International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Advanced Content Based Mobile Search Engine

Patil Anuja A.[1], M. M. Naoghare[2]

[1]M.E. Computer Engineering, SVIT Chincholi, Nashik, Maharashtra, India

[2]Assistant Professor, M.E. Computer Engineering, SVIT Chincholi, Nashik, Maharashtra, India

Abstract: ***A personalized search engine that captures the users preferences in the form of concepts by mining their click through data. Due to the value of location information in search, in this technique the user preferences are organized in an ontology-based, versatile user profile, which are used to adapt a personalized ranking function for rank adaptation of future search results. Based on the client-server model, presented is a detailed architecture and design for implementation of Personalized Search Engine. In design, the client collects and stores locally the click through data to protect privacy, whereas important tasks such as concept mining training, and re-ranking are performed at the Personalized Search Engine server. Moreover, the privacy issue is addressed by restricting the information in the user profile exposed to the server. This technique of Personalized Search Engine is prototyped on the Google Android platform.***

Keywords: Click through data, location search, PMSE, Mobile search engine, ontology, user, personalized profiling

1. Introduction

A major problem in personalized mobile search is that the contacts between the users and search engines are limited by the small form factors of the mobile devices. As a result, mobile users be liable to submit shorter, hence, more confusing queries compared to their web search counterparts. In order to revisit highly related results to the users, personalized mobile search engines must be able to profile the users' benefit and personalize the search results according to the users' profiles.

A practical approach to capturing a user's interests for personalization is to analyze the user's clickthrough data. Earlier developed a search engine personalization method based on users' concept preferences and showed that it is more effective than methods that are based on page preferences. Observing the need for different types of concepts, presented in the seminar is a personalized mobile search engine (PMSE) which represents different types of concepts in different ontologies. In particular, recognizing the significance of location information in mobile search, the concepts are separated into location concepts and content concepts. For example, If a user is planning to visit India may issue the query "Theater," and click on the search results about Theaters in India. From the clickthroughs of the query "Theater," PMSE can learn the user's content preference (e.g., "Ticket" and "Movies") and location preferences ("India"). Accordingly, PMSE will favor results that are concerned with Theater information in India for future queries on "Theater." The introduction of location preferences offers PMSE an additional element for capturing a user's interest and an opportunity to develop search quality for users.

To incorporate background information exposed by user mobility, the visited physical locations of users in the PMSE are also taken into account. Since this information can be easily obtained by GPS devices, GPS locations play an important role in mobile web search. For example, if the user is searching for Theater information, is currently located in "Pune, Maharashtra," his/her position can be used to personalize the search results to favor information about nearby Theaters. Here, observation is that the GPS locations (i.e., "Pune, Maharashtra") help reinforcing the user's location preferences (i.e., "India") derived from a user's search activities to provide the most relevant results. The presented framework is capable of combining a user's GPS locations and location preferences into the personalization process. To the best of knowledge, this technique is the first to propose a personalization framework that utilizes a user's content preferences and location preferences as well as the GPS locations in personalizing search results.

In this paper, a practically design for PMSE by adopting the meta search approach which relies on one of the commercial search engines, such as Google, Bing, Yahoo to perform an actual search is presented. The client is responsible for receiving the user's requests and submitting the requests to the PMSE server, showing the returned results, and collecting his/her clickthroughs in order to get his/her personal preferences. The PMSE server is responsible for handling important tasks such as forwarding the requests to a profitable search engine, as well as training and re ranking of search results before they are returned to the client. The user profiles for particular users are stored on the PMSE clients, thus preserving confidentiality to the users.

The main observation of this paper is as follows:

- This paper studies the exclusive characteristics of content and location concepts, and provides a reliable strategy using client-server architecture to join together them into a identical solution for the mobile situation.
- This system of personalized mobile search engine is an inventive approach for personalizing web search results. it utilizes both the content and location preferences By mining content and location concepts for user profiling, to personalize search results for a user.
- PMSE include a user's physical locations in the personalization process.
- A practical system is designed for PMSE. The design gains the server-client model in which user queries are forwarded to a PMSE server for processing the training and re ranking quickly.

