Collaborative Filtering Based Recommender System For E- Commerce

[1]URVASHI CHITRANSH SHRIVASTAVA and [2]M.A. RIZVI

[1]M. tech National Institute of Technical Teachers Training and Research Bhopal (M.P.)

[2] Associate Professor National Institute of Technical Teachers Training and Research Bhopal (M.P.)
Email of Corresponding Author :- marizv@nitttrbpl.ac.in

Abstract

Collaborative filtering is a powerful technology which is used in this huge adaptive web technology to evaluate the opinions based on others mind or we can say that it is a technique to suggest, evaluate or filter the items or data according the mood or interest of others. Here in this paper we suggest a concept of a recommender system based on collaborative filtering for e-commerce or online shopping portal in which a probable suggestion is filtered for the user and a brief on collaborative filtering and previous work done.

Keywords : Ecommence, recommender system, collaborative filleming

Introduction

Now in this modern era of computer science and internet, everybody has option to choose best thing according to their mood or to go with others opinion with the help of internet. For example if we want to see a movie, we prefer the movie which has grater ratings or we look for a particular item every time when we have to purchase from any e- commerce. In last few year recommender system is comes in existence to reduce the widespread problem of finding the relevant items and to help people to manage things in this large amount of information and to manage the time[2,3,4,5]. Recommender system relies on ratings in other word it predict the rating of any items from given N items according to the probability, it suggest the item which is relevant to the user, according there is a study that can improve the recommendation system with prediction and accuracy.

Some studies shows that the quality of recommender system can be evaluated along dimensions number and relaying on accuracy are not enough to find most relevant item for a particular user[6,7]. Few studies[8,10,11,12,13,14] argued that the aim of recommender system is to provide a user with highly idiosyncratic or personalized item and more diverse result in more opportunities for user to have a relevant item on this basis some studies increase the diversity of recommendation sets for particular user, often measure by average difference or dissimilarity between the item that are to be recommended having maintain accuracy level for particular/individual user[8,13,14]. Here we are suggesting a recommender system in which we will identify the similar data set with the help of K means or nearest neighbor algorithm then after call some prediction method for relevant item suggestion.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature on traditional recommendation algorithms and the evaluation of recommendation quality. Section 3 describes our core

concept for alternative recommendation ranking techniques, such as prediction of item based on the individual customers mind.

Related work study :

Recommender systems are usually classified into three categories based on their approach to recommendation: content based, collaborative, and hybrid approaches[2]. **Recommender system of content based type suggests items similar to the items preferred in the past by an individual.** Collaborative filtered recommender system recommends the items based on the others same preferences users. And Hybrid combines both the technologies.

There exist many variations of neighbor based collaborative filtering techniques[9,15]. A neighborhood based CF can be of any types from user based or item based here in this concept we use user based IP techniques[1].
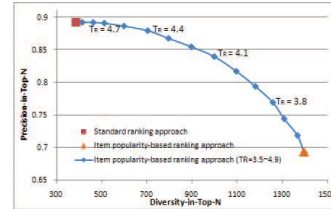
According to work of GediminasAdomavicius and YoungOkYown recommender systems predict unknown ratings on the basis of known ratings, using any neighborhood-based or matrix factorization CF techniques. Then, the predicted ratings are used to support the user's decision-making. In particular, each user u gets recommended a list of top-N items, LN(u), selected according to some ranking criterion. More formally, item ix is ranked ahead of item iy (i.e., ix □ iy) if rank (ix) < rank (iy), where rank: I -> R is a function representing the ranking criterion. The vast majority of current recommender systems use the predicted rating value as the ranking criterion:

rankStandard(i)=R*(u, i)-1.

