method decrease exponentially with m, where m is iteration. However there are several drawbacks of Monte Carlo but still in the case where high precision is needed Monte Carlo is preferred because Monte Carlo provides simplicity in implementation and allows parallel implementation.

3.4 Recommendation of Web Pages for Online Users Using Web Log Data

For better understanding of the user, the concept of data mining is introduced, which is not got by the statistical user and online analytical processing (OLAP). Data mining is a process of discovering useful information from the large amount of data stored in databases, data warehouses and other information repositories, similarly finding the useful information from the web data is known as web mining. Web mining is classified into three categories web structure mining, web content mining and web usage mining. V. Citrate et.al [16] proposed a recommendation system based on the web usage mining which mines the user access patterns from usage logs to keep the record of every year's clicks and shows user interest in a web site. Web usage mining consists of three distinct phases: pre-processing, pattern discovery and pattern analysis. To increase the efficiency of mining, pre-processing is performed which takes the usage of data recorded in the server log, and convert them into data abstraction. Data abstraction is necessary for the pattern discovery, which achieved by extracting, decomposing, combining and deleting raw data. Data mining technique applies to the pre-processed logs data in pattern discovery phase, to identify the some useful pattern. In pattern analysis phase the main aim is to analyze some of mode rule that have exhumed, to find out the patterns and interesting rules. This paper focused on the web user clustering and recommendation for personalizing the users. The fuzzy clustering is used to handle the ambiguity in the data during pattern discovery phase and LCM (longest common subsequence algorithm) [17] is used for recommendation which classify the current user activities. The recommendation given by authors can help website owners to provide personalized service to users for their effective browsing.

4. Recommendation of Movie

Various methods are discussed below related to recommendation of movie by different researchers. In this section we discuss the different methods used for the movie recommendation.

4.1 ERPM

One of the novel method of recommendation is collaborative filtering (CF) which is classified into two categories namely user based CF and item based CF, but there is a certain limitation in both of classification. The user based CF, lacks in scalability and item based CF suffers from the data sparsity problem and fails in understanding the dynamic changes of the relationship between items. Jian. et.al [5] overcomes the above mentioned problem by proposing a method based on probability model named as ERPM (Easy Recommendation based on Probability Model). EPRM generate the prediction more efficiently and precisely as compare to collaborative filtering algorithm and also saves the storage dramatically. Instead of comparing user or item similarity, EPRM find out

the probable rating with the help of probability model, in which the probability of each user with respect of rating assigned to the movie and the probability of each movie with respect of rating assigned by different users is computed and with the help of these calculated probabilities, predicted rating is found. For checking the quality of recommendation, mean absolute error (MAE) is used, which gives a better prediction accuracy of recommendation than CF algorithm.

4.2 Contextual Walk

The Recommendation System is useful in solving the problem of information overloading, but most of the existing recommended system only focussed on recommendation through rating, not by the context. The Context can be time and location of the recommendation, actor, director, writer in the movie etc. So, Toine Bogers [6] proposed a context walk recommendation algorithm that includes different type of contextual information. The main aim of proposing this algorithm is to solve the difficulty arises during the processing of contextual data such as contextual information is difficult to collect and difficult to produce a computable formalization of contextual information. Toine applies the context walk algorithm on the datasets of movie where context is connected by a link that produce a contextual graph on which random walk is applied. The advantage of the context walk is that it can support many recommendation tasks with the same random walk model without retaining the information such as recommending movie for the group of users or tag recommendation.

4.3 Typically based Collaborative Filtering Recommendation

Current CF suffers from such problem of data sparsity, recommendation inaccuracy and big error in prediction. To deal with these problems Yi Cai [7] proposed a novel method named TyCo which uses the idea of Object typicality from cognitive psychology, where neighbour of users were found based on user typicality degrees in user groups. The TyCo method cluster all items into several item groups, and create a user group corresponding to each item group. These user groups contain multiple users having a different typicality degree in each of the user groups. With the help of user typicality matrix, user's similarity is measured, in all user groups, and finds the neighbors of each user. Based on the neighbors rating of user on the items, the author predicts the unknown rating of user on an item. Tyco shows better performance in accuracy (improvement of 6.35%), sparse training data (improvement of 9.89%) and has lower time cost than the traditional CF algorithm in movie lens data sets apart from this the algorithm obtains more accurate prediction in comparison to less number of big – error prediction.

