ISSN:2348-2079

Volume-5 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Mining web graphs for recommendations identification for query suggestion with efficient search re-ranking

[1]Mr R.Navinkumar MCA.,M.Phil., Assistant professor,
[2]Ms A.Vijayalakshmi Final MCA,
Department of MCA, Nandha Engineering College (Autonomous), Erode-52.
E-mail ID: navinsoccer07@gmail.com, vijimca29@gmail.com

Abstract—As the exponential explosion of various contents generated on the Web, Recommendation techniques have become increasingly indispensable. Innumerable different kinds of recommendations are made on the Web every day, including movies, music ,images, books recommendations, query suggestions, tags recommendations, etc. No matter what types of data sources are used for the recommendations, essentially these data sources can be modeled in the form of various types of graphs. In this paper, aiming at providing a general framework on mining Web graphs for recommendations, 1) we first propose a novel diffusion method which propagates similarities between different nodes and generates recommendations; 2) then we illustrate how to generalize differentiate commendation problems into our graph diffusion framework. The proposed framework can be utilized in many recommendation on the World Wide Web, including query suggestions, tag recommendations, expert finding, image recommendations, image annotations, etc. The experimental analysis on large data sets shows the promising future of our work.

Index Terms—Recommendation, diffusion, query suggestion, image recommendation.

1 INTRODUCTION

With the diverse and explosive growth of Web information, how to organize and utilize the information effectively and efficiently has become more and more critical. This is especially important for Web 2.0 related applications since user-generated information is more freestyle and less structured, which increases the difficulties mining useful information from these data sources. In order to satisfy the information needs of Web users and improve the user experience in many Web applications, Recommender Systems, have been well studied in academia and widely deployed in industry.

Typically, recommender systems are based on Collaborative which is technique that automatically predicts the interest of an active user by collecting rating information from other similar users or items. The underlying assumption of collaborative filtering is that the active user will prefer those items which other similar users prefer. Based on this simple but effective intuition, collaborative filtering has been widely employed in some large, well-known commercial systems, including product recommendation at Amazon,1 movie recommendation at Netflix,2 etc. Typical collaborative filtering algorithms require a user-item rating matrix which contains user-specific rating preferences to infer users' characteristics. However, in most of the cases ,rating data are always unavailable since information on the Web is less structured and more diverse .Fortunately, on the Web, no matter what types of data sources are used for recommendations, in most cases, these data sources can be modeled in the form of various types of graphs. If we can design a general graph recommendation algorithm, we can solve many recommendation the Web.

However, when designing such a framework for recommendations on the Web, we still face several challenges that need to be addressed. The first challenge is that it is not easy to recommend semantically relevant results to users. Take Query Suggestion as an example, there are several outstanding issues that can potentially degrade the quality of there commendations, which merit investigation. The first on this the ambiguity which commonly exists in the natural language. Queries containing ambiguous terms may confuse the algorithms which do not satisfy the information needs of users. Another consideration, as reported in and, is that users tend to submit short queries consisting of only one or two terms under most circumstances ,and short queries are more likely to be ambiguous. Through the analysis of a commercial search engine's query logs recorded over three months in 2006, we observe that19.4 percent of Web queries are single term queries, and further 30.5 percent of Web queries contain only two terms.

Third, in most cases, the reason why users perform a search is because they have little or even no knowledge about the topic they are searching for. In order to find satisfactory answers, users have to rephrase their queries constantly. The second challenge is how to take into account the Personalization feature. Personalization is desirable for many scenarios where different users have different information needs. As an example, Amazon.com has been the early adopter of personalization technology to recommend products to shoppers on its site, based upon their previous purchases. Amazon makes an extensive use of collaborative filtering in its personalization technology. The adoption of personalization will not only filter outer relevant information to a person, but also provide more specific information that is increasingly relevant to aperson's interests.

The last challenge is that it is time consuming and inefficient to design different recommendation algorithms for different recommendation tasks. Actually, most of these are commendation problems have some common features, where a general framework is needed to unify there commendation tasks on the Web. Moreover, most of existing methods are complicated and require to tune a large number of parameters.

