A Ranking Based Recommendation Framework for Optimized Search Result

[1] **Ch.Nagini,** [2] **M.Srinivasa Rao,** [3] **Dr. R.V.Krishnaiah**
[1]M.Tech Student, [2]Associate Professor, [3]Principal

[1]M.Tech Student,Dept. of CSE, DRK College of Engineering & Technology, Hyderabad, AP, India
[2]Associate Professor, Dept. of CSE, DRK College of Engineering & Technology, Hyderabad, AP, India
[3]Principal, Dept of CSE, DRK Group of Institutions, Hyderabad, Andhra Pradesh, India

Abstract

Web mining is resulted in millions of documents against a search query. Users cannot get intended results easily. To overcome these problem recommender systems came into existence. Theyhave become popular as they can provide the users with intended search results instead of returning huge number of documents. We believe that even the recommendations provided by such recommender frameworks can be improved further. Generally the results are presented based on certain ranking mechanism. In this paper we present "Rank Improvement" algorithm that will re-rank the results by classifying them into relevant groups. We built a prototype application that demonstrates the re-ranking process and search results optimization. The empirical results revealed that the proposed algorithm is effective and can be used in tandem with recommender systems in the real world.

Index Terms *–Recommendations, recommender systems, ranking, re-ranking*

I.INTRODUCTION:

World Wide Web (WWW) has become resource rich with multimedia content. Mining such content over Internet is known as web mining which applies various data mining techniques. Clustering is one of the data mining techniques that can group related objects so as to help making decisions. The web mining process can discover underlying patterns in web documents. These patterns result in business intelligence and help in taking well informed decision making. Thus the WWW has become a goldmine for the researchers. Many recommendation systems came into existence that use some sort of mining and also ranking. The ranking models are required to present results in more meaningful way. This is because the information retrieval systems return huge number of documents and finding the intended documents is time consuming and user needs to browse for related results. To overcome this problem many ranking models came into existence. These are basically machine learning algorithms used for ranking. They include SVM [1], [2], RankNet [3], RankBoost [4], LambdaRank [5] and ListNet [6]. The ranking algorithms improve the performance of the recommender systems over WWW.

The WWW has volumes of information that can help users to get required information throughrecommender systems and the results are presented through ranking. There are many search engines exist. They include Yahoo, Google, Bing, Ask and so on. It is said that search engines return huge number of documents. Out of them users only view first two pages only [7], [8]. It does mean that user is not willingto view all pages. He wants intended results in the first or second page of the search results. For this reason all search engines use certain mechanisms for ranking the search results. For instance, they use page ranking algorithm to present results in such a way that the most important documents appear in first and second pages. All these things come under web mining discipline. The web mining can be categorized into three types. They are web usage mining, web content mining and web structure mining[9]. Web content mining does mining on the content of web documents. The web structure mining focuses on the structural summary of web pages and web sites. Web usage mining on the other hand focuses on discoveringnavigational patterns of end users.

In this paper we continue our prior work that focused on giving recommendations through mining web graphs. The results of that framework are used as input to the proposed algorithm in this paper. Then the results (which are already ranked) are processed further for re-ranking. We built a prototype application that demonstrates the re-ranking process and search results optimization. The empirical results revealed that the proposed algorithm is effective and

can be used in tandem with recommender systems in the real world. The remainder of the paper is structured as follows. Section II reviews relevant literature. Section III provides information about the proposed algorithm. Section IV presents experimental results while section V concludes the paper.

II.RELATED WORK:

Web mining has been around for many years as researchers can discover knowledge from the huge volumes of data present in WWW. The information retrieval is done most of the time using search engines like Google, Bing, Yahoo, Ask and so on. The web crawlers can provide required information. However, the search results are very huge and users are not interested to view results which are not relevant. Instead they are interested to view few pages results that are presented as per ranking. The results in first few pages are expected to beintended results. Generally page ranking algorithm [10], [11] is used by search engines to improve search results. In other words, the search engines present results based on the ranking associated with the web pages being returned as part of query processing. Search engines maintain log of user searches that include user id, query issued by the user the results returned and the URLs clicked by user, rank associated, time at which the query was submitted. These query logs are widely used for research in order to unearth hidden patterns [12], [13], [14], [10]. The information which is in the form of query logs is used in many real time applications. Search engines and researchers use this information to learn about process of search and also improving search results [10], [11]. For automatically improvingquality ofsearch results an approach is presented in [13]. Extraction of relation between query logs and the user submitted queries is done in an accommodation system [14].

Of late, many recommendation systems came into existence [15], [16], [17], [18]. In our previous work we also proposed a recommender framework which is as shown in fig. 1.