- Privacy preservation is a challenging issue in PMSE, where users send their user profiles along with queries to the PMSE server to obtain personalized search results. The results show this proposal facilitates soft privacy preserving control, while maintaining good ranking quality.

Clickthrough data have been used in determining the users' preferences on their search results. clickthrough data for the query "Theater," composes of the search results and the ones that the user clicked on the content concepts and lis are the location concepts extracted from the corresponding results. Many presented personalized web search systems are based clickthrough data to decide users' preferences. Joachim's proposed to mine document preferences from clickthrough data. Later, Ng et al. planned to combine a spying technique together with a novel voting procedure to find out user preferences. More recently, Leung et al. introduced an effective approach to expected users' conceptual preferences from clickthrough data for personalized query suggestions.

2. Literature Survey

The differences between existing system and this PMSE are:

- Mainly existing location-based search systems, such as, require users to manually define their location preferences, or to manually prepare a set of location responsive topics. PMSE profiles equally of the user's content and location preferences in the ontology based user profiles, which are automatically learned from the clickthrough and GPS data without requiring extra efforts from the user.
- Presented is a new and realistic design for PMSE. To train the user profiles quickly and professionally, the design forwards user requests to the PMSE server to handle the training and re-ranking processes.
- Existing system on personalization do not address the issues of privacy protection. PMSE addresses this issue by controlling information in the client's user profile being exposed to the PMSE server using two privacy methods, which can control privacy easily, while maintaining good ranking quality.

Sr.No	Paper name	Advantages	Disadvantages
I.	**Personalized Concept-Based gathering of Search Engine Doubt.** - Introduced an effective approach to guess users' topic preferences from clickthrough data for personalized doubt suggestions.	- Proposed is a new personalized concept-based clustering technique that is able to obtain personalized query suggestions for individual users based on their conceptual profiles. –The technique makes use of clickthrough data and the concept relationship graph mined from web-snippet, which can be taking at the back end and do not add extra depress to users.	In the paper, instead of considering only query-concept pairs in the clickthrough data the relationships between users, queries, and concepts to obtain more personalized and accurate query suggestions can be considered. - Clickthrough data and concept link graphs can be directly included into the ranking algorithms of a search engine so that it can rank results accepted to individual users' interests.
II	**Accomplished doubt Processing in Geographic Web Search Engines** The particular problem is accomplished doubt processing in location-based search systems. A doubt is allowed with footprint that specifies the geographical area of into the user. Manifold algorithms are engage to rank the search results as a combination of a textual and a geographic score.	- In paper, results show that in many cases geographic query processing can be performed at about the same level of efficiency as text-only queries.	Need to study pruning techniques for geographic search engines which could combine early termination approaches from search. - Need a plan to study parallel geographic query processing on **gathering** of machine, it may be preferable to assign documents to participating nodes not at random, as commonly done by standard search engines, but based on an appropriate partitioning of the underlying geographic domain.
III	**Expansion Search Engines Using Clickthrough Data** - It is proposed to mine document preferences from clickthrough data.	The paper presented an approach to mining log files of WWW search engines with the goal of improving their retrieval performance automatically. -Here clickthrough data can provide training data in the form of relative preferences. - The paper derives an algorithm for learning a ranking function.	It is not clear in how far a single user could unkindly control the ranking function by repeatedly clicking on particular links.
IV	**Assay of Geographic Queries in a Search Engine Log** - Developed a classifier to classify geo and non-geo queries. In order to handle the queries that focus on position information, a number of position-based search systems designed for position queries have been proposed.	- In the paper, with improved understanding of users' query goals and websites' informational content, search engines can take solution to develop response relevance.	-There is a need to explore additional properties of the web sites associated with geographic queries, and of geographic search sessions, -Study need to be done on how user behavior on geo queries (particularly clickthrough data) can be harvested for better geographic search.
V	**PMSE: A Personalized Mobile Search Engine**	- The user preferences are organized in an versatile user profile, properly create	- To further enhance the personalization usefulness of PMSE; there is need for

