AN OBJECT ORIENTED APPROACH TO MINING WEB GRAPHS FOR RECOMMENDATIONS

T Murali MOhan,
Kaushik college of Engineering.
Gambhiram, Visakhapatnam,
AP, INDIA

P Suresh babu,
Associate Professor,
Kaushik college of Engineering.
Gambhiram, Visakhapatnam,
AP, INDIA.

N T Radha,
Associate Professor.
Kaushik college of Engineering.
Gambhiram, Visakhapatnam,
AP, INDIA.

T V Satya SHella -
Kaushik college of Engineering.
Gambhiram, Visakhapatnam,
AP, INDIA.

Abstract—Web mining is the application of data mining techniques to extract knowledge from Web. Web mining has been explored to a vast degree and different techniques have been proposed for a variety of applications that includes music, images, books recommendations, query suggestions, etc. In this paper, we highlight the significance of studying the evolving nature of the Web personalization. Web usage mining is used to discover interesting user navigation patterns and can be applied to many real-world problems, such as improving Web sites/pages, making additional topic or product recommendations, user/customer behavior studies, etc. The proposed framework can be utilized in many recommendation tasks on the World Wide Web, including query suggestions, image recommendations, etc. The experimental analysis on large datasets shows the promising future of our work.

Keywords- Web Mining, World wide web, Web Personalization;

I. INTRODUCTION

With the dramatically quick and explosive growth of information available over the Internet, World Wide Web has become a powerful platform to store, disseminate and retrieve information as well as mine useful knowledge. Due to the properties of the huge, diverse, dynamic and unstructured nature of Web data, Web data research has encountered a lot of challenges, such as scalability, multimedia and temporal issues etc. As a result, Web users are always drowning in an “ocean” of information and facing the problem of information overload when interacting with the web. A user interacts with the Web, there is a wide diversity of user’s navigational preference, which results in needing different contents and presentations of information.

To improve the Internet service quality and increase the user click rate on a specific website, thus, it is necessary for a Web developer or designer to know what the user really wants to do, predict which pages the user is potentially interested in, and present the customized Web pages to the user by learning user navigational pattern knowledge [1,2,3].

In this paper, aiming at solving the problems analyzed above, we propose 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: (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) This model provides a natural treatment for personalized recommendations; (4) The designed recommendation algorithm is scalable to very large datasets. The empirical analysis on several large scale datasets (AOL clickthrough data and Flickr image tags data) shows that our proposed framework is effective and efficient for generating high quality recommendations.

II. BACKGROUND

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 recommendations including collaborative filtering, query suggestion techniques, image recommendation methods, and clickthrough data analysis.

Two types of collaborative filtering approaches are widely studied: neighborhood-based and model-based. The neighborhood-based approaches are the most popular prediction methods and are widely adopted in commercial collaborative filtering systems [4,5]. The most analyzed examples of neighborhood-based collaborative filtering include user-based approaches [6,7] and item-based approaches[8,9].

In the model-based approaches, training datasets are used to train a predefined model. Examples of model-based approaches include the clustering model, the aspect models [10,11] and the latent factor model [12] presented an algorithm for collaborative filtering based on hierarchical clustering.

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!3, Live Search4, Ask5 and Google6. However, due to commercial reasons, few public papers have been released to reveal the methods they adopt. The goal of query suggestion is similar to that of query expansion [13], query substitution [14] 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 original query with new search terms to narrow down the scope of the search. 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.

In the field of clickthrough data analysis, the most common usage is for optimizing Web search results or rankings [15]. In [16], 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. In [17], a ranking function is learned from the implicit feedback extracted from search engine clickthrough data to provide personalized search results for users. Besides query suggestion, another interesting recommendation application on the Web is image recommendation. Image recommendation systems, like Photoree8, 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. In the academia, few tasks are proposed to solve the image recommendation problems since this is a relatively new field and analyzing the image contents is a challenge job.

In general, comparing with previous work, our work is a general framework which can be effectively, efficiently and naturally applied to most of the recommendation tasks on the Web.

III. PROPOSED SYSTEM ARCHITECTURE AND IMPLEMENTATION

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. Clickthrough data record the activities of Web users, which reflect their interests and the latent semantic relationships between users and queries, as well as queries and clicked Web documents as shown in Fig 1.

A. Existing System

The last challenge 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. Moreover, most of existing methods are complicated and require tuning a large number of parameters.

Disadvantages: It is becoming increasingly harder to find relevant content and what user recommends the actual thing.

B. Proposed System:

In order to satisfy the information needs of Web users and improve the user experience in many Web applications, Recommender Systems. This is a 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 the proposed method consists of two stages: generating candidate queries and determining “generalization/specialization” relations between these queries in a hierarchy. The method initially relies on a small set of linguistically motivated extraction patterns applied to each entry from the query logs, then employs a series of Web-based precision-enhancement filters to refine and rank the candidate attributes.

