User profile Personalization Based on Link-Click Concept

S.Geetharani, S.Priyadharshini, M.Sorana Mageswari

Assistant Professor, Dept. of Computer Science, PSG College of Arts & Science, Coimbatore, India

Research Scholar, Dept. of Computer Science, PSG College of Arts & Science, Coimbatore, India

Assistant Professor, Dept of Computer Science, Government Arts College, Coimbatore, India

ABSTRACT: In the past few years, the growing number of personal information shared on the Web. Generally web search engines are built to serve all users, independent of the individual needs of any user. Personalization of web search is to carry out retrieval for each user incorporating their interest. A better user profiling strategy is an important and primary component in search engine personalization. This has become an important factor in daily usage as it improves the retrieval effectiveness on topics that the user would look for. In order to enhance the relevance of search results, personalized search engines form user profiles which capture the user's personal preferences and by using those preferences find out the actual goal of the input query. There are several studies that have been done in this field. This paper proposes a new idea on personalized web search which is based on user profile and link-click concept.

KEYWORDS: Personalization, user profile, search engine, click through data, input query

I. INTRODUCTION

As the amount of information on the Web increases rapidly, it creates newer challenges for Web search. When similar queries are submitted by multiple users, a general search engine could return the same result, In the current scenario, where users expect systems to be intelligent and support their information search, the general search engine is no longer workable. Systems have to be made sensitive to user likes by understanding their search patterns over a period of time. This will enable users to obtain more relevant results to searches on similar queries posted in the internet search engines. Learn the user profiles from users search histories. The user profiles are then used to improve retrieval effectiveness in web search [2].The user profile includes technical, demographical and geographical information of the users.

Personalization of web search has been area of research in the recent times several efficient approaches have been derived based on related studies and observations. Of the several methods that have been proposed and implemented, this paper focuses on two methods User Clicks and User Profile. This paper proposes an novel idea of combining two crucial methods. User profile and user clicks to enhance the web based search and increase the level of personalization. This paper proposes a new method for evaluating the quality of retrieval functions. Unlike traditional methods that require relevance judgments by experts or explicit user feedback, it is based entirely on click through data. This is a key advantage, since click through data can be collected at very low cost and without overhead for the user.

User feedback can provide powerful information for analyzing and optimizing the performance of information retrieval systems. Unfortunately, experience shows that users are only rarely willing to give explicit feedback. To overcome this problem, this paper explores an approach to extracting information from implicit feedback. The user is not required to answer questions, but the system observes the user's behavior and infers implicit preference information automatically. The particular retrieval setting studied in this paper is web search engines. In this setting, it seems out of question to ask users for relevance judgments about the documents returned. However, it is easy to observe the links the user clicked on. With search engines that receive millions of queries per day, the available quantity of such click through data is virtually unlimited. This paper shows how it is possible to tap this information source to compare different search engines according to their effectiveness.

II. RELATED WORK

So far, personalization techniques have been developed in diversified ways. The techniques can be classified into three Categories: content based personalization and function based personalization and link based personalization.

A. Content Based Method

Content based personalized mainly uses a set of <word, weight> to represent the users interest [2].The most Representations are: Match user query to a category which would be a users context to provide users with personalized Web result [2]. Gather query words or page summary and then put these data to the matched ODP category in order to develop the user's personal model [3].

B. Function based Method

Function-based personalization is mainly by designing new sorting algorithms or transforming the existing algorithms to achieve the personalized services. The most representations are: Using SVM support vector machine algorithm to sort the results of web pages; [4] An algorithm called Cube SVD is used to calculate the relationship among the three dimensional data <user, query word, page>, then obtains the weight of each page and re-sorts them. In this article, we try to combine content based method and function based method. The user profile is the content based method and the users click through data is the function based method.