In this paper, aiming at solving the problems analyzed above, we propose a general framework for the recommendation the Web. This framework is built upon the heat diffusion on both undirected graphs and directed graphs ,and has several advantages.

1. It is a general method, which can be utilized to many recommendation tasks on the Web.
2. It can provide latent semantically relevant results tot he original information need.
3. This model provides a natural treatment for personalized recommendations.
4. The designed recommendation algorithm is scalable to very large data sets.

The empirical analysis on several large scale data sets (AOL Click through data and Flicker image tags data) shows that our proposed framework is effective and efficient for generating high-quality recommendations. The rest of the paper is organized as follows. We review related work in Section 2. Section 3 presents the diffusion models on both undirected graphs and directed graphs. In Section 4, we demonstrate the empirical analysis of our models and recommendation algorithms on several diversified data sources.

2 RELATED WORK

Recommendation on the Web is a general term representing a specific type of information filtering technique that attempts to present information items (queries, movies, images, books, Web pages, etc.) that are likely of interest to the users. In this section, we review several work related to recommendation, including collaborative filtering, query suggestion techniques, image recommendation methods, and click through data analysis.

The existing system is aimed at solving the recommendation problem and proposed a general framework for the recommendations on the Web. This framework is built upon the heat diffusion on both undirected graphs and directed graphs, and has several advantages. It is a general method, which can be utilized to many recommendation tasks on the Web.It can provide latent semantically relevant results to the original information need.This model provides a natural treatment for personalized recommendations.The designed recommendation algorithm is scalable to very large data sets.

The existing system introduced a graph diffusion model for recommendation. It shows how to convert different Web data sources into correct graphs in the models; It conducts several experiments on query suggestions since Query Suggestion is a technique widely employed by commercial search engines to provide related queries to users' information need. **Graph construction** and **Query Suggestion algorithm** is used to provide related queries.

All the existing system approaches are implemented in proposed system also. Graph construction and Query suggestion algorithm is implemented. Similarity measure between two queries is also considered. Abbreviation based query suggestion is also considered. Personalized recommendations are given importance. Previous search query words are also taken into query suggestion calculation.

3 DIFFUSION ON GRAPHS

In this section, we first introduce a novel graph diffusion model based on heat diffusion. This model can be applied to both undirected graphs and directed graphs. We then present how to infer the parameter based on the graph structure. Last, we analyze the computational complexity of our model.

3.1 Heat Diffusion

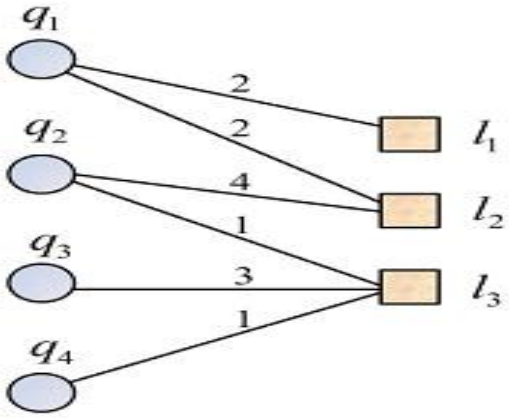
Heat diffusion is a physical phenomenon. In a medium ,heat always flows from a position with high temperature to a position with low temperature. Recently, heat diffusion-based approaches have been successfully appliedin various domains such as classification and imensionality reduction problems [6]. Lafferty and Lebanon approximated the heat kernel for amultinomial family in a closed form, from which

great improvements were obtained over the use of Gaussian or linear kernels. In Kondor and Lafferty proposed the use of a discrete diffusion kernel for categorical data, and showed that the simple diffusion kernel on the hypercube can result in good performance for such data. Belk in and Niyogi employed a heat kernel to construct the weight of a neighborhood graph, and apply it to a reduction algorithm in [6]. Yan get al. proposed a ranking algorithm known as the Diffusion Rank using heat diffusion process; simulations showed that it is very robust to Web spamming. In this paper, we use heat diffusion to model the similarity information propagation on Web graphs. In Physics, the heat diffusion is always performed on metric manifold with initial conditions. However, it is very difficult to represent the Web as a regular some try with a known dimension. This motivates us to investigate the heat flow on a graph. The graph is considered as an approximation to the underlying manifold, and so the heat flow on the graph is considered as an approximation to the heat flow on the manifold.

