International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 3, Number 6 (2013), pp. 511-520

http://www. irphouse.com /ijict.htm

Search Results Improvement through Ranking Approach

T. Naheena[1] and K.F. Bharathi[2]

[1]Department of Computer Science and Engineering, JNTUA College of Engineering, Anantapur (Autonomous), India.
[2]Department of Computer Science and Engineering, JNTUA College of Engineering, Anantapur (Autonomous), India.

Abstract

Recommender systems provide plenty of benefits to both users and the businesses. Due to Web related applications, user-generated information is more freestyle and less ordered, which increases the difficulties in mining helpful information from these data sources. Innumerable dissimilar types of recommendations are ready on the Web daily, with movies, music, images, books recommendations, query suggestions, and tags recommendations, etc. It is not concerned about types of data sources old for the recommendations, basically these data sources can be constructed in the type of dissimilar types of graphs. Then it illustrates to generalize different recommendation problems into graph diffusion framework. The existing system does not focus on improving search results. In order to suit the information needs of Web users and recover the user practice in many web applications, Recommender Systems. The proposed framework can be using in many recommendation errands on the World Wide Web (WWW). The search results are improved by heat diffusion based ranking. By using this process we can satisfy the user needs in web application.

Keyword: Graphs construction, search results, query suggestion, heat diffusion, and ranking.

1. Introduction

Web mining is the data mining method that repeatedly discovers or retrieve the information from web documents. It is the incorporation of information gathered by conventional data mining processes and techniques with information retrieved larger than the WWW [1]. Web-mining is a multi-disciplinary attempt that draws techniques from fields like in-formation retrieval, information, machine learning, common language processing, and others. Data mining aims at discovering precious information that is unseen in conventional data bases, the emerging field of web mining focuses at result and extracting relevant information that is unseen in web-linked data.Web mining particularly aimed in hyper-text papers available on the web [2]. Web mining has new characters compared to traditional data mining. The major concerns of them are first, the items of web mining are a big number of web documents which are heterogeneously dispersed and every data source is heterogeneous. Second new character of web mining is that the web document itself is semi-structure or un-structured and lack the semantics the appliance can know, this area of research has been alert due to the great growth of information sources obtainable on the web and the new attention in e-commerce [3].

Web mining categories :Generally web mining is categorized into three based on the web data used as input[4].They are **web content mining**, **web structure mining** and **web usage mining**. Web content mining is the procedure of retrieving the information from WWW into additional structured forms and indexing the information to recover it quickly. It attention mainly on the structure inside a document which is usually called internal document level. Web content mining is linked to data mining, because it is possible to apply several data mining techniques in web content mining. Web structure mining is the procedure of determine the process of connection arrangement of the web pages. We list the links by generate the information such as the likeness and relations among them by taking the benefit of hyperlink topology. Page rank and hyperlink analysis also fall in web structure mining category. The primary objective of web structure mining is to make ordered review about the website and web page. Web usage mining is the process by which it is possible to recognize the browsing patterns by analyzing the navigational performance of user. The major focus of web usage mining is on techniques that can be used to forecast the user activities while the user communicate with the web. It uses the secondary data on the web. This action involves the mechanical discovery of user entrée patterns from one or more web servers. Follow this mining technique we can define the concern of users on Internet. The search engine has no infrastructure or matching techniques to give correct or related information for the query raised. The semantic web has the ability to give the solution for this problem. Each page contains meta data with notes, meanings, list of words, definitions, etc. The popularity of search engines on the internet has led to a nearly uniform interface for searching a single input box that accepts keywords. This simple interface has pushed the burden of inferring the intent of the user's information need down into the search engine. With the enlarge of bulk and popularity of the World Wide Web, numerous users discover it's complicated to get the preferred information, although they utilize most efficient search engines (e.g. Google, yahoo).Actually theses search engines permit users to identify queries simply as lists of keywords, following the approach of conservative information systems [5].

