

Predicting User Goal in Personalized Web Search & Improvement Using Categorization

Mrs.Geetha.S¹, Prof.RachanaDhannawat²

¹(Mtech CST student, UMIT/ SNTD Womens university, Mumbai, India)

²(Asst Professor, CST Department, UMIT/ SNTD Womens university, Mumbai, India)

Abstract: For the web users retrieving the most relevant information from the Web become difficult because of the huge amount of documents available in various formats. Personalized Web search approach can be used to satisfy the requirements of the user. It is a process that adapts information or services provided by a Web to the needs of each specific or set of users, taking the facts of the knowledge gained from the users. In this paper a method for deriving user search objective using user feedback sessions and mapping of these feedback sessions to pseudo documents has been proposed. Clustering of these documents and applying cosine based similarity algorithm for calculating the rate of matching between user query and pseudo document for web URL can also be done in the later stages. The evaluation of the rate of performance of prediction is done using classified average precision algorithm.

Keywords: pseudo-documents, feedback sessions, clustering, classified average precision.

I. Introduction

Personalized Web search

Personalized Web search is a customization of search engine results created by a filter that takes into account potentially relevant information such as the user's history, location and preferences. Google introduced personalized search in 2004, promoting it as a way to understand exactly what you mean and give you back exactly what you want. Google uses a combination of information from user search history, bookmarks and personalized Google+ pages etc. Web Sites we tend to visit more frequently appear higher on search engine results page. Google also attempts to provide personalized search to users that do not have a Google account and do not have search history enabled. Microsoft Bing is another personalized search provider. Microsoft has many pages and services a user might already interact with. Those sites and services yield information for customization of both search and advertisement results. Personalized search is more useful for some types of searches than others. It's convenient, for example, for finding local services and finding information on topics that we have researched before. One of the limitations of personalized search is that it limits the user's view of the Web.

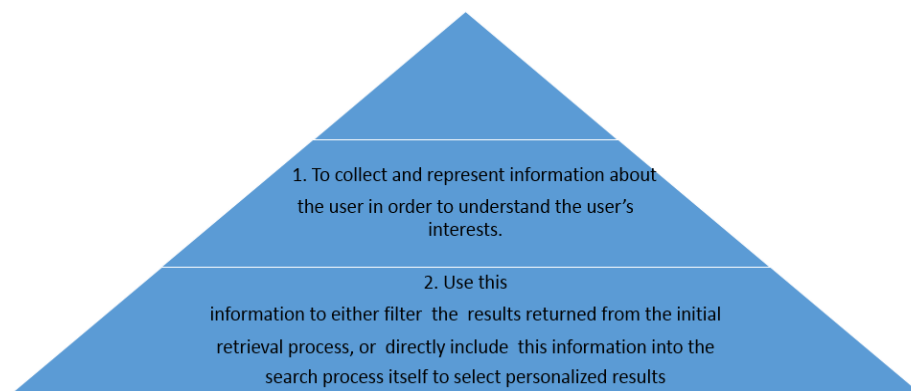


Fig 1: Steps in Personalized Search

II. Related Work

Zheng Lu, HongyuanZha, et all [1] presents an approach for deriving user search goal using feedback sessions. Each of the clicked URLs and therefore the unclicked ones before the last click are considered as user implicit feedbacks and taken under consideration to construct feedback sessions and map this to pseudo documents. The pseudo-documents will enrich the URLs with additional textual contents as well as the titles and

snippets. Based on these pseudo-documents, user search goals will be discovered and represented with some keywords. Better Web-page recommendation can be provided through semantic-enhancement by integrating the domain and Web usage knowledge of a website [2]. Domain knowledge can be represented through ontology or semantic network. A conceptual prediction model can be used to generate a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and Web usage knowledge.

Web dynamics can be analyzed based on web-related data browsed, collected, tagged, and semi-organized by end users using a bottom-up approach [3]. It produces search results based solely on user provided web-related data and their sharing among users. A bottom-up search engine starts from a core consisting of the most interesting part of the Web according to user opinions and incrementally and measurably improves its ranking. The user behavior detailing while interacting with the web are based on a top-down approach, where the entire Web, viewed as a vast collection of pages and interconnection links, is used to predict how the users interact with it. In the community web directory context, the Web directory is viewed as a thematic hierarchy and personalization is realized by constructing user community models on the basis of usage data [4]. The user communities can be modelled by combining the users' browsing behavior with thematic information from the Web directories. Enhancement of this modeling can be done by combining clustering and probabilistic approaches. The resulting community models take the form of Community Web Directories.

Different underlying levels of semantic uncertainty in terms of Web events can be identified, and then these can be utilized for recommending Webpages. Web event can be considered as a system composed of different keywords, and the uncertainty of this keyword is related to the uncertainty of the particular Web event. Based on keyword association linked network, Web event representation and Shannon entropy, the identification of different levels of semantic uncertainty and construction of a semantic pyramid (SP) [5] to express the uncertainty hierarchy of a Web event can be produced. Finally, a SP-based Webpage recommendation system can also be developed. Enhanced User Profile technique can be used for suggesting relevant pages to the user. The suggestions provided to the user with Enhanced User Profile are better than those obtained by using a User Profile. An Enhanced User Profile improves the User Profile by using the Domain Knowledge. For preparing the Enhanced User Profile we need to consider each URL of the User Profile, match it with Domain Knowledge URLs and add most relevant URLs to the Enhanced User Profile [6].

