Volume 3, Issue 5, May 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A General Framework for Recommendations on the Web

Ch.Nagini* *Department of CSE & JNTUH India.*
M.Srinivasa Rao *Department of CSE & JNTUH India.*
R.V.Krishnaiah *Department of CSE &JNTUH India.*

Abstract - As the content over World Wide Web is rapidly increasing, web users get huge amount of data in response to their query. To overcome this problem recommendation techniques are being used to recommend books, images, music, movies etc. to the users. There are many recommender systems exist in the real world. This paper presents a novel approach in making recommendations. It builds a graph based on the dataset being used. Afterwards, it employs diffusion method to find the similarities among nodes in the graph to generate recommendations. This paper presents a general framework that is meant for making recommendations such as query suggestions and image recommendations. We built a prototype web application to test the efficiency of the proposed framework. The empirical results revealed that the proposed framework is effective and can be used in real world applications.

Index Terms – Data mining, recommendations, diffusion method, web graphs

1. INTRODUCTION

With the invention of Web 2.0, there is a sea change in the way web applications are built and operated. Users are now allowed to store and query less structured data over World Wide Web (WWW). Thus every day lot of information is being added to the web. Many technologies such as AJAX (Asynchronous JavaScript and XML) [1]came into existence to ensure rich user experience. However, mining such data has become a tedious task. To overcome this problem and to satisfy theneeds of Internet users many web applications are integrated with recommender systems that help users and businesses alike. Recommender systems have been well studied and widely used in the industry. Many recommender systems existed in the real world are based on collaborative filtering [2], [3], and [4]. Collaborative filtering is a technique in data mining which can predict the interest of the active user by considering the rating information from other items and users which are similar. This is based on the assumption that the preferences of others users and active user are same [5]. The collaborative filtering technique is widely used as it is simple and effective. In is used in real world web applications such as amazon.com for generating product recommendations. For movie recommendations it is being used in Netflix. The collaborative filtering needs the item and user data rankings in order to generate recommendations. However, such ranking data is sometimes not available especially when data is less structured and more diverse in nature. In spite of this, it is possible to model all kinds of datasets in the form of graphs. Therefore it is useful to build a graph recommendation algorithm, effective recommendations can be provided. There are many challenges when a general framework is built with that generate such graphs. Obtaining latent semantics and showing to user is not easy; query suggestion also has many issues such as natural language ambiguity that are to be addressed; another challenge is that users generally provide short queries as described in [6] and [7] and they are ambiguous naturally. As per the survey done in 2006 on web search logs around 30 percent of queries contained two words whereas around 20 percent of the queries contained single words. As users who search for information do the search as they do not have good knowledge in that area. Personalization feature is another challenge in many search cases. Amazon.com is one web application that adopted the personalization early while recommending products to its customers based on the history of previous purchases. Such adaptation helps in filtering out unnecessary information and gives useful information pertaining to the person in question. Another challenge is that, it takes lot of time to develop different recommender algorithms. Though many recommendation problems have common features, the general framework has to unify all recommendations considering number of dynamic parameters. In this paper the above challenges are analyzed and general framework is proposed for recommendations. Especially it focuses on the question suggestion and image recommendations with AOL click through data and also the Flicker image data respectively. A prototype web application is built to demonstrate the efficiency of the proposed frame work.

The rest of the paper is structured into some sections. Section 2 provides review of literature on relevant topics. Section 3 presents the proposed framework used for generating recommendations. Section 4 provides information about experimental results while section 5 concludes the paper.

2. PRIOR WORK

This section review literature on recommender systems in terms of collaborative filtering, AOL click-through data analysis, query suggestions, and image recommendations. The collaborative filtering is a process of obtaining recommendations based