III. PROPOSED ALGORITHM

In this article, a new technique is proposed by combining the User Profile and Click through data. Each of this method is designed as a separate module and the output of these two modules are mapped to find the more personalized web pages**.**

A. Architecture of the Proposed System

This Proposed system consists of two main modules. One is User Profile and the second is Click through data. When user enters a query through the interface, The key terms extraction will done and the terms will be grouped with weight. This term and weight will be compared with different categories which is stored in the database, to find the user intention. As a result of this module, it is possible to find the user interest. On the other hand in the click through data module, the Bayesian classification algorithm is used to generate patterns and the result of these two modules will be mapped to find the data, which the user needs. From the extracted key terms and weight, it is possible to find the user.s intension and need. From the click through data, it is possible to find the frequently needed data and the results from these two modules are compared to arrive at the more personalized web page. The architecture of the proposed system is as shown in figure 1.

1) Content Extraction

Due to the huge number of web pages and the high growth rate of the Web, it is impractical to analyze each web search result pages directly and separately. Therefore, the web snippets of the query are used for extracting concepts instead of web pages. Snippets are useful information resources provided by Web search engines. It contains a brief window of text selected by a Web search engine around the query in a web page[11]. We employ the following support formula, to find the frequent item sets in data mining [12] to measure the frequency of a particular keyword c_i extracted from the Web snippets arising from the query q.

Support(c_i)= sf(c_i)/n . $\left| c_i \right|$ where sf(c_i) is the snippet frequency of the keyword. n is the number of snippets returned. $\left| c_i \right|$ is the number of terms in the phrase.

2) User Profile Module

This strategy includes two steps. The first step is to map a user query to a set of categories, which represents the users search intention and serve as a context for the query. The second step is to utilize both the query and its context to retrieve web page. For example for the query .Orange., some users may be interested in documents dealing with "Orange "as "fruit", while some users may want documents related to "Orange flavor "In our approach, a category in a user profile is a weighted term vector, in which a high weight of a term indicates that the term is of high significance in that category for the user and low weight of the same in another category indicates that the term is not important in that

category. We utilize the Weights of terms in different categories to identify the interest of the user. As an example if the user is interested in both "Fruit" and "Flavor" and has previously used "Orange" in retrieving relevant documents in "Fruit" but has not used the Same word in retrieving relevant documents in the category "Flavor "As a consequence, the user profile should have a weight for the word .Orange
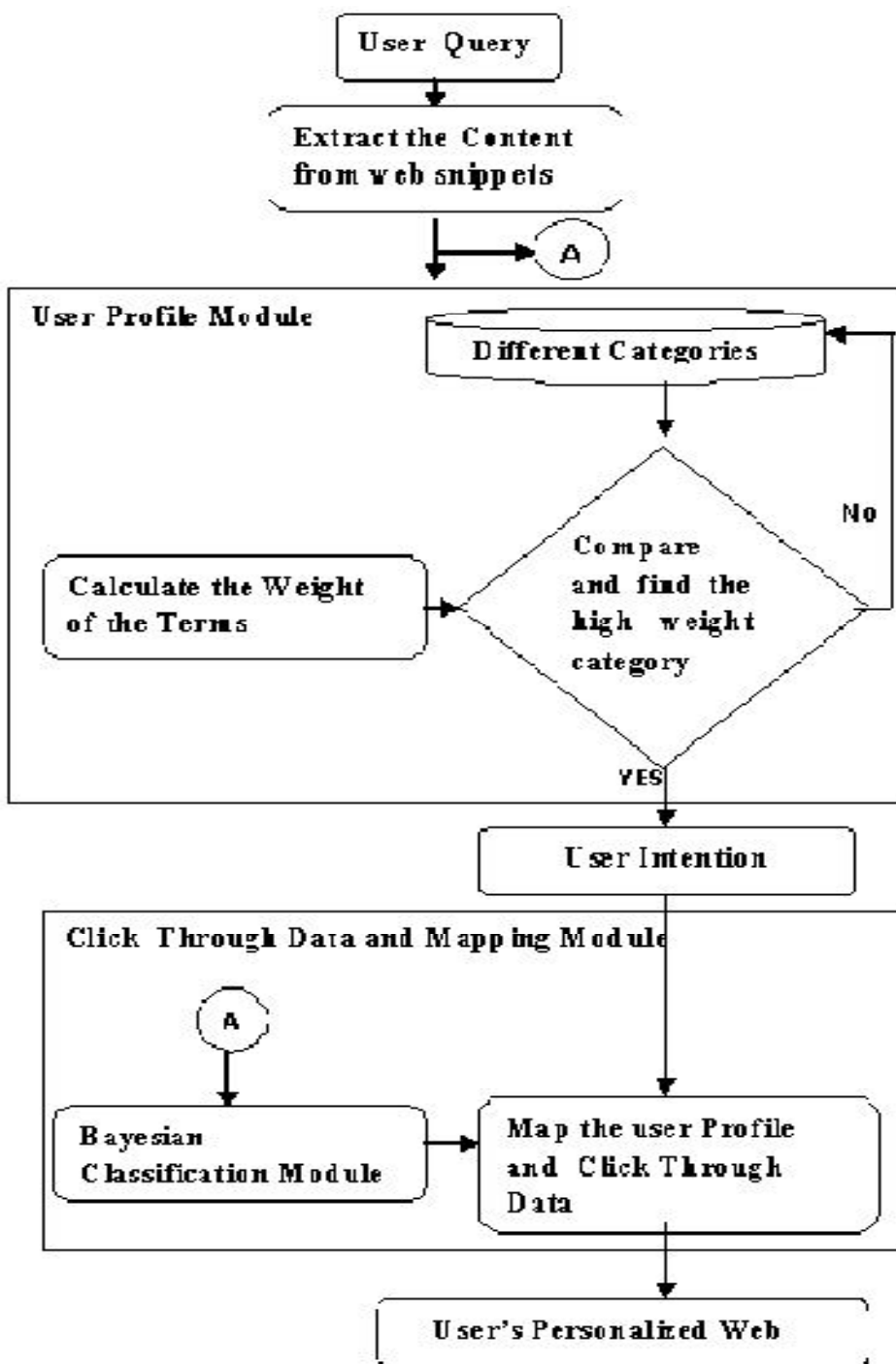