Dynamic user profile can be used to automatically update user profile. Combining it with collaborative filtering, the efficiency of search engine can be improved to great extent and it gives result according to user need and relevance information [7]. This technique can be considered for page recommendation which helps to retrieve search result and relevant document to user according to its need and preferences by diagnosing its web search behavior based on previous search history. Dynamic user profile can also be combined with clustering techniques [8] for personalized web search. Dynamic user profile can be build using long term user search histories in order to understand the behavior of user search which helps the search engine to know his interests and preferences. The dynamic user profiling can be done without explicit involvement of user.

Query reformulation and user profiling can be used in identifying relevant search term for particular user from his previous search history by analyzing web log file maintained in the server. These terms can be appended to user's ambiguous query [9]. Re-ranking of retrieved result can be done by identifying interest value of user on retrieved links. The identification of user interest on retrieved links can be done by combining the user interest value generated from Vector Space Model and actual rank of that link. Since users have diverse background on same query, it is difficult for some informative query to identify user's current intention. Based on user navigation pattern can facilitate user profile creation. Clustering and Classification techniques are used for discovering user navigation pattern [10]. Click through data which is recorded in search engine logs can be used to simulate user experiences in Web search. Time consumed for personalized search and normal search can also be compared.

Ref Paper	Methodology used	Advantage	Algorithm/method
Ref#1	Content Based and Link Based	More Accurate result and Low Complexity	K-means clustering & Classified Average precision
Ref#2	Semantic web personalization	Enhanced web page recommendation in terms of precision and user satisfaction	Ontology based
Ref#3	Bottom up approach	Incremental and measurable improvement in web page ranking	UserRank/PageRank
Ref#4	Web directory	Construction of community web directories based on web usage data.	cluster mining algorithm/ rule mining algorithm
Ref#5	Semantic web personalization	Recommendation of appropriate Webpages of Web events to visitors.	Apriori algorithm/ keyword extraction

			algorithm
Ref#6	Enhanced User Profile	Enhanced User Profile suggestions are better than User Profile.	K-nearest method
Ref#7	users browsing history	Provide search result relevant to the user by learning the of user search behavior.	clustering algorithm/ query expansion algorithm
Ref#8	Different information retrieval techniques for efficient personalized web search	Suggest the techniques for personalized web search according to the merits and demerits of various available techniques.	Hierarchical Algorithm/clustering algorithm/Content based filtering
Ref#9	Personalize web search result through query reformulation and user profiling.	Identify current interest of user by suggesting some relevant and irrelevant keywords to user.	Content based filtering/Re-Ranking Algorithm
Ref#10	User navigation pattern.	Search results based on user's interest by ranking the user clicks. Takes lesser time.	Ant based clustering algorithm/Apriori Algorithm

Table 1: Comparison of personalized Web search techniques

III. Proposed System

Web users normally issue keyword queries to search engines to fetch relevant information on specific topics. As users may have diverse backgrounds and different expectations for a given query, some search engines try to personalize their results to better match the overall interests and preferences of an individual user. In most of existing search engine, search result is retrieved by evaluating relative importance of links. Ranking algorithm calculate rank of links by their in edge and out edge links. Higher the in edges or out edges, higher the rank is assigned to link. But this approach might not be good as some links may disappear from the choice of links given to the user over a period of time because the ranking of the page would have been changed by the search engine. The search engine's perspective towards the user behavior will be the set of queries users submit based on factors such as the overall length of a query. This approach prevents them from looking beyond the query to get why the users are performing their searches. The "why" of user search behavior is actually essential for satisfying the user's information needs. Searching is a way to satisfy the user with the underlying goal they are expecting to achieve. User search goals can be considered as the set of information needs for a query. The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience. Due to its efficiency, many works about user search goals analysis have been invented.

In this paper, a method has been proposed for discovering the search objective of the user when he types a query on the search engine. After discovering the user objective, each user goal can be represented with some keywords. Interpreting user search goals for a query can be done by clustering the feedback sessions. The feedback session is defined as the series of both clicked and unclicked URLs and ends with the last URL that was clicked in a session from user click-through logs. Next these feedback sessions can be mapped into pseudo-documents which can accurately reflect user information needs. The existing approaches can be improved by categorizing the topics and creating pseudo documents according to the categorization. The clustering of these pseudo documents can be done to predict user search goals and these can be represented with some keywords. The pseudo document clustering can be achieved using k-means clustering algorithm. Prediction of user search goals can be mapped with some keywords in the pseudo document.