The explosive increase of web information has not only created a critical challenge for search engine to handle great size data, but also improved the complexity for a user to manage his information need. It is complex for a user to create exact query to near his search need. Instead of pushing this load to the users, it is ordinary practice for a search engine to give some types of query suggestions.

2. Related Work

Search result is a specific type of technique which contains the user information need in the web. In this section, we review several works related to search result including graph construction, heat diffusion, query suggestion, image recommendation methods.

2.1. Graphs construction

The exponential explosion of different contents produced on the Web, Recommendation techniques have become more and more crucial. Innumerable various types of recommendations are made on the Web each day, as well as movies, music, images, books recommendations, query suggestions, tags recommendations, etc. In this no issue what types of data sources are utilized for the recommendations, basically these data sources can be designed in the form of different types of graphs [6]. In graph building consider an undirected graph G={V,E},where V is the vertex set, and $V=\{V_1,V_2,\ldots\ldots,V_n\}$, E is the set of all edges. In this node contain query and edge contain user resource location (URL). The value on the edges is specify how many times a query is clicked on a URL. This module is responsible to take click-through data of American online (AOL) search engine [7] or Flickr data and extract bipartite graph from data which is undirected in nature. The bipartite graph is converted into another form of bipartite graph where each undirected edge becomes two directed edges. That web graphs are normally very huge, and so that generally algorithm will be performed on a sub graph extracted from the original graph. Hence, it is necessary to evaluate the size of the sub graph which affects the recommendation accuracy. The performance changes with different sub graph sizes. It is observed that when the size of the graph is very small, like 500, the performance of our algorithm is not accurate since this sub graph must ignore some very relevant nodes. However, when the size of sub graph is increasing, the performance also increases. It is also noticed that the performance on sub graph with size of 5,000 is very close to the performance with size of 1, 00,000. This indicates that the nodes that are far away from the query node are normally not relevant with the query node. The sub graph is designed by using depth-first search in the original graph. The search stops when the number of nodes is larger than a predefined number.

2.2. Heat diffusion process

Heat diffusion is a physical fact. In a medium, heat forever flows from a high temperature area to low temperature area. Recently, heat diffusion-based approaches have been successfully applied in different domains such as classification and dimensionality reduction problems [9, 10, 11]. In our process, we use heat diffusion to model the similarity information propagation on web graphs. In Physics, the heat diffusion is always performed on a geometric manifold with initial conditions. However, it is very difficult to represent web as a regular geometry with known dimensions. This motivates us to investigate the heat flow on a graph. The graph is considered as an approximation to the underlying manifold, and thus the heat flow on the graph is considered as an approximation to the heat flow on the manifold.

2.2.1 Diffusion on undirected graphs

Consider an undirected graph G = (V, E), Where V is the vertex set, and V = (V_1, $V_2 \ldots V_n$). E is the set of all edges. The edge (v_i; v_j) is considered as a pipe that connects nodes v_i and v_j. But in many cases, the Web graphs are directed, specifically in online recommender systems and knowledge distribution sites. Each user in knowledge distribution sites typically has a belief list. The users in the belief list can influence this user deeply. These relationships are directed since user “a” is in the belief list of user “b”, but user “b” might not be in the belief list of user “a”. At the same time, the extent of trust relations is various since user u_i may belief user u_j with belief score “1” while belief user u_k only with belief score 0.2. Hence, there are various weights related with the relations. Δt is the time period, H is heat matrix, α is heat diffusion coefficient, D is diagonal matrix, calculation of heat diffusion based bellow equation.

$$(f(t+\Delta t)-f(t))/(\Delta t)=\alpha(H-D)f(t) \qquad (1)$$

2.2 2 Diffusion on direct graphs

Based on consideration of diffusion in undirected graphs, we adjust the heat diffusion model for the directed graphs as follows. Consider a directed graph, G = {V; E; W}, where V is the vertex set and W = {w_{ij} / where w_{ij} is the probability that edge (v_i, v_j) exists} or the weight that is associated with this edge. E is the set of all edges. Form the consideration of diffusion for direct graphs; it is strongly felt that direct graph is better for construction of web graphs comparative to indirect graphs.