on the active user's interest and also the interest or preferences of the other users. There are many researches with regard to collaborative filtering [8], [9]. There are specifically user based approaches which are also known as neighborhood based collaborative filtering approaches [10], [11]. There are some approaches specifically for item-based filtering [12], [13]. The algorithms used by them include VSS (Vector Space Similarity Algorithm) [10] and Pearson Correlation Coefficient Algorithm (PCC) [9]. Out of them PCC based one has provided higher performance since it makes use of differences in user rating style. There are many collaborative filtering approaches which are model-based. They include clustering models [14], aspect models [15], [16] and latent factor models [17]. Collaborative filtering needs user-item rating matrix. However, rating data is sometimes not available. Therefore for many tasks collaborative filtering cannot be applied directly. With regard to query suggestion, this feature is being used by many web search engines including Google, Ask, Yahoo, and Live Search. Query expansion [18], [19], query refinement[20], [8] and query substibution [21]are similar to query suggestions since all make use of user's search intentions. Two query recommendation methods are explored in [22] and [23]. The drawback of these methods is that they ignore the content in bipartite graph. By using session and clickthrough data Cao et al. [24] proposed a context aware query suggestion method. It provides suggestions by analyzing the data present in click-through content. Based on the hitting time on the click-through data Mei et al. proposed a query suggestion method [25]. There are many ranking models that can be used to improve search results. HITS [26], PageRank [27] are some of the examples. With respect to image recommendations, it is being used by many existing web applications like Photoree etc. Based on the users' preferences, the image recommender systems recommends images to end users. These recommender systems take rating from users initially. As users given response such as they like or dislike, the rating is built on top of it. Recently Yand et al. [28] employed context based image search using Flickr dataset. This impoved quality of image recommendations. Howver, it is computationally expensive as it is context based. Diffusing in huge datasets using this approach is very time consuming. Therefore the proposed system performs diffusion on the bipartite graph generated and the empirical results revealed that our approach is effective and efficient.

3. PROPOSED RECOMMENDATION FRAMEWORK

This section provides information about the proposed framework used for generating recommendations. It has many components involved. They are schematically presented in fig. 1.

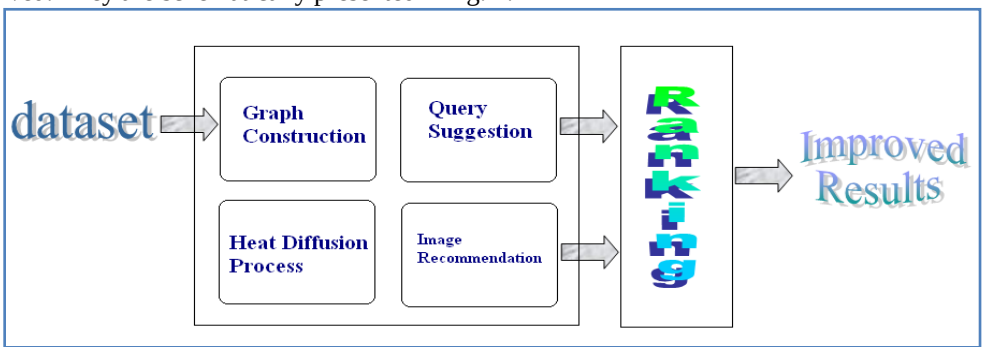

Fig. 1 –Proposed Framework for Recommendations

As can be seen in fig. 1, the proposed framework takes dataset as input. The datasets and experimental results are described in the next section. The framework has provision for query suggestions and image recommendations as of now. However, it can be extended further to support recommendations in other areas. The graph construction component constructs query – URL bipartite graphs which are used in heat diffusion process for both query suggestion and image recommendation algorithms. Once graph is constructed, the diffusion process starts which takes care of finding similar items and making recommendations. The diffusion results are used either by query suggestion component or image recommendation component based on the dataset provided by the user in the prototype web application. For the results of either operation, ranking is applied while presenting the final results to end user. The ranking is based on the feedbacks given by the users which is associated with dataset.

Query Suggestion Algorithm

Question suggestion algorithm needs bipartite graph as input and generated top-k queries based on the ranking associated with input dataset.

1. A converted bipartite graph G = (V+U V^2, E) consists of query set V^+ and URL set V^*. The two directed edges are weighted using the method introduced in previous section.
2. Given a query q in V^+ , a subgraph is constructed by using depth-first search in G.
   The search stops when the number of queries is larger than a predefined number.
3. As analyzed above, set α=1, and without loss of genarility, set the initial heat value of query q fq(0)=1 (the

choice of initial heat value will not effect the suggestion results). Start the diffusion processing $F(1) = e^{aR}f(0)$.