	-PSME: captures the user's preferences in the form of concepts by mining their clickthrough data.	a personalized ranking function for rank adaptation of future search results. - To adapt to the user mobility, the user's GPS locations are included in the personalization process, which helps to develop recoup usefulness, especially for location queries	methods to develop regular travel patterns and query patterns from the GPS and clickthrough data.
VI	**Exploit user Basically Using Spy Voting for Search Engine Personalization** - It is proposed to combine a spying technique together with a novel voting procedure to determine user preferences.	- Proposed here is a SpyNB preference exploit algorithm, which is more useful and formative than the existing algorithms.	- There is essential to develop a more superior voting strategy by incorporating continuous probability into the voting procedure to replace the current binary voting method. - Also, a need to ensure the scalability of the training and optimization processes.
VII	**A Probability Topic-Based Ranking Structure for Position-sensitive Domain Information Betterment** - Proposed a probability topic-based structure for position-delicate domain information betterment.	-AlterNet of modeling position in width-height pairs, the model assumes that users can be interested in a set of position sensitive topics. - It recognizes the geographical effect distributions of topics, and models it using probability modifying Process is part.	Track
VIII	**Kokono Search: A Position Based Search Engine** - Proposed a position-based search system for web documents. Position information was show from the web documents, which was converted into width-height pairs.	- When a user submits any doubt together with a width height pair, the system creates a search circle centered at the specified width-height pair and retrieves documents containing position information within the search circle.	

3. Existing System

A mostly problem in mobile search engine is that the relations between the users and search engines are limited by the short factors of the mobile devices. Most of the previous work certain that all conception are of the same type. As a result, mobile users be liable to submit shorter, hence, more confusing queries compared to their web search counterparts. Basically demand for different types of conception, we present in this paper a personalized mobile search engine (PMSE) which represents different types of conception in different ontologies. In particular, known the importance of the location information in mobile search, we separate conception into location intellection and content intellection. To add in context information opens by user mobility, we also into account the visited physical locations of users in the PMSE. Since this information can be easily obtained by GPS devices it is thus suggest to as GPS locations. GPS technology is the main part for mobile web searching.

Disadvantages of Existing System

1) In an existing system, Trace GPS location is problematic.
2) Much difficulty in the privacy.
3) Mostly usable search engines return unexpected results to all users. But different users may have different information needs still for the same query.

4. Proposed System

But most of the previous work implicit that all concepts are of the same type. We divide concepts into location concepts and content concepts to recognize information valuable So far there have been many papers written & researched on search engines. There is great development in this field. But there is only one such paper written so far on Personalized Mobile Search Engine [PMSE]. In this paper, we propose a realistic design for PMSE by gain the Meta search approach which relies on one of the commercial search engines, like Yahoo, MSN, Google or Ask, to perform an original search. The user is responsible for receiving the user's information, finding the information to the PMSE server, showing the results, and saving his/her clickthrough in order to get his/her personal preferences. The PMSE server, on the other hand, is liable for handling heavy tasks such as forwarding the requests to a commercial search engine, as well as guidance and re ranking of search results before they are returned to the users. The user profiles for exact users are stored on the PMSE clients, thus secure privacy to the users. PMSE has been old typed with PMSE clients on the Google Android platform and the PMSE server on a PC server to accept the proposed ideas Studies the unique characteristics of content and location concepts, and provides a consistent strategy using client-server architecture to integrate them into a uniform solution for the mobile environment.

By exotic content and location concepts for user profiling, it utilizes both the content and location preferences to personalize search results for a user.

