IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 3, March 2016.
www.ijiset.com
ISSN 2348 – 7968

Enhancement in web service provision using combination of collaborative technique and association rule generation

P.Kavitha[1]
Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India
E-Mail: kaviamrita@gmail.com

Dr.G.N.K.Suresh babu[2]
Associate Professor, GKM College of Engineering and Technology, Chennai, Tamil Nadu, India.
gnksureshbabu@gmail.com

Abstract:

Search engine provide query recommendation which suggest user to choose any one query from their recommendation list according to their necessities. Such systems had been expected to have a good result, especially in E-commerce and E-learning environments. However, although they have been intensively explored in the Web Mining and Machine learning fields, and there have been some commercialized systems, the quality of the recommendation and the user satisfaction of such systems are still not conclusive, since the search result provided by the search engine will not be a satisfactory one in most of the cases. In this paper we proposed a robust model that leverages most appropriate and promising search result for user queries. The system uses useful information on user behavior and their activities when searching online and the content they are interested in viewing for the result which they obtained in response to their query by search engine. The proposed technique able to provide required promising result as a response rather than returning the result which has maximum hits for their search queries.

Keyword: E-commerce, E-Learning, Recommendation system, Search engine

Introduction:

Web search is an area of the information technology industry where artificial intelligence and particularly knowledge engineering techniques can be applied with potentially significant impacts. Currently users face a still increasing amount of data for their queries, since the search engine provide the result which has maximum hit rate. Such result cannot able to satisfy the user as their interested content may not present in those results.

It is quite reasonable assumption that a series of queries characterize the effort of what the user want to find better than the single query. Therefore the search engines were designed in such a way to guide the user for selecting queries which results in the appropriate thing they need.

Recently, some of the major search engines have introduced a new “Search History” feature, which allows users to track their online search by recording their queries and clicks. This history includes a sequence of the queries displayed in reverse chronological order together with their corresponding clicks.

The proposed system which organize user search histories and learning of user behavioral patterns not for providing relevant query pattern but for getting the result in which the user are already interested to know, by the way it could identify the hidden data which is

ISSN 2348 – 7968

more appropriate than the actual result provide by the search engine.

Related works

Search engines, such as Yahoo! [8] and AltaVista [9], are useful for finding information on the World Wide Web. However, these general-purpose search engines are subject to low precision and/or low coverage. The main stumbling block for a user who wants to utilize topic-specific search engines is about how they find the appropriate result as response for any given query.

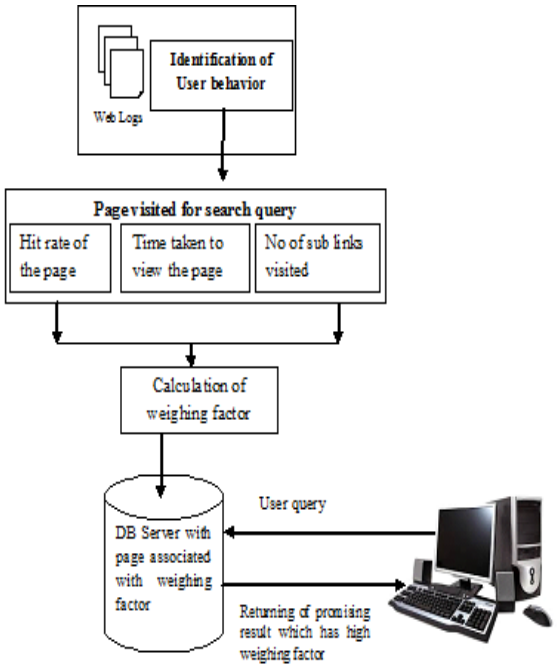
Usually the systems in search engine can be built in three ways: content based filtering, collaborative filtering, and hybrid systems. Content-based recommender systems, sometimes called information filtering systems, use behavioral user data for a single user in order to try to infer the types of item attributes that the user is interested in. Collaborative filtering compares one user's behavior against a database of other user's behaviors in order to identify items that like-minded users are interested in. Even though content-based recommender systems are efficient in filtering out unwanted information and generating recommendations for a user from massive information, they find few if any coincidental discoveries.

On the other hand, collaborative filtering systems enable serendipitous discoveries by using historical user data. Collaborative filtering algorithms range from simple nearest neighbor methods [3, 6] to more complex machine learning based methods such as graph based methods [1, 5], linear algebra based methods [2,4] and probabilistic methods. These studies are about the recommending system of search engine. Once user selects the query, it will go for respective servers for fetching result. Search engines match queries against an index that they create. The index consists of the words in each document, plus pointers to their locations within the documents. This is called an inverted file. A search engine system comprises four essential modules as document processor, query processor, search and matching function and ranking capability [7]. Studies has been done for routing and generating perfect query pattern for obtaining result from the server, but the proposed system deals in returning best result document to the user as response than sending page which has simply high hit rate.

The page which is more appropriate for the user query is calculated using Weighted Associative Class rule generation. Syed et al., [10] have proposed weighted associative class (WCARs), which integrates weighted association rule mining and classification to construct the efficient weighted associative classifier. Weighted associative classifier extracts special subset of association rules called weighted class association rules (WCARs). Weighted association rule mining uses weight as one parameter but here weights for each item are assigned randomly but it is very difficult to assign weights to each item.

