

An Enhanced Web Graph Search Engine Based on User Profiles and Clickthrough Patterns

Rushikesh M. Shete

*M. E. Computer Science Engineering Department
Sant Gadge Baba Amravati University, India
Sipna COET, Amravati, India*

Prof. V. S. Gulhane

*Associate Professor
Department of Computer Science & Engineering
Sipna COET, Amravati, India*

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. In this paper, aim is to providing a general framework on user profiles & clickthrough patterns. Firstly proposing a method which propagates similarities between different nodes i.e. from user profiles and generates recommendations from clickthrough data. The proposed framework can be utilized in many recommendation tasks on the World Wide Web, including query suggestions, tag recommendations, expert finding, image recommendations etc. The experimental analysis on large data sets will show the promising future of our work.

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

1. INTRODUCTION

A key factor for the popularity of today's Web search engines is the friendly user interfaces they provide. 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 [1]. This is especially important for Web 2.0 related applications since user-generated information is more freestyle and less structured, which increases the difficulties in 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 and widely deployed in industry.

Web mining is the technique of data mining. In this report we propose the web graph mining. The directed links between pages of the World Wide Web are described by the web graph. A graph, in general, consists of several vertices, some pairs connected by edges. In a directed graph, edges are directed lines or arcs. The web graph is a directed graph, whose vertices correspond to the pages of the WWW, and a directed edge connects page X to page Y if there exists a hyperlink on page X, referring to page Y. The degree distribution of the web graph strongly differs from the degree distribution of the classical random graph model. The web graph is an example of a scale-free network. The web graph is used for computing the Page Rank of the WWW pages.

Recommender systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that a user would give to an item, such as music, books, or movies or social element (e.g. people or groups) they had not yet considered, using a model built from the characteristics of an item or the

user's social environment [4], [6]. Typically, recommender systems are based on Collaborative Filtering which is a technique that automatically 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 or movie recommendation etc. Typical collaborative filtering algorithms require a user-item rating matrix which contains user-specific rating preferences to infer users' characteristics [7].

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 clickthrough data analysis.

2.1 Collaborative Filtering

Neighborhood-based and model-based are two types of collaborative filtering [5]. The most analyzed examples of neighborhood-based collaborative filtering include user-based approaches and item-based approaches. User-based approaches predict the ratings of active users based on the ratings of their similar users, and item-based approaches predict the ratings of active users based on the computed information of items similar to those chosen by the active user. Recently, several matrix factorization methods have been proposed for collaborative filtering. These methods all focus on fitting the user-item rating matrix using low-rank approximations, and use it to make further predictions [7].

2.2 Query Suggestion

In order to recommend relevant queries to Web users, a valuable technique, query suggestion, has been employed by some prominent commercial search engines, such as Yahoo!, Live Search, Ask, and Google.

The goal of query suggestion is similar to that of query expansion, query substitution, and query refinement which all focus on understanding users' search intentions and improving the queries submitted by users. Query suggestion is closely related to query expansion or query substitution, which extends the

predicts the interest of an active user by collecting rating information from other similar users or items. The underlying assumption of collaborative filtering is original query with new search terms to narrow down the scope of the search [4]. But different from query expansion, query suggestion aims to suggest full queries that have been formulated by previous users so that query integrity and coherence are preserved in the suggested queries. Query refinement is another closely related notion, since the objective of query refinement is interactively recommending new queries related to a particular query.