The power of -1 in the above expression indicates that the items with highest-predicted (as opposed to lowest-predicted) ratings R*(u, i) are the ones being recommended to user. In their paper they refer to this as the standard ranking approach, and it shares the motivation with the widely used probability ranking principle in information retrieval literature that ranks the documents in order of decreasing probability of relevance [*8]. Using this approach they have suggested on item popularity based ranking which ranks the items directly based on their popularity from low to high, there popularity is represented by the number of known ratings which is held by each item there ranking function were as follows

*rank*ItemPop(*i*) = |*U*(*i*)|, where *U*(*i*) = {*u*□*U* |□*R*(*u, i*)}.

They compared the performance of the item popularity based ranking approach with the standard ranking approach using Movie Lens dataset and item-based CF, and they present this comparison using the accuracy-diversity plot in Fig.[1,16]

Panagiotis Symeonidis Alexandros Nanopoulos Apostolos Papadopoulos Yannis Manolopoulos contributes I their paper contributions of this paper are summarized as follows:

• To disclose the duality between users and items and to capture the range of the user's preferences, they suggest the application of an exact biclustering algorithm to the CF area.
• They propose a novel nearest-biclusters algorithm, which uses a new similarity measure that achieves partial matching of users' preferences.
• Their extensive experimental results have the effectiveness and efficiency of the proposed algorithm over existing techniques.[17]

Core concept :

Two types of CF algorithms have been proposed in the literature: (a) nearest-neighbors (a.k.a. memory-based) algorithms, which recommend according to the preferences of the same preferences users; and (b) model-based algorithms, which suggest by first developing a model of user ratings. Related research study concluded that nearest-neighbor algorithms are better if we talk about accuracy but that are not for handling a huge data. Therefore the requirement is to develop a technique of a recommendation system which recommend accurately or scalable too.

The proposed technique is a collaborative filter based recommender system design. In this presented work the hidden markov model and K-mean clustering is used to recommend the next user navigation. In this process we use three different algorithms in which we filter or evaluate the items using the opinions of user, cluster the items on price criteria, and find the next matched items/price etc. We also predict the probability of URL's, with the help of one model.

In this, the system produces the predictions or recommendations to a particular given user and suggest item to them. We used some of the known algorithms widely used in collaborative filtering. Or simply we suggest or recommend item by filtering the items from the cluster for which we used K means and then filtered that data or item through HMM which may provide accurate relevant thing to the user.

For this recommendation we select data set as a input which is taken from server side access log file after

that we store desired data input in table after selecting attributes, then we take a particular item's IP address or we can say we select input IP for particular item based recommendation. Our next move is to create a data list form data set of selected IP's followed by selection of clusters center then find out the distance of item from each cluster for finding out that the particular item belongs to which cluster or which cluster is nearest. We create observation matrix and transition matrix by using dataset list and clusters centers after this we find out the probability of an item which is to be suggested by our recommendation system. A step by step procedure is written here:
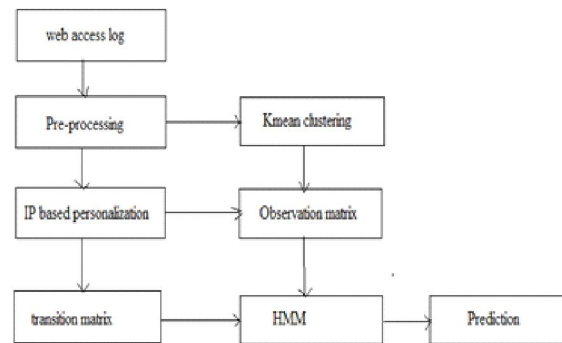
i) experimental procedure to be done for getting result:

We will create a portal to collect dataset or to have log files of a user. This portal is like other e-commerce portal, it has registration process for user authentication, and user can create their own login-id and set their own password. After this registration process, user can able to surf the items of their desire. This procedure is for first time user only. Once a user got registered we filtered the data on the content based it has the item which is last surfed by user after this we will choose the items from the clusters, which we create to find the same preference user item. As we study that the KNN is used to find out the similarity between the user or item or simply we can apply the KNN for Item Based similarity or User based similarity. We will use K- means to cluster the Item This algorithm is used to filter the numerical values. This helps to cluster the price criteria, produce different levels of price of items. With K-means we select the nearest matched price. It can fetch that price to their matched cluster and take an average of that and at the last we use Markov model to predict the item.