Proposed work:

The proposed system involves in identifying best solution for user query at database than simply sending some results using behavioral pattern of the users who do online searching.



ISSN 2348 – 7968

S. No	Links
1	http://all-free-download.com/free-photos/red-rose.html
2	http://hdw.eweb4.com/search/red+rose/
3	http://photobucket.com/images/red%20rose
4	http://www.justourpictures.com/roses/redroses.html
5	http://en.wikipedia.org/wiki/Red_rose
6	http://www.fanpop.com/clubs/roses/images/11662034/title/red-red-rose-wallpaper
7	http://www.poetryfoundation.org/poem/173068
8	http://www.redrosetea.com/
9	http://www.robertburns.org/works/444.shtml
10	http://sonsofanarchy.wikia.com/wiki/Red_Rose

Fig.1. Architecture diagram of proposed system

The behavioral pattern involves in identifying not only the number of times that particular result or page has been viewed, but also the time factor about how long a user spends time over that page, number of sub links he/she interested in viewing that particular page etc.

Fig.1 shows the architecture diagram of the proposed system. The novel approach of proposed system is to store all the details about a page hit rate, time factor about how long it has been viewed, and number of sub link visited in that page. Using the above mention factor as parameter a weight factor will be generated using WCARs. Thus each page will be associated with its weighing factor for a specific query. The page which has higher weighing factor will be considered as the best result and that will be returned to the user when he/she give relevant query.

Weighted class association rule generation

Weighted Associative classifier construction is of two steps. In the first step, all the weighted class association rules are generated based on weighted association rule mining technique. To find the weighted association rule, weight is the important factor. Here weighted support and weighted confidence measures are used to evaluate the rule.

The w-support of a rule Q->R is defined as

$$\text{Sup(Q->R)}= \sum \frac{Weight(Q->R)}{Totalweight}$$

The w-confidence of a rule X C is defined as

$$\text{Conf(Q->R)}= \sum \frac{Conf\,(R->Q)}{Conf\,(R)}$$

Q refers to the query and R refers to result for that query, Sup(Q->R) indicate the support for returning the result for that query and Conf(Q->R) refers to the confident factor.

Table 1. Links used to find weighing factor

Experimental result:

The experimental is done by testing the system with different links with different parameter values, (e.g) Table 1. Shows the link visited for the query “red rose” in the proposed system, likewise the system has been tested with different search query.

S.NO	Hit count	Time factor	Sub links viewed	Support factor	Confident factor
1	200	3 min	5	.9	.9
2	153	.23 min	1	.5	.49
3	148	1 min	2	.39	.32
4	320	.5 min	1	.1	.01
5	298	2.2 min	2	.55	.5
6	173	1.9 min	1	.35	.3
7	247	.8 min	0	.1	.01
8	95	.3 min	0	.1	.01
9	70	1.3 min	0	.1	.01
10	103	.5 min	0	.1	.01

Table 2. Parameters for user behavior for table 1

Table 2 shows the relevant parameter to find the weighing factor of a web page that has to be send when user sends a query. In the proposed

ISSN 2348 – 7968

system the weighing factor is identified by finding the average of support factor and the confident factor. The page which has the maximum weighing factor will be return as a response for the user query.

Performance analysis:

The system efficiency has been calculated by testing each and every page parameter. The proposed system involves in extracting identifying the hit rate of the page, time taken by user to view the page, number of links clicked in that particular page. Then the weighing factor is calculated using WCARs. The proposed system shows accuracy of about 93.7% in responding most promising result for the user query.

Conclusion and future work:

In this paper we proposed a novel approach for responding most relevant result for the user by combining the parameters like page hit rate, time viewed, number of active links present in that page. The system is more efficient in identifying the appropriate result than sending too many associated page for the user. The system can be more efficient if it combines with some other heuristic approaches like feedback from user.

References:

[1] G J. Teevan, E. Adar, R. Jones, and M. A. S. Potts, “Information retrieval: repeat queries in yahoo’s logs,” in SIGIR. New York, NY, USA: ACM, 2007.

[2] R. Jones and K. L. Klinkner, “Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs,” in CIKM, 2008.

[3] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, “The query-flow graph: Model and applications,” in CIKM, 2008.

[4] R. Baeza-Yates and A. Tiberi, “Extracting semantic relations from query logs,” in KDD, 2007.

[5] K. Avrachenkov, N. Litvak, D. Nemirovsky, and N. Osipova, “Monte carlo methods in PageRank computation: When one iteration is sufficient,” SIAM Jthenal on Numerical Analysis, vol. 45, no. 2, pp. 890–904, 2007.

[6] P. Boldi, M. Santini, and S. Vigna, “Pagerank as a function of the damping factor,” in WWW, 2005.

[7] https://www.google.co.in/insidesearch/howsearch works/thestory/.

[8] Yahoo!, http://www.yahoo.com/

[9] AltaVista, http://www.altavista.com/

[10] Syed Ibrahim S. P, Chandran K.R, Abinaya M.S (2011) , “Integration of weighted association rule mining and Classification” in the proceedings of National Conference on Recent Advances in Computer Vision & Information Technology , March 7, 2011, pp 231-235.