4. Output the Top-K queries with the largest values in vector f(1) as the suggestions

Fig. 2–Question Suggestion Algorithm

As can be seen in fig. 2, the query suggestion algorihm has four steps. It is similar to both query suggestion and image recommendations.

4. EXPERIMENTS AND RESUTLS

A web application is built to test the efficiency of the proposed framework. The application is made using software such as Tomcat, JDK and Net Beans. Struts framework is used for systematic developing using MVC (Model View Controller) architecture.

Datasets Collection

AOL click-through dataset is collected from Internet for testing query suggestions while Flikr image database is collected from Internet for testing generation image recommendations.

Results of Query Suggestions

The proposed framework is tested to find query suggestions using various test queries. The results with top 5 query suggestions are presented in fig. 3.

Testing Queries	Suggestions				
Michael Jordan	Nda	nike	Jorfan xi	air jordans	Michael Jordan bio
Java	Sun java	Java download	Java updates	Virtual machine	Sun microsystems
Apple	Itunes	ipod	quicktime	Apple ipod	Apple stores
Fitness	Exercise	Fitness magazine	Muscle and fitness	Mens fitness	Weight loss
Solar systems	Planets	jupiter	saturn	Neptune	Pluto
sunglasses	Chanel sunglasses	oakley	Maui jum sunglasses	Designer sunglasses	Oakley sunglasses
Flower delivery	Flowers	florist	Gift baskets	Cheap flowers	Proflowers
wedding	Wedding channel	Wedding dresses	The knot	Wedding plans	Wedding poems
astronomy	Apod	Star charts	planets	Solar system	Skyandtelescope
Real estate	Remax	Realtor	Homes for sale	Coldwell banker	Houses for sale

Fig. 3 –Results of Query Suggestions

As can be seen in fig. 3, top 5 suggestions are presented for all test queries. For instance the test query "java" generated sun java, java download, java updates, virtual machine and sun Microsystems as top 5 suggestions.

Results of Image Recommendations

The algorithm used for image recommendation is same as that of query suggestions. Only difference is that the input dataset is changed here. The dataset is collected from Flikr image database.

Seed Image

Recommendations

Fig. 4 – Results of Image Recommendations using Flikr Dataset

As can be seen in fig. 4, seed image and recommendations are presented. Theseed image is taken by the proposed framework and the diffusion results are taken by image recommendation algorithm in order to generate recommendations.

Seed Image

Recommendations

Fig. 5 – Results of Image Recommendations using Flikr Dataset

As can be seen in fig. 5, seed image and recommendations are presented. The seed image is taken by the proposed framework and the diffusion results are taken by image recommendation algorithm in order to generate recommendations.

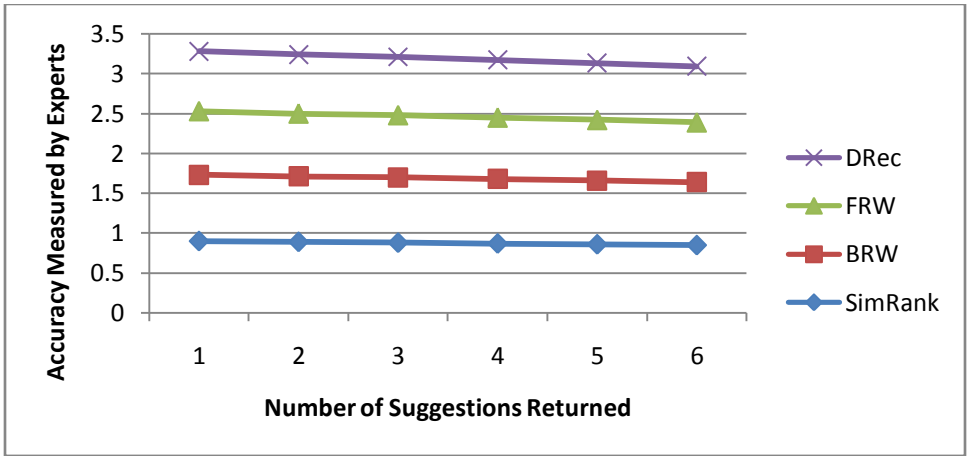

Fig. 6 – Accuracy Comparison Measured by Experts in Query Suggestions

As can be seen in fig. 6, it is evident that the proposed method (DRec) consistently outperforms other methods such as SimRank, BRW, and FRW.