3.2 Diffusion on Undirected Graphs

Consider an undirected graph G ¼ ðV ;EÞ, where V is the vertex set, and V ¼ fv1; v2; . . . ; vng. E ¼ fðvi; vjÞ j there is an edge between vi to vjg is the set of all edges. The edge ðvi; v jÞis considered as a pipe that connects nodes vi and vj. The value fiðtÞ describes the heat at node vi at time t, beginning from an initial distribution of heat given by fið0Þ at time zero. fðtÞ denotes the vector consisting of fiðtÞ.We construct our model as follows: suppose, at time t, each node i receives an amount Mði; j; t;_tÞ of heat from it neighbor j during a time period _t. The heat Mði; j; t;_tÞ should be proportional to the time period _t and the heat difference fjðtÞ _ fiðtÞ. Moreover, the heat flows from node j to node i through the pipe that connects nodes i and j. Based on this consideration, we assume that M ði; j; t; _tÞ ¼_ðfjðtÞ _ fiðt ÞÞ_t, where _ is the thermal conductivity—the heat diffusion coefficient. As a result, the heat difference at node i between time t þ _t and time t will be equal to the sum of the heat that it receives from all its neighbors. The curve for the amount of heat at each node with time is shown in Fig. 1b.

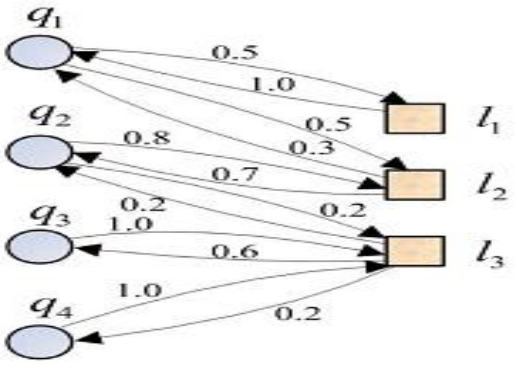
We can see that, as time passes, the heat sources nodes 1 and 2 will diffuse their heat to nodes 3, 4,and 5. The heat of nodes 3, 4, and 5 will increase respectively, and the trends of their heat curves are the same since these three nodes are symmetric in this graph. Another example is shown in Fig. 1c. Initially, at time zero, suppose node 1 is given 4 units of heat, then the vector of ð0Þ equals ½4; 0; 0; 0_T. The related heat curve is shown in Fig. 1b. We can see that the node 2, the closest node to the heat source, gains more heat than other nodes. This also indicates that if a node has more paths connected to the heat source, it will potentially obtain more heat. This is a perfect property for recommending relevant nodes on a graph.



3.3 Diffusion on Directed Graphs

The above heat diffusion model is designed for undirected graphs, but in many situations, the Web graphs are directed, especially in online recommender systems or knowledge sharing sites. Every user in knowledge sharing sites typically has a trust list.

The users in the trust list can influence this user deeply. These relationships are directed since user a is in the trust list of user b, but user b might not
be in the trust list of user a. At the same time, the extent of trust relations is different since user ui may trust user uj with trust score 1 while trust user uk only with trust score0.2. Hence, there are different weights associated with there relations. Based on this consideration, we modify the heat diffusion model for the directed graphs as follows. Consider a directed graph G ¼ fV ;E; Wg, where V is the