5. Tag Recommendation

The recommendation of the tag can be done using several methods which are discussed below as follows.

5.1 LDA

Now a day tagging system has a major contribution for recommendation on web. Tags basically help in searching the data and other tag's which is belonging to a topic can be recommended for the new resources. For recommending tags of resources Ralf. et.al [8] proposed an approach based on Latent Dirichlet Allocation (LDA). LDA is a generative model that shows why some parts of data are similar with the

help of observation explained by unobserved groups. The goal of this method is to overcome the cold start problem ("reduces the usefulness of tags in particular for resources annotated only by a few users") for tagging new resources. LDA takes three input parameters the number of terms to represent latent topics, the number of latent topics to represent a document, and the overall number of latent topics to be identified in the given corpus. This method shows better precision and recall in comparison of association rules, for a high threshold value precisions increases and recall decreases in LDA.

5.2 Personalized Tag Recommendation

The Tag recommendation helps in bridging the semantic gap between human and features of media object by allowing user to add more tags which provide a feasible solution for content based multimedia information retrieval. A Tag based recommendation proposed by Jun. et. al. [9] is used for retrieving information in an easy and convenient way. By using the social relationship information of the user, an online social network is produced based on a network the author proposed a topology where nodes are used for characterizing the user's social influence. The recommendation is performed based on tagging history and the latent personalized preference, which is learned from whose have a most influenced in the user's social network. The experiment is performed on large scale real world data and result shows that the proposed method can outperform the non-personalized global co-occurrence method and other two state-of-art-personalized methods using social networking.

6. Recommendation of Others

In this Section method for the recommendation of the Venue and Video are discussed.

6.1 Venue Recommendation

Noulas. et. al [10] proposed a random walk method for ranking the venues. Previously proposed collaborative filtering (CF) algorithms have problems in working with mobility data while for online recommendation scenario collaborative algorithm is best. This random walk approaches overcome the problem of relationship between check-in, social and spatial data with others. It has linked structure of connected items in which every item has transitioned probabilities, which helps for the random walker to choose the nodes. Random walker moves from one node to another node according to transition probabilities and stays on every node for different amount of time, after that steady state is reached. Probabilities at steady state is a steady state probability and these probabilities are the output of the Random walk model. Now random walk starts from a node and find out the constant probabilities for a node, if found then return to the target node. Node those are closer to target node rank high compare to other nodes. Rank each place in decreasing order of steady state probabilities. Random walk performs shows 5-18% improvement in accuracy compared to other algorithms.

6.2 Video Recommendation

With the increase in the number of videos in YouTube or any other site, the choice for the user to select videos of their interest also increases. Baluja. et. al [11] proposed a novel method named Adsorption which provides personalized video suggestion for users based on the analysis of the user video graph with the help of random walk. The adsorption algorithm is used for classification and learning, of the labelled object and a graph structure defines the universe of labels and

unlabeled objects. The proposed algorithm constructs a personalized page which provides user recommendations as per their viewing habits along with the latest and most popular videos. With the help of adsorption algorithm the author tries to improve the efficacy of suggestion in YouTube. The author also performs a recommendation test done in three months snapshots of the live data from YouTube.

7. Conclusion

Some measure is needed, which helps the users in finding the valuable information among all the available information in the world. One such measure among them is Recommendation which helps the users in finding data of their interest. These systems are popping up everywhere from movies to news, to travel and leisure etc. They provide the valuable personalized information that can greatly influence the way we use the web. Any recommendation system consists of two basic entities namely users and items .Which help in decision making and actually based on real activity. The Recommendation system doesn't use any guess mark but find prediction based on an objective reality and applied in almost every domain, like web pages and items, movie, video, tag recommendation, etc. This survey paper is primarily focusing on these domains and work done by different researchers regarding the recommendation, which further helps in decision making.

8. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed, valuable comments and constructive suggestions.

9. REFERENCES

- [1] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item Based Collaborating Filtering Recommendation Algorithms", ACM, 2001 Hong Kong.
- [2] Shiramshetty Gouthami, Golamari. Jose Mary and Pulluri Srinivas Rao, "Ranking Popular Items by Naive Bayes Algorithm", International Journal of computer science and information technology (IJCSI) Vol 4. No 1, Feb 2012, pp 147-163.
- [3] Gediminas Adomavicius, YoungOk Kwon, "Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 5, MAY 2012.
- [4] Xiang Cui, Guisheng Yin, "Method of collaborative filtering based on uncertain user interests cluster", JOURNAL OF COMPUTERS, VOL. 8, NO. 1, JANUARY 2013, PP 186-193.
- [5] Jian Chen, Jin Huang, Huaqing Min, "Easy Recommendation Based on Probability Model", Proceeding in Fourth International Conferences on Semantics, Knowledge and Grid, IEEE, 2008, pp 441-444.
- [6] Toine Bogers, "Movie Recommendation uses Random Walk Over Contextual Graph", 2010.
- [7] Yi Cai, Ho-fung Leung, Qing Li, Huaqing Min, Jie tang and Juanzi Li, "Typicality-based Collaborative Filtering Recommendation", IEEE TRANSACTION ON KNOWLEDGE AND DATA ENGINEERING, Jan 2013.
- [8] Ralf Krestel, peter Fankhauser, Wolfgang Nejdl, "Latent Dirichlet Allocation for tag Recommendation", ACM 2009.
- [9] Jun Hu, Bing Wang, Yu Liu, De-Yi Li, "Personalized Tag Recommendation Using Social Influence", JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 27(3): 527-540 May 2012.
- [10] Anastation Noulas, Salvatore Seellaoto, Neal Lathia, Cecilia Mascolo, "Random Walk Around the City :New Venue Recommendation in Location-Based Social Networks", 2012.
- [11] Shumeet Baluja, Rohan Seth, D. Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, Mohamed Aly, "Video Suggestion and Discovery for YouTube: Taking Random walks Through the View Graph", ACM, Beijing, China, April 2008.
- [12] Wenpu Xing and Ali Ghorbani, "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04), 2004 IEEE
- [13] Debajyoti Mukhopadhyay, Pradipta Biswas, "FlexiRank: An algorithm offering flexibility and accuracy for ranking the web pages". Springer-Verlag Berlin Heidelberg 2005, pp-308-313.
- [14] Catherine Benincasa, Adena Calden, Emily Hanlon, Matthew Kindzerske, Kody Law, Eddery Lam, John Rhoades, Ishani Roy, Michael Satz, Eric Valentine and Nathaniel Whitaker, "Page Rank Algorithm", 2006.
- [15] K. Avrachenkov, N. Litvak, D. Nemirovsky, N. Osipova, "Monte Carlo method in PageRank computation: when one iteration is sufficient". SIAM Journal on Numerical Analysis, Volume 45, Issue 2, 2007.
- [16] V. Chitraa, Antony Selvadoss Thanamani, "Recommendation of Web Pages for Online users using Web Log Data", International journal of science and research (IJSR), Jan 2013, pp 345-349.
- [17] Mehrdad Jalali, Norwati Mustapha, Ali Mamat, Md. Nasir B Sulaiman, 2009, A Recommender System for Online Personalization in the WUM Applications, Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II, San Francisco, USA pp 741- 746.
- [18] C. Ding, X. He, P. Husbands, H. Zha, and H. Simon, "Link Analysis: Hubs and Authorities on the World". Technical report: 47847, 2001.