Fig. 1. Architecture of the Proposed System

In the category .Fruit., but the word has low weight in the category "Flavor "Using this user profile, when the user submits a query containing the word .Orange. ,the category .Fruit. Will be predicted for this user. User profile consists of a set of categories and for each category, a set of terms with weights. Each category represents a user interest in that category. The weight of a term in a category reflects the significances of the term in representing the users' interest in that category [2]. For example the term .Orange. has a high weight in the category .Fruit., then the occurrence of the word .Orange. in a future query of the user has a tendency to indicate that the category .Fruit. is of interest. A users profile will be learned automatically from users search history. A user profile are learned from the users search history and a category hierarchy. The query is submitted to a search engine without specifying any category. The system determines the three categories which are most likely to match the interests of the user with the given user query. From these three categories, the user can pick the ones which are more suitable or he/she can decide to see the next three Categories. From this approach, it is possible to find the users intention.

3) Inference of User Intention
The personalized search is accomplished by mapping a user query to a set of categories, which reflect the users intention and serves as a context for the query, based on the user profile. The mapping is carried out as follows. First, the similarities between a user query and the categories representing the users' interests are computed. Next the categories are ranked in descending order of similarities.

4) User Click through Data Module
Click through data is the search result clicked by the user and the ranking function is employed to present the search results in some proper order according to the user's preferences. The first step in this method is preference mining, which discovers user's preferences of search results from click through data. The second step is the ranking function optimization, which optimizes the ranking according to the users preferences [1].Existing click through can be categorized into document based and concept based approaches. In both the methods, the user clicks can be used to infer the users Interest. Document based profiling methods try to estimate users. Document preferences.[5],[6],[7],[8]. Most document based methods focus on analyzing users. Clicking and browsing behaviors recorded in the users. Click through data [9].The proposed system first, extracts the concepts from the web snippets. The maximum number of keywords can be limited to six or seven. Secondly, it uses the Naive Bayesian classifier. Since the search engines can't understand the real intention, query expansion is very important and the system records the title, keyword, and description as the text contents of the page and it maintains an index file for each user. Since the click through data is spread overall the returned pages, it is essential to extract the user related pages from the all returned pages .This task can be accomplished with the help of the Naïve Bayes algorithm.