3.4 Random Jump

The heat can only propagate through the links that connect nodes in a given graph, but in fact, there are random relations among different nodes even if these nodes are not connected. For an example, in the click through data, people of different cultures, genders, ages, and environments, mayimplicitly link queries together, but we do not know these latent relations. Another good example is the trust relations in a social network. On online social network sites, users always explicitly state the trust relations to other users. Actually, there are some other implicit hidden trust relations among these users that cannot be observed. Hence, to capture these relations, we propose to add auni form random relation among different nodes. More specifically, let _ denote the probability that such phenom enahappen, and ð1 _ _Þ is the probability of taking a"random jump." Without any prior knowledge, we set g ¼ 1n 1, where g is a uniform stochastic distribution vector,1 is the vector of all ones, and n is the number of nodes .Based on the above consideration, we modify our model to

MA ET AL.: MINING WEB GRAPHS FOR RECOMMENDATIONS 1055

Fig. 1.Two simple heat diffusion examples on an undirected graph.
(a) Example 1. (b) Curve of heat change with time.
(c) Example 2.
(d) Curve of heat change with time.

3.5 Complexity Analysis

When the graph is very large, a direct computation of e_R is very time consuming. We adopt its discrete approximation to compute the heat diffusion equation where P is a positive integer. In order to reduce the computational complexity, we introduce two techniques: the size of Web information is very large ,the graph built upon the Web information can become extremely large. Then, the complexity OðPMÞ is also too high, and the algorithm becomes time consuming and inefficient to get a solution. To overcome this difficulty, we first extract a sub graph starting from the heat sources. Given the heat sources, the sub graph is constructed by using depth-first search in the original graph. The search stops when the number of nodes is larger than a predefined number. Then, the diffusion processes will be performed on this sub graph efficiently and effectively. Generally, it will not decrease the qualities of the heat diffusion processes since the nodes too far away from the heat sources are normally not related to the sources.

4 EMPIRICAL ANALYSIS

We introduced our graph diffusion models for recommendations. In this section, 1) we show how to convert different Web data sources into correct graphs in our models; and 2) we conduct several experiments on query suggestions, and image recommendations.9

4.1 Query Suggestion

Query Suggestion is a technique widely employed by commercial search engines to provide related queries to users' information need. In this section, we demonstrate how our method can benefit the query suggestion, and how to mine latent semantically similar queries based on the users' information need.

4.2 Image Recommendation

Finding effective and efficient methods to search and retrieve images on the Web has been a prevalent line of research for a long time. The situation is even tougher in the research of Image Recommendation. In this section, web resent how to recommend related images to the given images using Flicker data set.

4.2.1 Personalized Image Recommendation

Personalization is becoming more and more important in many applications since it is the best way to understand different information needs from different users. Actually, our method can be easily extended to the personalized image recommendations. In the query suggestions conducted in Section 4.1 and image suggestions performed in this section, we only employ one node (either a query or an image) as the heat source. In the personalized image recommendations, we can set all the images submitted by a specified user as the heat sources, and then start the diffusion process. This ensures that the suggested images are of interests of this user. In order to evaluate the quality of our personalized image recommendation method, we create 10 groups: Given1, Given2, , and Given10, where Given1 means in this group, all the users only submitted 1 images. We then randomly select 50 users from the user list for each group, hence totally we have 500 users. For each of these users, we start the diffusion processes once with the submitted images as the heat sources. After generating the results, we ask three experts to rate these recommendations. We again define a 6-point scale (0, 0.2, 0.4, 0.6, 0.8, and 1) to measure the relevance between the testing images and the suggested images, in which 0 means "totally irrelevant"while 1 indicates "entirely relevant."

4.1.2 Efficiency Analysis

As analyzed in Section 3.5, our algorithm is very efficient, and can be applied to large data sets. Our algorithm hassimilar complexity with FRW and BRW methods. The computation time for the query suggestion task of these three method (subgraph size is 5,000) is normally around 0.10 seconds. However, Sim Rank is not very efficient since it

has a high computational complexity. It takes more than 15 minutes to compute a query suggestion task in our data set.