While deducing user search goals clustering of user feedback sessions gives more decisive results. Direct clustering of user search results or clicked URLs yield inadequate results. The divergent propagation of user search goals can be obtained smoothly after feedback sessions are clustered. The formation of the pseudo-document can be done by combining the enriched URLs in a feedback session. The pseudo-document can adequately mirror the information need of a user. This help us to identify the user search goal precisely. CAP (Classified average precision) is used to evaluate the performance of user search goal interpretation based on rearrangement of web search results. Using CAP the number of user search goals for a query can be determined. The user search goal is the information on different aspects of a query that user groups want to obtain. User search goals can be considered as the clusters of information needs for a query.

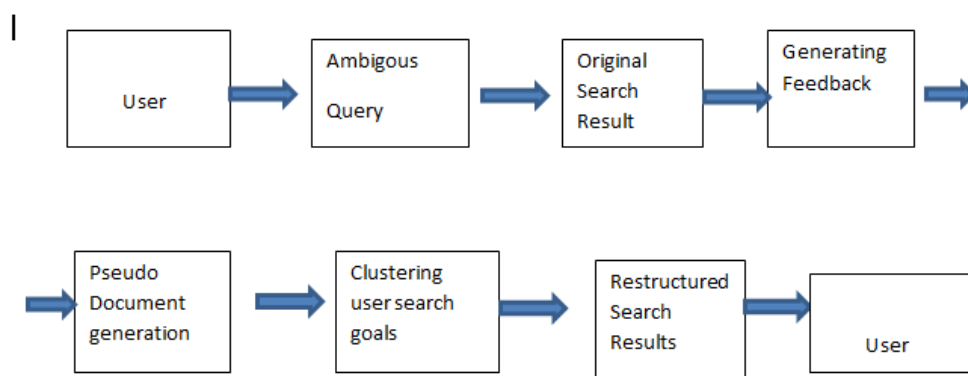


Fig 2: Flow Diagram of the Proposed System

IV. Conclusion

In this paper, a novel method has been proposed to predict user search goals for a query by gathering its feedback sessions represented by pseudo-documents. First, we Present feedback sessions to be analysed to deduce user search goals rather than hunting for results or clicked URLs. Both the clicked URLs and the unclicked ones before the last click are considered as user implicit responses and taken into account to concept feedback sessions. Therefore, FeedbackSessions can reflect user info needs more professionally. Second, we map feedback sessions to pseudo documents to approximate goal texts in user attentions. The pseudo documents can enrich the URLs with extra textual contents including the titles and brief descriptions. Based on these pseudo-documents, objective of user web pursuit can be discovered and depicted with some keywords.

References

- [1]. "A New Algorithm For Inferring User Search Goals With Feedback Sessions" By Zheng Lu, HongyuanZha, Xiaokang Yang, Weiyao Lin, And ZhaohuiZheng , pp.502-513,2013 IEEE
- [2]. "Web-Page Recommendation Based on Web Usage and Domain Knowledge" byThanh Sang Nguyen, Hai Yan Lu, and Jie Lu, pp.2574-2587, 2014 IEEE
- [3]. "A Collaborative Decentralized Approach to Web Search"by AthanasiosPapagelis and Christos Zaroliagis,pp 1271-1290 , 2012 IEEE
- [4]. "Personalizing Web Directories with the Aid of Web Usage Data" by Dimitriosierrakos and George Palioura, pp 1331-1343, 2010 IEEE
- [5]. "Uncertainty Analysis for the Keyword System of Web Events, byJunyuXuan, XiangfengLuo, Guangquan Zhang, Jie Lu, and ZhengXupp 829-842,2016 IEEE
- [6]. "Personalized web search using browsing history and domain knowledge"by Rakish Kumar and AditiSharan, pp. 2161-2174, 2014 IEEE
- [7]. "Efficient Technique for personalized web search using users browsing history" by Anoj Kumar and Mohd. Ashraf, 2015 IEEE
- [8]. "Personalized Web Search Engine using Dynamic User Profile and Clustering Techniques" by Anoj Kumar and Mohd. Ashraf, 2015 IEEE
- [9]. "A Novel Approach to Personalize Web Search through User Profiling and Query Reformulation" by KamleshMakvana, Pinal Shah and Shah, 2014 IEEE
- [10]. "User Profile Creation Based On Navigation Pattern for Modeling User Behavior with Personalized Search "byJosnaJojo and Dr.N.Sugana, 2013 IEEE
- [11]. "User Profile for Personalized Web Search", by C Liang, pp. 1847-1850, 2011 IEEE
- [12]. "Community Aware Personalized Web search", by O. Shafiq, R. Alhaji and I. G. Rokne, pp. 3351 - 355, 2010 IEEE.
- [13]. "Personalized Web Search with Location Preferences," by K.W.T. Leung, D.L. Lee, and W.-C.Lee, 2010 IEEE
- [14]. "A Comparative Study of Page Ranking Algorithms for Information Retrieval", by Ashutosh Kumar Singh, Ravi Kumar P, 2009 IEEE.
- [15]. "A Hybrid Approach of Personalized Web Information Retrieval", by Namita Mittal, RichiNayak, M.C. Govil, K.C. Jain, pp. 308-313 2010 IEEE.