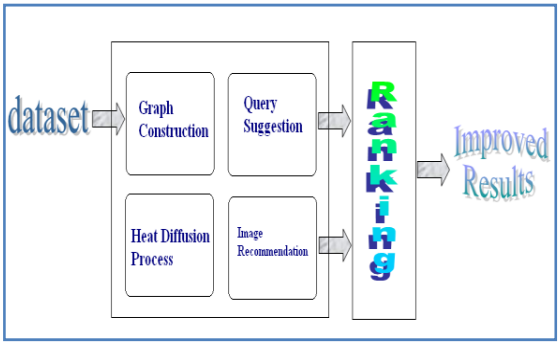

Fig. 1 –Recommender system presented in our previous work [19]

As can be seen in fig. 1, the framework for recommendations takes dataset as input and creates web graphs that represent documents over WWW. Then it performs query suggestion and image commendation based on the heat diffusion process. Finally the results are ranking and improved results are presented. In this paper, we further improve the framework with re-ranking process. This will result in more optimized search results that can help users to get more optimal results straight away in first two pages. For query suggestion AOL quick through dataset is taken while Flikr dataset is used for image recommendations.

III.PROPOSED FRAMEWORK and ALGORITHM:

In this paper we improve our framework presented in fig. 1 further to support re-ranking of results. We focused on re-ranking of ranked results of the recommender system proposed by us earlier. After re-ranking more optimized results are presented to user thus improving user satisfaction further.

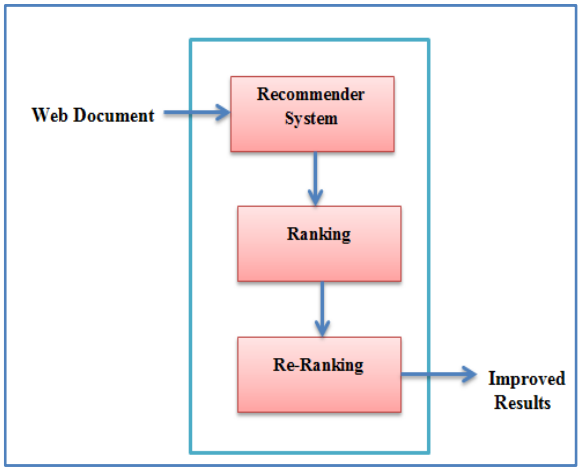

Fig. 2 –Illustrates Search Results Improvement through re-ranking

As can be seen in fig. 2, the recommender system has been improved with re-ranking process. More details about the recommender system with ranking algorithm for query suggestions and image recommendations as shown in fig. 1 can be found in [19]. The ranking results are presented to re-ranking algorithm known as "Rank Improvement". The algorithm is presented in fig. 3.

Algorithm: rank improve (Q,n)

Given: A set of n queries and corresponding clicked URLs stored array Q [qi, URL1……., URLm], 1≤i≤n
Output: A set C= {C1, C2…, Ck} of k query.
// Start of algorithm

K=0;
For (each query P in Q)
Set Clusterid (P) = NULL;
For (each P∈ Q with clustered (P) = NULL)
{
I =n, page= Q (n);
Clusterid (p) =ck;
Weight(X) =In (lenpar(X))
level(X)
Page_ rank(X) = (1-d) +d $\sum_{V \in B(X)} \frac{PR(v)}{Nv}$
New Page _rank(X) =Page_ rank + Weight(X)

While (i>1) and (Q [i/2] <New Page _rank(X)) do
{
Q[i] = Q [i/2];
I=i/2;
}
Q[i] =New Page _rank;
return true;
}
K=k+1;
}

Fig. 3 – Rank improvement algorithm

As can be seen in fig. 3 the algorithm takes a set of queries and corresponding ranked results as input and generates re-ranked results. The results thus can be found more optimal and user satisfaction increases.

IV.EXPERIMENTAL RESULTS:

Web built a prototype web application to demonstrate the proof of concept. In fact we improved our previous work [19] in order to re-rank results based on the algorithm presented in fig. 3. The environment used to build application is Java platform, a PC with 4 GB RAM, Core 2 Dual processor running in Windows 7 operating system. The IDE used to build the application is NetBeans. We evaluated the results through feedback system. We have taken feedback from 1000 end users. They used the proposed application and provided feedback with respect to their feedback. The average satisfaction level of users of both ranking and re-ranking is computed and the results are presented in fig. 4.

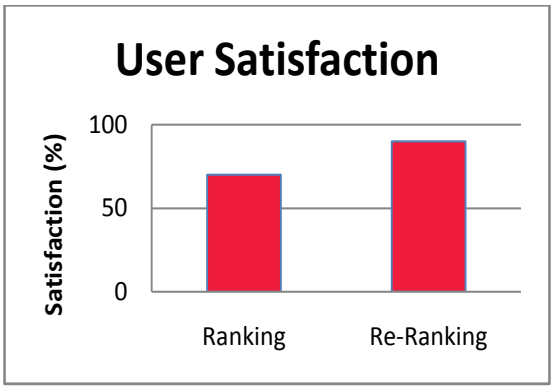

Fig. 4 –User satisfaction with two recommender systems

The recommender system which is based on mining web graphs and its improved form with re-ranking were evaluated through feedback given by 1000 users. As can be seen in fig. 4, the 70% satisfaction is expressed by users with prior system which presents recommendations with ranking. Around 90% satisfaction is expressed by users with improved form of recommendations with re-ranking algorithm. This shows that the algorithm proposed in this paper is able to improve the user satisfaction level further.