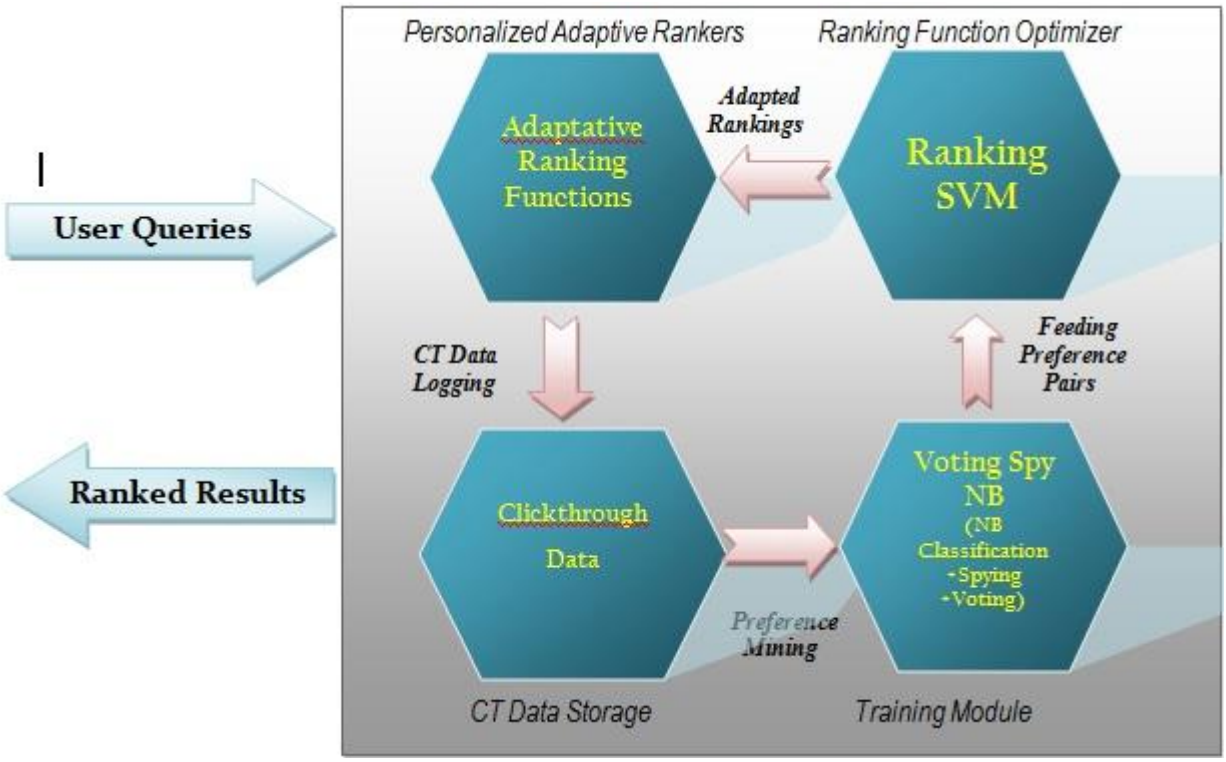

Figure1. General Process of Search Engine Adaptation

5. Algorithm

The concepts and clickthrough data are collected from past search activities; user's preference can be learned. In this section, we review two alternative preference mining algorithms, namely, Joachim's Method and SpyNB Method that we accept in our personalization framework.

Joachim's Method

This Joachim's method assumes that a user would scan the search result list from start to end. If a user avoid a document bs at rank r but clicks on document bs at rank r where r < s, he/she must have read br' web small topic and decided to avoid it. If Joachim's method concludes that the user prefers bs to document br (denoted as br <v0 bs, where v0 is the user's preference order of the documents in the search result list).we can get a set of document preference pairs.

The document preference pairs are then employed in a ranking SVM algorithm to learn a linear feature weight vector, which is composed of either content or location concept features, to rank the search results according to the user's content and location preferences.

Spy Naive Bayes (SpyNB) Method

The problem now can be formulated as how to identify the reliable negative examples from an unlabeled set using only positive and unlabeled data. Recently, partly organized classification provides a novel paradigm for constructing classier using positive examples and a large set of unlabeled examples.