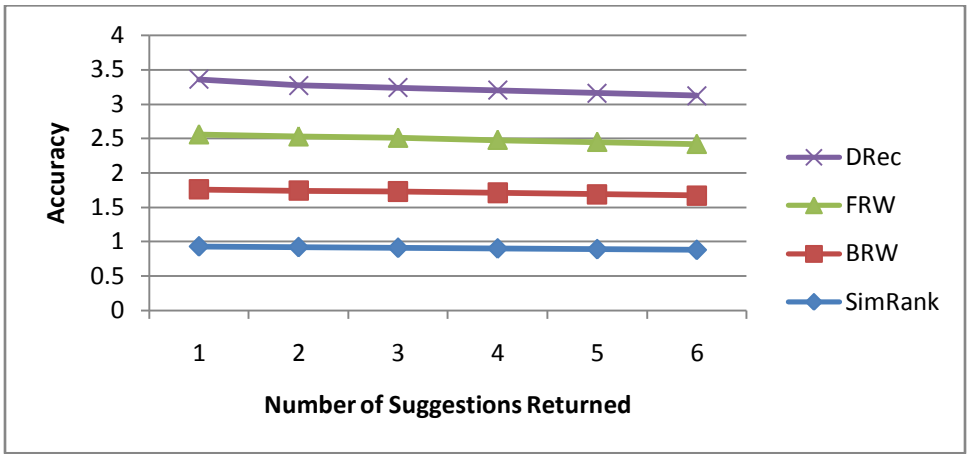

Fig. 7 –Accuracy Comparison Made by Experts in Image Recommendation

As can be seen in fig. 7, the DRec method (proposed) consistently outperforms the other methods such as SimRank, BRW, and FRW.

5. Conclusion

This paper presents a recommender system based on a general framework for generating various kinds of recommendations. For instance it supports query suggestions and image recommendations. It is based on the web graphs generated based on the data sets. After generating the graph, the proposed algorithm traverses, the graph using diffusion method to ascertain similarities among the items presented in the form of graph. Based on the similarities, it generates recommendations. In order to test the efficiency of the proposed framework we built a web based application that shows query suggestions and image recommendations in the proposed approach.

References

[1] http://www.w3schools.com/ajax/ (additional ref, not from base paper)

[2] 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.

[3] J.D.M. Rennie and N. Srebro, "Fast Maximum Margin MatrixFactorization for Collaborative Prediction," ICML '05: Proc. 22nd Int'l Conf. Machine Learning, 2005.MA ET AL.: MINING WEB GRAPHS FOR RECOMMENDATIONS 1063

[4] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization,"Advances in Neural Information Processing Systems, vol. 20,pp. 1257-1264, 2008.

[5] H. Ma, I. King, and M.R. Lyu, "Effective Missing DataPrediction for Collaborative Filtering," SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 39-46, 2007.

[6] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "Real LifeInformation Retrieval: A Study of User Queries on the Web," ACMSIGIR Forum, vol. 32, no. 1, pp. 5-17, 1998.

[7] C. Silverstein, M.R. Henzinger, H. Marais, and M. Moricz,"Analysis of a Very Large Web Search Engine Query Log,"ACM SIGIR Forum, vol. 33, no. 1, pp. 6-12, 1999.

[8] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations:Item-to-Item Collaborative Filtering," IEEE Internet Computing,vol. 7, no. 1, pp. 76-80, Jan./Feb. 2003.

[9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl,"Grouplens: An Open Architecture for Collaborative Filtering ofNetnews," CSCW '94: Proc. ACM Conf. Computer SupportedCooperative Work, 1994.

[10] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis ofPredictive Algorithms for Collaborative Filtering," Proc. 14th Conf.Uncertainty in Artificial Intelligence (UAI), 1998.