V.CONCLUSION:

Web mining returns useful information based on the prior actions of users. Thus recommendation systems came into existence to help users get compact results for decision making. However, the search results can be improved further so as to help the user to grasp the results with ease. Though the results of web mining througha recommender system are ranked already, they can be re-ranking for improving search results. In this paper we implemented an algorithm to improve the search results by re-ranking the results. We also built a prototype web application for testing the efficiency of the proposed algorithm. The empirical results are encouraging and the algorithm can be used in real world recommender systems for improving search results.

REFERENCES:

[1] C.J.C. Burges, R. Ragno, and Q.V. Le, "Learning to Rank withNonsmooth Cost Functions," Proc. Advances in Neural InformationProcessing Systems (NIPS '06), pp. 193-200, 2006.
[2] Z. Cao and T. Yan Liu, "Learning to Rank: From PairwiseApproach to Listwise Approach," Proc. 24th Int'l Conf. MachineLearning (ICML '07), pp. 129-136, 2007.

[3] C.J.C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N.Hamilton, and G. Hullender, "Learning to Rank Using GradientDescent," Proc. 22th Int'l Conf. Machine Learning (ICML '05), 2005.
[4] Y. Freund, R. Iyer, R.E. Schapire, Y. Singer, and G. Dietterich, "An Efficient Boosting Algorithm for Combining Preferences,"J. Machine Learning Research, vol. 4, pp. 933-969, 2003.
[5] R. Herbrich, T. Graepel, and K. Obermayer, "Large Margin RankBoundaries for Ordinal Regression," Advances in Large MarginClassifiers, pp. 115-132, MIT Press, 2000.
[6] T. Joachims, "Optimizing Search Engines Using ClickthroughData," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discoveryand Data Mining (KDD '02), pp. 133-142, 2002.
[7] A.Spink, D.Wolfram, B.J.Jansen, T.Saracevis,"Searching the Web: The public and their queries".jouurnel of the amercianSocitey for information Science and technology 52(3), 2001, 226-234.
[8] R.Cooley, B.Mobasher and J.Srivastava, "Web mining: Information and pattern discovery on the World Wide Web,".
In 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 97) 1997.
[9] M.H.Dunham, Companion slides for the text ," Data mining Introductory and advanced topics ".Prentice Hall,2002.
[10] IsakTaka, SarahZelikovitz, AmandaSpink,"Web Search log to Identify Query Classficationterms"Proceeding of IEEE International Conference on Information Technology (ITNG'07),pp:50-57,2008.
[11] L. Page,S.Brin ,R.Motwani,T.Winograd,"The pagerankcitration ranking :Bringing order to the Web ".Technical report ,Stanford Digital Libraries SIDL-WP-1990-0120,1999.
[12] J.Wen, J.Mie, and H.zhang," Clustering user queries of a search engine".In Proc of 10th International WWW Conference .W3C,2001.
[13] Thorsten joachims,"optimizing search engine using clickthough data" Proceeding of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, 2002, pp:133-142, New York.
[14] H.Ma, H.Yang,I.King,and M.R.Lyu,"learning latent semantic relations from clickthough data from query suggestion ".InCIKM'08:Proceeding oh the 17th ACM conference on information and knowledge management ,pages 709-708,New York,ny,USA,2008,ACM.
[15] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Trans. Information Systems, vol. 22, no. 1, pp. 5-53, 2004.
[16] T. Hofmann, "Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis," SIGIR '03: Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 259-266, 2003.
[17] T. Hofmann, "Latent Semantic Models for Collaborative Filtering," ACM Trans. Information Systems, vol. 22, no. 1, pp. 89-115, 2004.
[18] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," ACM Trans. Information Systems, vol. 22, no. 1, pp. 116-142, 2004.
[19] Ch.Nagini* M.SrinivasaRaoR.V.Krishnaiah**, "** A General Framework for Recommendations on the Web". Volume 3, Issue 5, May 2013

Author:

Nagini Chikati completed B.Tech (CSE) from Mother Teresa Institute of Science & Tech, AP, and pursuing M.Tech (CSE) in DRK College of Engg and Technology, JNTUH, Hyderabad, Andhra Pradesh, India. Her main research interest includes data mining, Databases and DWH.

M.Srinivasa Rao is working as an Associate Professor in DRK College of Engineering and Technology, JNTUH, Hyderabad, Andhra Pradesh, India. He is pursuing Ph.D in Information Security. He has completed M.Tech (C.S.E) from JNTUH. His main research interest includes Information Security and Computer Ad-Hoc Networks.

Dr.R.V.Krishnaiah, did M.Tech (EIE) from NIT Waranagal, MTech (CSE) form JNTU, ,Ph.D, from JNTU Ananthapur, He has memberships in professional bodies MIE, MIETE, MISTE. His main research interests include Image Processing, Security systems, Sensors, Intelligent Systems, Computer networks, Data mining, Software Engineering, network protection and security control. He has published many papers and Editorial Member and Reviewer for some national and international journals.