

ALGORITHM FOR RESEARCH PAPER RECOMMENDATION SYSTEM

Megha Jain*

Abstract: In daily life we need many things to be searched over the internet, for search purpose there are many search engines are available. Whenever we search something we try to get the most relevant results. Recommender systems, generally uses COLLABORATIVE FILTERING ALGORITHM to fetch the results. This paper will discuss, how we can actually use different algorithms and developed such a powerful tool, as can provide more useful search results for the users. A technique will be using many things like users and their data, their queries, to find relevance between two users on the basis of their queries, as we can say if two users are searching for the same thing they may be interested in same field. by having this information we can suggest some more related items to other users. Discussion will be related to different algorithms like Check URL Algorithm, Subspace Clustering Algorithm.

Keywords: Collaborative filtering, Users' Habits, Potentially Interesting

1. INTRODUCTION

The main problem arises when we need to recommend something to the user, at that time we have to be very sure that the recommendation is relevant. for those reasons CF systems work by collecting user feedback in the form of ratings for items in a given domain and exploit similarities and differences among profiles of several users in determining how to recommend an item. Recommender systems help overcome information overload by providing personalized suggestions based on a history of a user's likes and dislikes. The filtering and retrieval of useful information has become a challenging consequence of the rapidly increasing world wide web. Search engines like Google, MSN have been extremely essential tools for such retrieval of information as per user query what is most valuable to him and they have been incredibly successful. One area in which the growth in online publication has been exponential is the research paper domain. The number of research papers published is increasing rapidly. That is why a researcher suffers a lot through a large quantity of research papers searched during the searching procedure to find an appropriate one.

Now days when we are dealing with a great use of internet the growing use of the web and the emergence of e-commerce has caused the collection of data to outpace the analysis necessary to extract useful information. Recommender systems were developed to help close the gap between information collection and analysis by filtering all of the available information to present what is most valuable to the user. One area of the web that has seen continued growth is the online publication of research papers. The number of research papers published continues to increase, and new technology has allowed many older papers to be rapidly digitized. A typical researcher must sift through a large quantity of articles manually, relying on keyword-based searches or paper citations to guide them. The search results of researchers with similar interests can help direct a more effective search, but the process of sharing search results is often too cumbersome and time consuming to be feasible. A recommender system can help by automatically recommending papers based on the preferences of other researchers with similar interests.

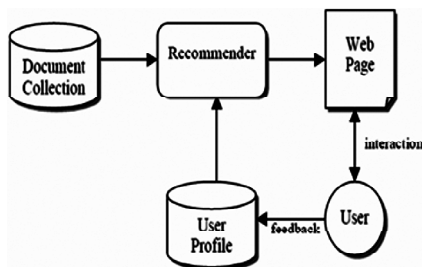


Figure 1

2. RELATED PREVIOUS WORK

Based upon the two major methods that is content based filtering and collaborative filtering a lot of work has been carried out in this field to develop the commercial recommender websites like amazon.com and many more but there are very few in research paper domain. CiteSeerx is an example of research paper recommender system but it has a limited domain research papers related to computer science field only. CiteSeerx uses content based filtering along with citations whereas Scienstein which is also a research paper recommendation system uses the hybrid technique. PRES (Personalized Recommender System) is one of the systems developed by using the content based filtering technique.

* Department of Computer Science & Information Technology Engineering, JSSATE, GBTU, Noida, India, E-mail: meghajain12@gmail.com

There are several techniques like probability (Bayesian probability networks) have also been useful to develop the user model or subspace clustering which is a data mining application is used to develop the user model in collaborative filtering and to optimize the performance of the recommender systems. The forthcoming collaborative filtering system improved their performance with the help of dependency networks and bipartite graph techniques.

There has been a lot work carried out in this era but still it seems to be out of minds of computer researchers. One of the reasons could be that its application is not visible from the front, if we talk about it as a quality of future coming search engines. But it is really very important field to work with. There is great scope of success in this field as it is still very new to the project developers and software engineers

3. PROBLEM

The main challenges during the development of the project are: First challenge is to deal with high dimensional data that is the number of research papers is much greater than the number of researchers working in a particular field. This leads to have an effective management of research papers as they are not limited and being published more and more per day.