The filtering technique is works on the following basis.
1. Input log data: that is a server side web access log file which is produced as input to the system. System pre-process the log file and extract the important information from log file that is stored in a temporary data table for using with the recommender engine.
2. Data personalization: in order to recommend the product for a user that required to system have the prior information about the user behavior thus that is required to personalize the data first. Thus the IP based data grouping technique is utilized.

3. K-mean clustering: that approach help to identify the similar data which is belongs to the targeted user. After search the similar pattern from the data base the HMM (hidden markov model) is called for prediction.
4. HMM: the hidden markov model is a probabilistic technique for finding the next pattern from the previous sequences of navigations. Thus the user navigated web pages are used as the states of hidden markov model and using these states the transition matrix is prepared. For preparing the observation matrix the time domain of data is used. Therefore data is arranged in three different time groups.
5. The constructed transition matrix and the observation matrix is used with a decision function (π, A, B). That returns the next most probable URL for user.

Conclusion

These kinds of system are helpful for the user to find the relevant item based on their interest or they want to buy from e commerce. These systems are not only helpful for the customer or any individual rather then this they help the business by increasing their sales. Recommender systems are a crucial tool in E-commerce on the Web.

In this paper we suggest a concept based on collaborative filtering and here a minor part of content based filtering at the final step to recommend the most probable suggestion to the individual we can further improve the rank to improve the probable or relevant suggestion.

References

1. Gediminas Adomavicius, Member, IEEE, and YoungOk Kwon "Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL. *24* NO.5 YEAR (2012).
2. G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. On Knowledge and Data Engineering, *17(6)*, pp. 734-749, (2005).
3. D. Billsus and M. Pazzani, "Learning Collaborative Information Filters," Proc.Int'l Conf. Machine Learning, (1998).
4. B. M. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," ACM E-Commerce 2000 Conf., pp.158-167, (2000).
5. Y. Koren, "Collaborative filtering with temporal dynamics," Proc. of the 15th ACM SIGKDD Int'l Conf.

on Knowledge discovery and data mining, pp. 447-456, (2009).

6. S.M. McNee, J. Riedl, J. A. Konstan, "Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems," Conf. on Human Factors in Computing Systems, pp. 1097-1101, (2006).

7. J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, *22(1),* pp. 5-53, (2004).

8. K. Bradley and B. Smyth, "Improving Recommendation Diversity," Proc. of the 12th Irish Conf. on Artificial Intelligence and Cognitive Science, (2001).

9. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence, (1998).

10. E. Brynjolfsson, Y J. Hu, and, D. Simester, "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales", NET Institute Working Paper, (2007).

11. D. Fleder and K. Hosanagar, "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity", Management Science, *55(5),* pp. 697-712, (2009).

12. B. Smyth and P. McClave, "Similarity vs. Diversity," Proc. of the 4th Intl. Conf. on Case-Based Reasoning: Case-Based Reasoning Research and Development, (2001).

13. M. Zhang and N. Hurley, "Avoiding monotony: improving the diversity of recommendation lists," Proc. of the 2008 ACM Conf. on Recommender systems, pp. 123-130, (2008).

14. C.N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen, "Improving Recommendation Lists Through Topic Diversification," Proc. of the 14th Int'l World Wide Web Conf., pp. 22-32, (2005).

15. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," Proc. of the 10th Int'l WWW Conf., (2001).

16. A. Nakamura and N. Abe, "Collaborative Filtering Using Weighted Majority Prediction Algorithms," Proc. of the 15th Int'l Conf. Machine Learning, (1998).

17. Panagiotis Symeonidis Alexandros Nanopoulos Apostolos Papadopoulos YannisManolopoulos "Nearest Biclusters Collaborative Filtering" WEBKDD'06, August 20, 2006, Philadelphia, Pennsylvania, USA