2.3. Query suggestion

In sort to recommend related queries to web users, a precious technique, query suggestion, has been working by some famous saleable search engines, such as Yahoo, Live Search, Ask, and Google. The purpose of query suggestion is parallel to that of query expansion [12, 13, 14, 15], query substitution [16], and query refinement [17, 18], which all center on considerate users’ search intentions and improving the queries submitted by users. Query suggestion is strongly connected to query expansion or query substitution, which extends the original query with new search conditions to narrow down the reach of the search. Both query expansion and query refinements are query recommendation methods based on proposed click-through data. The main disadvantage of these two query refinement and query recommendation methods is that they disregard the loaded information rooted in the query-click bipartite graph, and consider only queries that emerge in the query logs, potentially losing the chance to recommend extremely semantically related queries to users.

Recently Mei et al. proposed a general query suggestion method using hitting time on the query-click bipartite graph [19]. This method can generate semantically relevant queries to users' information needs. The primary advantage of this work is that it can suggest some long tail queries (infrequent queries) to users. However, this approach has disadvantages because sometimes it may accidentally rank the infrequent queries highly in the results while potentially downgrades the ranks of the most related queries.

2.4. Image recommendation

Besides query suggestion, one more motivating recommendation request on the web is image recommendation. Image recommendation systems, similar to Photo tree, focus on recommending motivating images to web users based on user's liking. Recently, by employing the Flickr data set, Yang et al. proposed a context-based image search and recommendation technique to recover the image search value and recommend related images and tags. However, since it is a context-based technique, the computational difficulty is extremely high and it cannot size to large data sets. By diffusing on the image-tag bipartite graph with one or more images, we can exactly and efficiently suggest semantically related non personalized or personalized images to the users. These image recommendations are not make happy to user needs. It is not precisely give the relevantly images to the users.

3. Proposed System

The major aim of the proposed system is at improving the recommendation framework by incorporating a provision that enhances the usefulness of search results. It utilized in many recommendation tasks on the WWW. The search results can be found by heat diffusion based ranking. It gives good result in search engine. The process is explained by step by step in the bellow Figure.1.

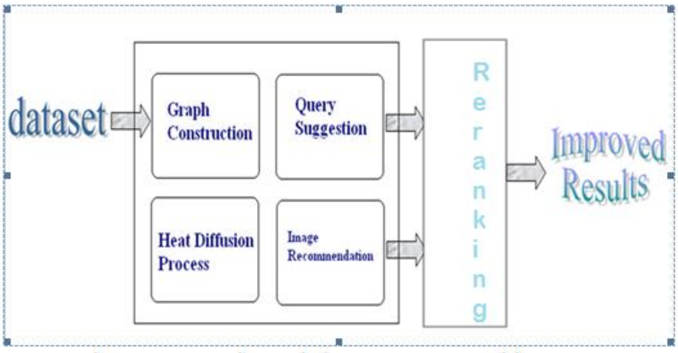

Figure 1: Search result improvement architecture.

3.1. Graphs construction

In this process contains direct graphs for web mining**.** The direct graphs give good performance in search engine. By using this type of graph, we can easily find out similar information in web mining.

3.2 Query suggestion

In the above figure query suggestion is one type of user recommendation in search engine is made on the web every day. Query suggestion involves utilized ranking algorithm for recommendation after heat diffusion process. This concept has been adopted in the process of query suggestion and image recommendation. Based on the heat values and their corresponding propagation, their relative values with respect to usefulness are determined. These values are used to estimate similarities among the nodes in the graph. The similarity nodes are filtering by using ranking. Query suggestion is done by following processes.

i. Queries beside with the text of their clicked URL's retrieved from the web log are clustered.
ii. Given an *input query* (i.e., a query submitted to the search engine) we first discover the cluster to which the input query belongs. Then we calculate a rank for each query by heat diffusion value in the cluster.
iii. Finally, the associated recommendations are returned ordered according to their rank.