The second challenge is to manage high sparseness (sparsity) of data means we need to keep track of record that which user has gone through which papers and form a large sparse matrix having much more number of zeros than ones as every user cannot read every paper published

4. ALGORITHMS

Check URL Algorithm: This algorithm will be basically used to create the database of the URL links of the research papers searched from the web by yahoo. However in initial cases when there will be no self database, this algorithm will be used to generate the ranking of the existing URL links as per user query and yahoo search relevance. For a user query first of all it will check whether the currently searched URL links pre-exists in database or not. If any of them exists previously then assuming that the current query is relevant to the query for which the pre-existing URL link queries has been searched by yahoo as yahoo might not refer the same links twice for two different

Subspace Clustering Algorithm: This algorithm has been mentioned in the literature survey chapter. This algorithm basically forms the subspace cluster from the given user-article matrix.

The runtime complexity of the algorithm is given as $O(m^2k)$.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a subspace clustering approach for recommender systems aimed at the research paper domain. A useful source of information when recommending research papers is the reading habits of other researchers who are interested in similar concepts. Thus, we adopted a collaborative filtering approach which allows us to use data collected from other researchers browsing patterns, and avoids issues with the interpretation of content. Such data consists of a small number of users (researchers) and a very large number of items (research papers). Our proposed approach takes advantage of the unique characteristics of the data in this domain and provides a solution which is fast, scalable and produces high quality recommendations. In order to improve the perceived quality and usefulness of the recommendations, a more sophisticated ranking scheme could be developed as an extension to the algorithm. The algorithm could also be extended to include the subjective user ratings rather than treating them as binary values and categorize recommendations as strong, mediocre and weak. A lot of work has been done in mixing the two models, content based filtering and collaborative filtering, to generate a hybrid model which tries to enhance the recommendation quality.

REFERENCES

- [1] May, R.M. 1997. "The Scientific Wealth of Nations", *Science*, **275(5301)**, pp. 793-796.
- [2] Torres, R. McNee, S.M. Abel, M. Konstan, J.A. and Riedl, J. 2004. "Enhancing Digital Libraries with TechLens", *Proceedings of JCDL'04*, pp. 228-236.
- [3] Pennock, D.M. Horvitz, E. Lawrence, S. and Giles, L.C. 2000. "Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach", in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence* (San Francisco).
- [4] Middleton, S.E. Shadbolt, N.R. and De Roure, D.C. 2004. "Ontological User Profiling in Recommender Systems", *ACM Transactions on Information Systems (TOIS)*, **22(1)**, pp. 54-88.
- [5] Fano, R.M. 1956. "Information Theory and the Retrieval of Recorded Information, in Documentation in Action", Shera, J. H. Kent, A. Perry, J. W. (Edts), New York: Reinhold Publ. Co., pp. 238-244.
- [6] Marshakova, I.V. 1973. "System of Document Connections based on References", *Nauchno-Tekhnicheskaya Informatsiya*, **2(6)**, pp. 3-8.
- [7] Small, H. 1973. "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents", *Journal of the American Society for Information Science*, **24**, pp. 265-269.
- [8] Giles, C.L. Bollacker, K.D. and Lawrence, S. 1998. "Cite Seer: An Automatic Citation Indexing System", in *Digital Libraries 98 - The Third ACM Conference on Digital Libraries*, pp. 89-98.

- [9] Meho, L.I. 2006. "The Rise and Rise of Citation Analysis", *Physics World*, <http://arxiv.org/abs/physics/0701012>.
- [10] Merton, R.K 1968. "The Matthew Effect in Science", *Science*, **159(3810)**, pp. 56-63.
- [11] Tagliacozzo, R. 1977. "Self-citations in Scientific Literature", *Journal of Documentation*, **33(4)**, pp. 251-265.