Finding reliable negative examples can be solved by partially supervised classification techniques, such as Spy technique, 1-DNF, and the Rocchio method In particular, we incorporate the spy technique with Naive Bayes to design a Spy Naive Bayes (SpyNB) algorithm for identifying the reliable negative examples. We choose the spy technique, because it has been shown to be effective for common text classification. However, clickthrough data have some unique characteristics compared to common texts. For instance, the titles and abstracts are both very short texts, and the size of positive set (the number of clicked links) is also very small. accordingly, the identified RN is not undependable if only a small portion of positive examples are used as spies. Thus we further employ a voting procedure to strengthen SpyNB In this section, we elaborate on the SpyNB algorithm in detail.

We first illustrate how the Naive Bayes (NB for short) is adapted in clickthrough analysis as follows. Let "+" and "-"denote the positive and negative classes, respectively. Let L = {l1, l2, ... ,lN} denote a set of N links (documents) in the search results. Each link li can be explaind as a word vector, W = (w1,w2,. . . . . ,wM), in the vector space model , where we count the occurrences of wi appearing in the titles, abstracts and URLs,Then, a NB classifier is built by estimating the prior probabilities (Pr(+) and Pr(-)), and likelihood (Pr(wj|+) and Pr(wj|-)), as shown in Algorithm 1.

In Algorithm 1, ±(+jli) indicates the class label of link li. Its value is 1 if li is positive; and 0 otherwise. Num(wj ; li) is a function counting the number of keywords wj appearing in link li. ¸ is the smoothing factor, where ¸ λ= 1 is known as Laplacian smoothing , which we use in our experiments.

6. Conclusion

To adapt to the user mobility, we incorporated the users GPS locations in the personalization process. We find out that

GPS locations help to improve recovery effectiveness, especially for location queries. In this paper we also proposed two privacy parameters, min Distance and exp Ratio, to tackle privacy issues in PMSE by allowing users to control the amount of personal information exposed to the PMSE server. The privacy parameters facilitate smooth control of privacy exposure while maintaining good ranking quality for future work, we will analyze methods to exploit regular travel patterns and query patterns from the GPS and clickthrough data to further enhance the personalization effectiveness of PMSE.

7. Future Scope

In the experiments, we personalized a meta search engine using SpyNB. Both the offline and online results showed that our approach and algorithm are effective: the personalized meta search engine improved the ranking quality and was able to cater for users species interests. we could further gear the spying technique toward the RSVM directly to mine preferences by voting on the rank order, which is a lightweight approach to the problem. Since the new direction of personalizing a search engine through adapting its ranking function has just emerged, many extensions can be further investigated. As evident in our experiments, the linear ranking function is quite effective for search engine personalization; however, the power of a linear ranking function is still limited compared to more sophisticated ranking functions, for example, a polynomial ranking function.

We also aim to develop the existing prototype into a full-edged adaptive search engine. We are considering incremental updates on the ranking function. In other words, whenever the user clicks on the result of a query, the training process is invoked, leading to the optimization of the corresponding ranker. The challenge is that we need to ensure the scalability of the training and optimization processes.

References

[1] KennethWai-Ting Leung, Dik Lun Lee, andWang-Chien Lee PMSE: A Personalized Mobile Search Engine IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013

[2] "Google Personalized Results Could Be Bad for Search". Network world Retrieved July 12, 2010

[3] E. Agichtein, E. Brill and S. Dumais, Improving web search ranking by incorporating user behavior information, in Proc. Of ACM SIGIR Conference, 2006.

[4] Ng,W,Deng, L. and Lee D.L. 2007 Mining user preference using spy voting for search engine personalization. ACM Trans. Intern

[5] Mr. V. Dilip kumar Mr. S. Ravi kumar School of Computing Sciences, Hindustan University, Chennai Ontology Based Android Mobile Search Engine International Journal of Engineering Research Technology (IJERT) Vol. 2 Issue 4, April 2013.

[6] National geospatial. http://earth-info.nga.mil/

[7] svmlight. http://svmlight.joachims.org/.

[8] Q.Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of Geographic Queries in a Search Engine Log," Proc. First Int'l Workshop Location and the Web (LocWeb) 2008

[9] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing In Geographic Web Search Engines," Proc. Int'l ACM SIGIR Conf Research and Development in Information Retrieval (SIGIR), 2006