Advantages:

a) It is a general method, which can be utilized to many recommendation tasks on the Web.
b) It can provide latent semantically relevant results to the original information need.
c) This model provides a natural treatment for personalized recommendations.
d) The designed recommendation algorithm is scalable to very large datasets.
C. Algorithm:

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>A converted bipartite graph ( G = (V + U \ V' \ E) ) consists of query set ( V + ) and URL set ( V' ).</td>
</tr>
<tr>
<td>2.</td>
<td>Given a query ( q ) in ( V + ), a sub graph is constructed by using depth-first search in ( G ). The search stops when the number of queries is larger than a predefined number.</td>
</tr>
<tr>
<td>3.</td>
<td>As analyzed above, set ( \alpha = 1 ), and without loss of generality, set the initial heat value of query ( q ) ( f_0(q) = 1 ) (the choice of initial heat value will not affect the suggestion results). Start the diffusion process using ( f(1) = e\alpha R(0) ).</td>
</tr>
<tr>
<td>4.</td>
<td>Output the Top-K queries with the largest values in vector ( f(1) ) as the suggestions.</td>
</tr>
</tbody>
</table>

IV. IMPLEMENTATION

In this paper our proposed method consists following modules.

a) Posting the opinion
b) Image Recommendation Technique
c) Rating Prediction
d) Ranking Approach
e) Collaborative Filtering
f) Query Suggestion

Posting the opinion: In this module, we get the opinions from various people about business, e-commerce and products through online. The opinions may be of two types. Direct opinion and comparative opinion. Direct opinion is to post a comment about the components and attributes of products directly. Comparative opinion is to post a comment based on comparison of two or more products. The comments may be positive or negative.

Image Recommendation Technique: Another interesting recommendation application on the Web is image recommendation. 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. However, the quality of recommendations can be evaluated along a number of dimensions, and relying on the accuracy of recommendations alone may not be enough to find the most relevant items for each

User, these studies argue that one of the goals of recommender systems is to provide a user with highly personalized items, and more diverse recommendations result in more opportunities for users to get recommended such items. With this motivation, some studies proposed new recommendation methods that can increase the diversity of recommendation sets for a given individual user. They can give the feedback of such items.

Rating Prediction: First, the ratings of unrated items are estimated based on the available information (typically using known user ratings and possibly also information about item content) using some recommendation algorithm. Heuristic techniques typically calculate recommendations based directly on the previous user activities (e.g., transactional data or rating values). For each user, ranks all the predicted items according to the predicted rating value ranking the candidate (highly predicted) items based on their predicted rating value, from lowest to highest (as a result choosing less popular items).

Collaborative Filtering: 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.

Ranking Approach: Ranking items according to the rating variance of neighbors of a particular user for a particular item. There exist a number of different ranking approaches that can improve recommendation diversity by recommending items other than the ones with topmost predicted rating values to a user. A comprehensive set of experiments was performed using every rating prediction technique in conjunction with every recommendation ranking function on every dataset for different number of top-N recommendations.

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. This extends the original query with new search terms to narrow down the scope of the search. 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.

V. RESULT ANALYSIS

In this paper we define a 6-point scale (0, 0.2, 0.4, 0.6, 0.8, and 1) to measure the relevance between the testing queries and the suggested queries, in which 0 means “totally irrelevant” while 1 indicates “entirely relevant”. The average values of evaluation results are shown in Fig. 2. We observe that, when measuring the results by human experts, our DRec algorithm increases the accuracy for about 19.81%, 13.0% and 7.5% comparing with the SimRank, BRW and FRW algorithm. As shown in Fig. 3, we observe that, when evaluating using ODP database, our proposed DRec algorithm increases the suggestion accuracy for about 22.45%, 11.9% and 7.5% comparing with the SimRank, BRW and FRW algorithm, respectively. The parameter \( \alpha \) plays an important role in our method. It controls how fast heat will propagation on the graph. Hence, we also conduct experiments on evaluating the impact of parameter \( \alpha \). The evaluation results are shown in Figure 4. We can observe that the best \( \alpha \) setting is 1. Figure 5 shows the performance changes with different sub graph sizes. We observe that
when the size of the graph is very small, like 500, the performance of our algorithm is not very good since this subgraph must ignore some very relevant nodes. When the size of sub graph is increasing, the performance also increases. We also notice that the performance on sub graph with size 5,000 is very close to the performance with size 100,000. This indicates that the nodes that are far away from the query node are normally not relevant with the query node.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel framework for recommendations on large scale Web. 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, personalized recommendations, etc. The generated suggestions are semantically related to the inputs. The experimental analysis on several large scale Web data sources shows the promising.

REFERENCES