3.3. Click-through

The mining related queries from a click-through bipartite build from search logs. Here, the basic statement is that two queries are similar to each other if they distribute a large number of clicked URLs .The set of URL's are clicked by users for the query. It utilized any of the following steps.

i. rooted in keywords or phrases of the query,
ii. rooted in string similar of keywords,
iii. rooted in common clicked URLs,
iv. rooted in the space of the clicked documents in some pre-defined hierarchy.

3.4. Image recommendation

In the above figure image recommendation is one type of user recommendation in search engine is made on the web every day. Image recommendation is similar to query suggestion and in this process images are recommended by using heat diffusion based ranking. This process gives the related images in efficiently and accurately with in short time. In this process web users are satisfy.

3.5. Heat diffusion based ranking

The search results are obtained by using heat diffusion based ranking in the recommendation framework. Heat diffusion based ranking contains ranking for each recommendation in the searching process. The ranking is given by its heat values (click-through data) of recommendation. The recommendations with higher value are seeing in top of the recommendation through ranking in search engine. After this process it provides user related recommendations in the top of the search engine.

3.6. Re-ranking

Re-ranking is the process of assigning the rank for the obtained search results which have been ranked already. The re-ranking process will optimize the result of a search engine by returning the more relevant and user desired pages on the top of search result list.

This re-ranking is applied to this project to obtain the more accurate and user-intended search results. We apply this re-ranking based on the existing ranking anc computed weights. The resultant ranking will be more opt to organize search results that reflect user intentions as much as possible

3.7. Search Result improvement

In our implementation we focused on query suggestions and image recommendations. Previously, query suggestions are generated based on the algorithm named "query suggestion algorithm" by taking AOL click-through data as input. Same algorithm is used for image recommendation which takes images of Flick as input. In either case, the results are improved by incorporating heat diffusion based ranking model. The search results are recommendations which have been optimized based on the heat value associated with resultant records. The heat values actually the wisdom of the crowd since it is based on the query-URL click data, which reflects the intelligent judgments of the web users. The ranking model computes heat values in order to present best ranked results at the top [8].After heat diffusion process applied Re-rank process it will give the good results. Search results are substantially improved by applying diffusion based ranking model. The heat values are used to determine ranking of search results. Thus the optimized results make more sense to end users.

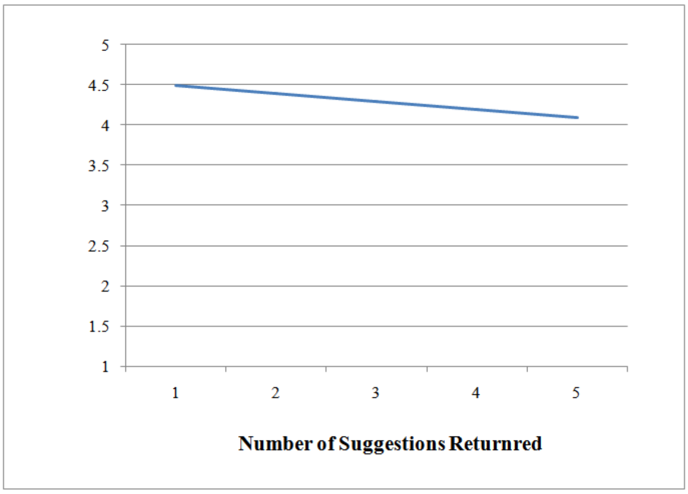

Figure 2: Accuracy of Search result improvement.

The average values of evaluation results are shown in Fig.2 .The accuracy is increases in heat diffusion based ranking process and return suggestions are related to query.

The table1 shows results of query suggestion through heat diffusion based ranking process. The below table explain ranking process for recommendation. Image recommendation process is done same like query suggestion process and results are also same like query suggestion process.

Table 1: Query suggestion result.