2.3 Clickthrough Data Analyses

In the field of clickthrough data analysis, the most common usage is for optimizing Web search results or rankings. Web search logs are utilized to effectively organize the clusters of search results by 1) learning "interesting aspects" of a topic and 2) generating more meaningful cluster labels[5],[6]. A ranking function is learned from the implicit feedback extracted from search engine clickthrough data to provide personalized search results for users. Besides ranking, clickthrough data is also well studied in the query clustering problem. Query clustering is a process used to discover frequently asked questions or most popular topics on a search engine. This process is crucial for search engines based on question answering. A typical relationship can be learning from clickthrough data is that "BMW" is a child of "car." The method proposed can extract attributes such as "capital city" and "President" for the class "Country," or "cost," "manufacturer" and "side effects" for the class "Drug." The method initially relies on a small set of linguistically motivated extraction patterns applied to each entry from the query logs, and then employs a series of Web-based precision-enhancement filters to refine and rank the candidate attributes [3].

2.4 Image Recommendation

Besides query suggestion, another interesting recommendation application on the Web is image recommendation. Image recommendation systems, like Photoree focus on recommending interesting images to Web users based on users' preference. Normally, these systems first ask users to rate some images as they like or dislike, and then recommend images to the users based on the tastes of the users. While framework proposed in this report, by diffusing on the image-tag bipartite graph with one or more images, we can accurately and efficiently suggest semantically relevant non-personalized or personalized images to the users [6]. Work which will be implemented in future is a general framework which can be effectively,

efficiently, and naturally applied to most of the recommendation tasks on the Web.

3. Analysis of Problem

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. If we can design a general graph recommendation algorithm, we can solve many recommendation problems on the Web. For recommendations on the Web, we have to face several challenges while designing framework that need to be addressed.

The first case is it is not easy to recommend latent semantically relevant results to users. Take Query Suggestion as an example; there are several outstanding issues that can potentially degrade the quality of the recommendations, which merit investigation. The first one is 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 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.

The second case is how to take into account the personalization feature. Personalization is needed for many scenarios where different users have different information needs.

The third case is that it is time consuming and inefficient to design different recommendation algorithms for different recommendation tasks. Actually, most of these recommendation problems have some common features, where a general framework is needed to unify the recommendation tasks on the Web.

4. Proposed Work And Objectives

In this proposed work, we are aiming at solving the problems analyzed above; we propose a general framework for the recommendations on the Web. This framework is built upon the user profiles and the clickthrough data patterns, and has several objectives.

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 to the original information need.
3. It can provide a long query to the user within short time.
4. It can provide the specific query suggestion to the user.

The empirical analysis on several large scale data sets (AOL clickthrough data and Flickr image tags data) shows that our proposed framework is effective and efficient for generating high-quality recommendations. Flowchart in Fig. 1 will show the execution of the process of this work

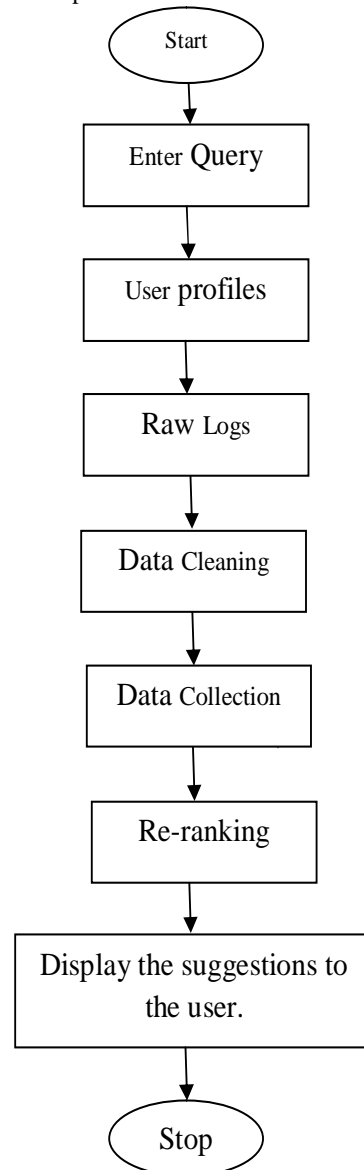


Figure 1. Flowchart Showing Basic Steps

4.1 Query Suggestion:

Query Suggestion is a technique widely employed by commercial search engines to provide related queries to users' information need [6], [7]. When user enters his/ hers query to the search engine, it suggests the query as per the users requirement. When query is suggested by the search engine user can select the query and can surf it. Query suggestion utilize the query logs from user profiles. From user

profiles required information is sorted from previous related queries. Query can be sometimes different from the users expectations. Query suggestion is necessary because from clicked data from previous user it does not give critical information. Effective query suggestion need the users query intent and then suggests query. It may help user to retrieve useful information. The aim of query suggestion is to use past information from previous user profiles.