5 CONCLUSION

In this paper, we present a novel framework for recommendation large scale Web graphs using heat diffusion .This is a general framework which can basically be most of the Web graphs for the recommendation tasks ,such as query suggestions, image recommendations, personalized recommendations, etc. The generated suggestion sare semantically related to the inputs. The experimental analysis on several large scale Web data sources shows the promising future of this approach.

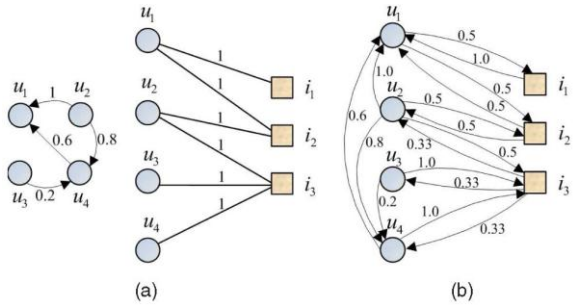
6 FUTURE WORK

6.1 Search Results Improvement

We list the heat values of the suggested queries. These values not only can be used in query suggestions, but also are very informative in the advertisement when customers bid for query terms. Actually, If this order is incorporated into the original results, the search results can be greatly improved since they are the representations of the implicit votes of all the search users. In the future, we plan to compare this ranking method with other previous Web search results ranking approaches.

6.2 Social Recommendation

Since our model is quite general, we can apply it to more complicated graphs and applications, such as Social Recommendation problem. Recently, as the explosive growth of Web 2.0 applications, social-based applications gain lots of traffics on the Web. Social recommendation, which produces recommendations by incorporating users' social network information, is becoming to be an indispensable feature for the next generation of Web applications. The social recommendation problem includes two different data sources, which are social network and user item relation matrices. We can see that in the social network graph, there are trust scores between different users, while in the user-item relation matrix, binary relations connect users and items. We can convert these two graphs into a single and consistent one, as shown in Fig. 10b.With the constructed graph, for each user (heat source),we can start the diffusion process and then recommend the Top-N items to this user. In fact, during the diffusion process on the graph as shown in Fig. 10b, there are two possible ways to diffuse heat from users to items.



The first out is within the user-item bipartite graph, which captures the intuition that similar users will purchase (or view) similar items. The second route is passing through the social network graph, which reflects the social interactions and influences between users. Hence, our DR ec diffusion method naturally fuses these two data sources together for social recommendations. We plan to conduct this social recommendation research in the future

REFERENCES

[1] E. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," SIGIR '07: Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 19-26, 2006.

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

[3] R. TiberiBaeza-Yates and A. Tiberi, "Extracting Semantic Relations from Query Logs," KDD '07: Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 76-85, 2007.

[4] R.A. Baeza-Yates, C.A. Hurtado, and M. Mendoza, "Query Recommendation Using Query Logs in Search Engines," Proc. Current Trends in Database Technology (EDBT) Workshops, pp. 588- 596, 2004.

[5] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," KDD '00: Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 407-416, 2000.

[6] M. Belkin and P. Niyogi, "LaplacianEigenmaps for Dimension- ality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

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

[8] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," Computer Networks and ISDN Systems, vol. 30, nos. 1-7, pp. 107-117, 1998.

[9] J. Canny, "Collaborative Filtering with Privacy via Factor Analysis," SIGIR '07: Proc. 25th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 238-245, 2002.

[10] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware Query Suggestion by Mining Click-Through and Session Data," KDD '08: Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 875-883, 2008.

[11] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 7- 14, 2007.

[12] N. Craswell and M. Szummer, "Random Walks on the Click Graph," SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 239-246, 2007.

[13] H. Cui, J.-R. Wen, J.-Y.Nie, and W.-Y. Ma, "Query Expansion by Mining User Logs," IEEE Trans. Knowledge Data Eng., vol. 15, no. 4, pp. 829-839, July/Aug. 2003.

[14] A.S. Das, M. Datar, A. Garg, and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering," WWW '07: Proc. 16th Int'l Conf. World Wide Web, pp. 271-280, 2007.

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