[11] E. Auchard, "Flickr to Map the World's Latest Photo Hotspots,"Proc. Reuters, 2007.

[12] M. Deshpande and G. Karypis, "Item-Based Top-n Recommendation,"ACM Trans. Information Systems, vol. 22, no. 1, pp. 143-177,2004.

[13] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-BasedCollaborative Filtering Recommendation Algorithms," WWW '01:Proc. 10th Int'l Conf. World Wide Web, pp. 285-295, 2001.

[14] A. Kohrs and B. Merialdo, "Clustering for Collaborative FilteringApplications," Proc. Computational Intelligence for Modelling, Controland Automation (CIMCA), 1999.

[15] T. Hofmann, "Collaborative Filtering via Gaussian ProbabilisticLatent Semantic Analysis," SIGIR '03: Proc. 26th Ann. Int'l ACMSIGIR Conf. Research and Development in Information Retrieval,pp. 259-266, 2003.
[16] T. Hofmann, "Latent Semantic Models for Collaborative Filtering,"ACM Trans. Information Systems, vol. 22, no. 1, pp. 89-115,2004.
[17] J. Canny, "Collaborative Filtering with Privacy via FactorAnalysis," SIGIR '07: Proc. 25th Ann. Int'l ACM SIGIR Conf.Research and Development in Information Retrieval, pp. 238-245, 2002.
[18] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized QueryExpansion for the Web," SIGIR '07: Proc. 30th Ann. Int'l ACMSIGIR Conf. Research and Development in Information Retrieval, pp. 7-14, 2007.
[19] H. Cui, J.-R. Wen, J.-Y.Nie, and W.-Y. Ma, "Query Expansion byMining User Logs," IEEE Trans. Knowledge Data Eng., vol. 15, no. 4,pp. 829-839, July/Aug. 2003.
[20] R. Kraft and J. Zien, "Mining Anchor Text for QueryRefinement," WWW '04: Proc 13th Int'l Conf. World Wide Web,pp. 666-674, 2004.
[21] R. Jones, B. Rey, O. Madani, and W. Greiner, "Generating QuerySubstitutions," WWW '06: Proc. 15th Int'l Conf. World Wide Web,pp. 387-396, 2006.
[22] R.A. Baeza-Yates, C.A. Hurtado, and M. Mendoza, "QueryRecommendation Using Query Logs in Search Engines," Proc.Current Trends in Database Technology (EDBT) Workshops, pp. 588-596, 2004.
[23] G. Dupret and M. Mendoza, "Automatic Query RecommendationUsing Click-Through Data," Proc. Int'l Federation for InformationProcessing, Professional Practice in Artificial Intelligence (IFIP PPAI),pp. 303-312, 2006.
[24] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li,"Context-Aware Query Suggestion by Mining Click-Through andSession Data," KDD '08: Proc. 14th ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining, pp. 875-883, 2008.
[25] Q. Mei, D. Zhou, and K. Church, "Query Suggestion Using HittingTime," CIKM '08: Proc. 17th ACM Conf. Information and KnowledgeManagement, pp. 469-477, 2008.
[26] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment,"J. ACM, vol. 46, no. 5, pp. 604-632, 1999.
[27] S. Brin and L. Page, "The Anatomy of a Large-Scale HypertextualWeb Search Engine," Computer Networks and ISDN Systems, vol. 30,nos. 1-7, pp. 107-117, 1998.
[28] Y.-H. Yang, P.-T.Wu, C.-W.Lee, K.-H.Lin, W.H. Hsu, and H.Chen, "ContextSeer: Context Search and Recommendation atQuery Time for Shared Consumer Photos," Proc. 16th ACM Int'lConf. Multimedia, pp. 199-208, 2008.

AUTHORS

	Nagini Chikati is student of DRK College of Engineering and Technology, Hyderabad, AP, INDIA. She has received B.Tech Degree computer science and engineering, M.Tech Degree in computer science and engineering. 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 (Ph.D) is working as Principal at DRK INSTITUTE OF SCINCE & TECHNOLOGY, Hyderabad, AP, INDIA. He has received M.Tech Degree EIE and CSE. His main research interest includes Data Mining, Software Engineering.