5) Term and Weight
Term frequency, is a numerical statistic which reflects how important a word is to a document in a collection. It is often used as a weighting factor in information retrieval and text mining. The value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to control for the fact that some words are generally more common than others. Variations of the weighing scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query. Can be successfully used for stop-words filtering in various subject fields including text summarization and classification.[3]. One of the simplest ranking functions is computed for each query term.

IV.SIMULATION RESULTS

In this section we evaluate and analyze the proposed system by conducting the required experiments. Tests were conducted module by module and the results are as shown in the following tables. The Result of the user profile module for the query with the word "Orange" is displayed in Table 1. From the below table we are able to understand that the user is very interested in the information related to "Orange Fruit" than the other items. The result generated from the click through data module is displayed in the Table 2.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

TABLE 1. Result Generated For the Query "Orange"

Document	Clicked Extracted	Data
1		Help and Support
2		Farmer
3		Flavor
4	$\sqrt{}$	Orange farms
5	$\sqrt{}$	Fruit
6		Business Service
7		Orange Tree
8	$\sqrt{}$	Image
9		Hotel
10		Bakery

TABLE 2. Result Generated From Click through Data Module

Extracted Data	Users Interest
Orange- Fruit	1
Orange- Image	2
Orange Farms	3

The above results were mapped in the third module and the results are as shown in the Table2. Final Result from the Mapping Module. From the Table 3 We are able to understand that the user is interested in information "Orange Fruit" than other items. From this new system it is possible to get more personalized web pages.

Concepts	Weight	Extracted	Data Weight
Fruit	0.67	Help and Support	0.24
Business Service	0.2	Farmer	0.43
Orange Tree	0.31	Flavor	0.44
Bakery	0.25	Orange farms	0.45
Image	0.5	Hotel	0.33

V.CONCLUSION AND FUTURE WORK

This new technique extracts the web pages based on two methods. The information extraction is based on User Profile and Click through data. The main advantage of this system is that it is possible to extract personalized web pages. An efficient user profile can improve the search engines performance by identifying the individual interest. Planned to conduct some more test to identify the drawbacks and to improve the efficiency of the system.

REFERENCES

1. Cheqian Chen, Kequan Lin, Heshan Li, Shoubin Dong, .Personalized Search Based on Learning User Click History ., Proc.9thIEEE International Conference on Cognitive Informatics[ICCI.10]978-1- 4222-1/10.
2. Fang Liu, Clement Yu, Weiyi Meng,. Personalized Web Search For Improving Retrieval Effectiveness., CIKM02.
3. M.Sperretta, S.Gauch,. Personalizing Search Based on User Search 4. Histories.Conference,2004.
5. Jian-Tao Sun,Hua-Jun Zeng,Huan Liu,.,CubeSVD:A Novel Approach to Personalized Web Search .. WWW 2005.
6. E.Agichtein, E.Brill, and S.Dumais,.Improving Web Search Ranking by Incorporating User Behavior information.,Proc.ACMSIGIR,2006
7. E.Agichtein, E.Brill ,and S.Dumais and R.Ragno,.Learning User Interaction Models for Predicting Web Search Result Preferences..Proc.ACM SIGIR,2006.
8. Z.Dou ,R.Song,and J.R Wen,.A Largescale Evaluation and Analysis of Personalized Search Strategies,.Proc.WWWConf.2007.
9. T.Joachims, .Optimizing Search Engine Queries., IEEE Trans. Knowledge and Data Eng.,Vol.20.no.11.pp.1505-1518,Nov.2008.
10. Kenneth Wai-Ting, Dik Lun Lee ,. Derving Concept-Based User Profiles from Search Engine Logs..,Digital Object Identifier no.10.1109/TKDE.2009.144.
11. R.Kraft, C.C.Chang, et al.. Searching with Contexts.In Proceedings of WWW.06, ACM Press, pp.477-486,2006.
12. Jie Yu, Fangfang Liu,. Mining User Context based on Interactive Computing for Personalized Web Search.In Proceedings of 2nd International Conference on Computer Engineering and Technology..978-1-4244-6349-7/10.