International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue 2, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Mining Web Graphs for Recommendations

Surendra Mahajan sa_mahajan@yahoo.com
Abolee Pande aboleepande@gmail.com
Sayalee Pande sayaleepande@gmail.com
Darshan Sanghvi darshan.sanghvi31@yahoo.com
Ronak Shah rontheking92@gmail.com

***Abstract* – Recommendation techniques have become increasingly essential. The different kinds of recommendations are made on the Web workaday, including images, music, books recommendations, query suggestions, etc. This paper, providing a common framework on mining Web graphs for recommendations using heat diffusion method, first propose a Recommendation algorithm the algorithm aggregates items from these similar customers eliminates items the user has already rated, and recommends the remaining items to the user. Which propagates similarities between different recommendations like image recommendation, the proposed algorithm can be utilized in many recommendation tasks on the World Wide Web, including image recommendations, etc. The observational Analysis on huge datasets shows the promising future of our work.**

***Keywords* - Diffusion, Collaborative Filtering, Image Recommendation, Query Suggestion, Recommendation.**

I. INTRODUCTION

The 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 modelled in the form of various types of graphs.

1) First propose a novel diffusion method which propagates similarities between different nodes and generates recommendations;

2) Then illustrate how to generalize different recommendation problems into graph diffusion framework. The proposed framework can be utilized in many recommendation tasks 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 the work.

II. RECOMMENDER SYSTEM

Recommender systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item.

Recommender systems typically produce a list of recommendations in one of two ways-through collaborative or content-based filtering. Collaborative filtering approaches build a model from a user's past behaviour (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to predict items (or ratings for items) that the user may have an interest in. Content based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often combined.

Architecture

The diagram given below shows the proposed architecture for our recommender system.

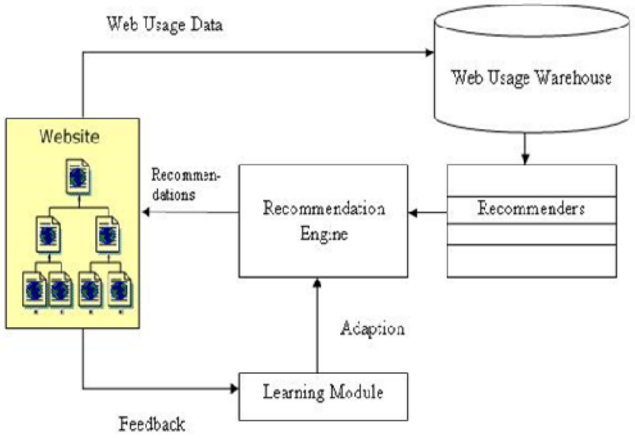

Fig.1. Architecture of Recommender System

III. COLLABORATIVE FILTERING

Collaborative filtering (CF) is a technique used by some recommender system. Collaborative filtering has two senses, a narrow one and a more general one. In general, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in electronic commerce and web applications where the focus is on user data, etc. The remainder of this discussion focuses on collaborative filtering for user data, although some of the methods and approaches may apply to the other major applications as well [5].

In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue *x* than to have the opinion on x of a person chosen randomly. For example, a collaborative filtering recommendation system for television tastes could make predictions about which television show a user

should like given a partial list of that user's tastes (likes or dislikes). Note that these predictions are specific to the user, but use information gleaned from many users. This differs from the simpler approach of giving an average (non-specific) score for each item of interest, for example based on its number of votes.

Collaborative filtering systems have many forms, but many common systems can be reduced to two steps:

1. Look for users who share the same rating patterns with the active user (the user whom the prediction is for).
2. Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user

This falls under the category of user-based collaborative filtering. A specific application of this is the user-based Nearest Neighbor Algorithm.Alternatively, item-based collaborative filtering proceeds in an item-centric manner:

1. Build an item-item matrix determining relationships between pairs of items
2. Infer the tastes of the current user by examining the matrix and matching that user's data

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender system.

IV. Data Processing

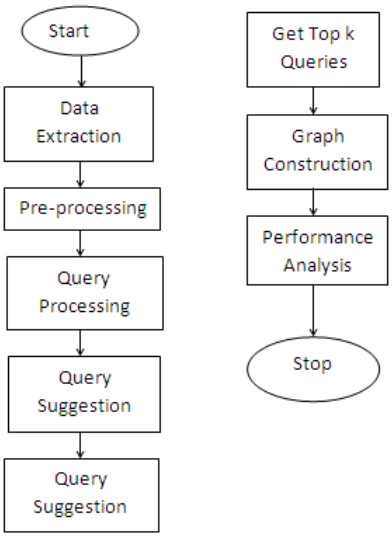

Fig.2. Data Flow

The first and foremost step in our system is data set extraction. Construct the query suggestion graph based on the click through data of the AOL search engine. Click through data record the activities of Web users, it shows their interests and the latent semantic relationships between users and queries as well as queries and clicked Web documents. Dataset is based on an existing a data source. A dataset specifies query parameters, query, filters, and a field collection. And also specify data options, such as collation, case, width and accent, for the data retrieved from the data source.

Every line of click through data contains: a user ID, a query issued by the user, a URL on which the user clicked, the rank of that URL, and the time at which the query was submitted for search. That data set is the raw data recorded by the search engine, and contains a lot of noise which will potentially affect the effectiveness of our query suggestion algorithm. So, in this module filter the data by only keeping those frequent, well formatted, English queries.