In this proposed work, the user profiles play an important role. From user profiles raw data or information is collected to form raw log. Raw log puts the data together from different user profiles who have previously searched the related data.

4.2 Clickthrough Data:

Clickthrough data includes different steps, data cleaning & data collection. From user profiles raw logs are maintained. These raw logs contain all the information related to user entered query. From this raw log only quality data is extracted. This extracted information contains all the information related to the users wishing information. After cleaning data all the data is collected and database is created. This step may reduce the size of data to the great extent.

Clickthrough data helps to recognize the patterns of data that whether it is information, image etc. this process is mining the data from web. In web graph all the data is stored in the form of nodes. Each node will proceed to the required information of the user. In web graph each node will be the link to the inserted query. Query can be compared with the link at the node, if it finds the require information it may be suggested to the user with the rank. [3], [4]

After data collection each link to the required information will be provided the rank. This rank can be a re-rank. Re-ranking is done on the basis of priority, which will be decided by the entered query. If the link in the graph is having quality information the rank of the node will be the highest. If the information is not up to the mark, rank will be lowest.

The next step is, user will select appropriate query suggested according to the requirement.

4.3 Image Recommendations:

Now a day's users frequently search for the images. Image gives the clear view to the thought of mind. The situation is even tougher in the research of Image Recommendation. In this we will collect data that is images from some famous website; and then we will apply image recommendation framework on it [6].

Basically, the graph construction for image recommendation is similar to that of query suggestion as explained in previous section. The only difference

will be instead of query here images and tags are the nodes. Image can be recommended on the file type as required to the user.

5. CONCLUSION

In this report, a novel framework for recommendations can be generated on large scale Web graphs using user profiles and clickthrough data. This is a general framework which can basically be adapted to most of the Web graphs for the recommendation tasks, such as query suggestions, image recommendations, etc. Related to the entered inputs suggestions can be generated. The experimental analysis on several large scale Web data sources shows the promising future of this approach. This model in general can be applied to more complicated graphs.

ACKNOWLEDGEMENT

I express my sincere gratitude to **Dr. A. D. Gawande** Head of the Department, Computer Science & Engineering and guide **Prof V. S. Gulhane** for providing their valuable guidance and necessary facilities needed for the successful completion of this seminar throughout. I am also obliged to our principal, **Dr. S. A. Ladhake** who has been a constant source of inspiration throughout.

REFERENCES

- [1] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "Real Life Information Retrieval: A Study of User Queries on the Web," ACM SIGIR Forum, vol. 32, no. 1, pp. 5-17, 1998.
- [2] 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.
- [3] 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.
- [4] D. Shen, M. Qin, W. Chen, Q. Yang, and Z. Chen, "Mining Web Query Hierarchies from Clickthrough Data," AAAI '07: Proc. 22nd Nat'l Conf. Artificial Intelligence, pp. 341-346, 2007.
- [5] Hao Ma, Irwin King, Michael Rung-Tsong Lyu, "Mining Web Graphs for Recommendations" IEEE Transactions on Knowledge Data Engineering, vol 24 June 2012.
- [6] 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.
- [8] J.-T. Sun, D. Shen, H.-J. Zeng, Q. Yang, Y. Lu, and Z. Chen, "Web- Page Summarization Using Clickthrough Data," SIGIR '05: Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 194-201, 2005.

IJERT