Query	Rank	Web site	Heat value	Re-rank
Sony	1	www.sony.com	14	15
	2	www.sonystyle.com	12	14
	3	www.sony.net	9	12
	4	www.sonypictures.com	8.32	12.32
	5	www.sonyericsson.com	4	9
Microsoft	1	www.microsoft.com	11	12
	2	Windowsupdate.microsoft.com	9.01	11.01
	3	Support.microsoft.com	7	10
	4	Office.microsoft.com	4	8
	5	www.msn.com	2.14	7.14
camera	1	www.dpreview.com	22	23
	2	www.adorama.com	18	20
	3	www.ritzcamera.com	16	19
	4	www.dcresource.com	13	13
	5	www.digitalcamera-hq.com	10.12	15.12
chocolate	1	www.chocolate.com	15	16
	2	www.hershys.com	12	14
	3	www.godiva.com	8.17	11.17
	4	www.ghirardelli.com	7.01	11.01
	5	www.scharffenberger.com	5	10

4. Conclusion and Future Work

The proposed framework give related results to inputs. The framework is useful for recommendation on large data sets, and it satisfy the user needs. Heat diffusion based ranking give related results for input uses. The recommendations are highly related to inputs .The query suggestion and image recommendation are web graph recommendations, it is done by heat diffusion based ranking. The future is implement to social recommendation on different web sites and knowledge sharing process.

Reference

[1] Web mining definition . available :http://en.wikipedia..org/wiki/Web_mining.

[2] About Web.avaliable:http://www.technicalsymposium.com/web_mining_notes.html.

[3] R. Kosala, H. Blockeel "Web mining research: A survey," ACM SIGKDD Explorations, Vol. 2 No. 1, pp. 1-15, June 2000.

[4] R. Kosala, and H. Blockeel, Web Mining Research: A Survey, SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining Vol. 2, No. 1 pp 1-15, 2000.

[5] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval chapter 3, pages 75–79. Addison-Wesley, 1999.

[6] Haoma, Irwin king, "Mining Web Graphs for Recommendations" IEEE Transactions on Knowledge and Data Engineering, vol.24, no.6,JUNE 2012.

[7] G. Pass, A. Chowdhury, and C. Torgeson, "A Picture of Search," Proc. First Int'l Conf. Scalable Information Systems, June 2006.

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

[9] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

[10] R.I. Kondor and J.D. Lafferty, "Diffusion Kernels on Graphs and Other Discrete Input Spaces," ICML '02: Proc. 19th Int'l Conf. Machine Learning, pp. 315-322, 2002.

[11] J.D. Lafferty and G. Lebanon, "Diffusion Kernels on Statistical Manifolds," J. Machine Learning Research, vol. 6, pp. 129-163, 2005.

[12] 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.

[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] M. Theobald, R. Schenkel, and G. Weikum, "Efficient and Self- Tuning Incremental Query Expansion for Top-k Query Processing," SIGIR '05: Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 242-249, 2005.

[15] J. Xu and W.B. Croft, "Query Expansion using Local and Global Document Analysis," SIGIR '07: Proc. 19th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 4-11, 1996.

[16] R. Jones, B. Rey, O. Madani, and W. Greiner, "Generating Query Substitutions," WWW '06: Proc. 15th Int'lConf. World Wide Web, pp. 387-396, 2006.

[17] R. Kraft and J. Zien, “Mining Anchor Text for Query Refinement,” WWW ’04: Proc 13th Int’l Conf. World.
Wide Web, pp. 666-674, 2004.

[18] B. Ve´lez, R. Weiss, M.A. Sheldon, and D.K. Gifford, “Fast and Effective Query Refinement,” ACM SIGIR Forum, vol. 31(SI) pp. 6-15, 1997.

[19] Q. Mei, D. Zhou, and K. Church, “Query Suggestion Using Hitting Time,” CIKM ’08: Proc. 17th ACM Conf. Information and Knowledge Management, pp. 469-477, 2008.