V. Query Processing

In this module want to implement query suggestion algorithm for graph construction in the next module. This query suggestion algorithm contains diffusion methodology to suggest recommendation [1]. To get the exact query for search means to calculate and get the pre-processed results for

- Query Set
- Redundant
- New Query Set

Query Diffusion

Query suggestion is associated to query substitution or query expansion, which extends the unique query with original search terms to narrow down the range of the search. But dissimilar from query expansion, query suggestion ambitions to suggest full queries that have been expressed by previous users so that query integrity and coherence are preserved in the suggested queries. Data cleaning is the very important process in the data mining process and also improves the quality of the data. It is used to detection the irrelevant data and removing it. Data quality problem is solved by using data cleaning method.

In Query processing removes the unformatted data and duplicates data. In this data set having query Id, URL, rank and time. Based on this data set removes the unformatted data in the data set. Calculating the rank values for every values for URL.

Based on that rank values removes duplicates are removed from data set. And also shows the number of rows removed in this process. Select the Query search for calculating the optimization values for that query search.

VI. Query Suggestion Algorithm

Algorithm

1: A converted bipartite graph G =(V+∪ V*,E) consists of query set V+ and URL set V.

2: Given a query q in V+*, a sub graph is constructed using depth-first search in G. The search stops when number of queries is larger than a predefined number

3: As analyzed above, set a =1, and without loss of generality, set the initial heat value of query q f(0) = 1 (the choice of initial heat value will not the suggestion results). Start the diffusion process qf(1) = eaRf(0).

4: Output the Top-K queries with the largest values in vector f(1) as the suggestions [3].

VII. IMAGE RECOMMENDATION

In this system users are first asked to rate the images and then recommend images to the users based on the tastes of the users.Image recommendation focus on recommending interesting images to users based on users' preferences. In image recommendation system users are asked first to like or dislike images. Based on likings users are recommended images.The system provides users with top suggested images and then the user chooses the best image out of the given suggestions.

VIII. Graph

We are using a bipartite graph. A bipartite graph (or bigraph) is graph whose vertices can be divided into two disjoint sets *U* and *V* (that is, *U* and *V* are each independent sets) such that every edge connects a vertex in *U* to one in *V*. Equivalently, a bipartite graph is a graph that does not contain any odd-length cycles. One set consists of query and the other set consists of URL's for query processing.The graph is so constructed such that every edge connects a vertex query to one in URL.The edge of the graph contains weights. These weights are nothing but the rank or the heat diffusion value.

IX. CONCLUSION

We have presented a novel framework for recommendations on large scale Web graphs using heat diffusion. 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.
Several large scale Web data sources shows the promising future of this approach.

REFERENCES

[1] Hang Cui, Ji-Rong Wen, Jian-Yun Nie and Wei-Ying Ma, "Query Expansion by Mining User Logs", IEEE Transactions on Knowledge and Data Engineering, Vol.

[2] Yong Zhen Guo, K. Ramamohanarao and L. A. F. Park, "Personalized PageRank for Web Page Prediction Based on Access Time-Length and Frequency", IEEE/WIC/ACM International Conference on Web Intelligence, pp. 687– 690, 2007.

[3] Daniel Fogaras and BalazsRacz, "Practical Algorithms and Lower Bounds for Similarity Search in Massive Graphs" IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 5, pp. 585 – 598, 2007.

[4] Ming-Sheng Shang, Yan Fu and Duan-Bin Chen, "Personal Recommendation using Weighted Bipartite Graph Projection", International Conference on Apperceiving Computing and Intelligence Analysis, pp. 198 – 202, 2008.

[5] Jinbo Zhang, Zhiqing Lin, Bo Xiao and Chuang Zhang, "An Optimized item-based collaborative filtering Recommendation Algorithm", Proceedings of the IEEE International conference on Network Infrastructure and Digital Content, pp. 414 – 418, 2009.

[6] K. Kazama, M. Imada and K. Kashiwagi, "Characteristics Estimation of Information Sources by Information Diffusion Analysis", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Vol. 1, pp. 484 – 491, 2010.

[7] Hao Ma, Irwin King and Michael Rung-TsongLyu, "Mining Web Graphs for Recommendations", IEEE Transactions on Knowledge and Data Engineering, Vol. [8] L. Si and R. Jin. Flexible mixture model for collaborative filtering. In Proc. of ICML, 2003.

[9] C. Silverstein, M. R. Henzinger, H. Marais, and M. Moricz. Analysis of a very large web search engine query log. SIGIR Forum, 33(1):6–12, 1999.

[10] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In Proc. of ICML, pages 720–727, 2003.

[11] J.-T. Sun, D. Shen, H.-J. Zeng, Q. Yang, Y. Lu, and Z. Chen. Webpage summarization using clickthroughdata. In Proc. of SIGIR, pages194–201, Salvador, Brazil, 2005.

[12] M. Theobald, R. Schenkel, and G. Weikum. Efficient and self-tuning incremental query expansion for top-k query processing. In Proc. OfSIGIR, pages 242–249, Salvador, Brazil, 2005.

[13] B. Velez, R. Weiss, M. A. Sheldon, and D. K. Gifford. Fast and effective query refinement. SIGIR Forum, 31(SI):6–15.

[14] L. von Ahn and L. Dabbish. Labeling images with a computer game. In Proc. of CHI, pages 319–326, Vienna, Austria, 2004.