Analysis of Personalized Queries in a Mobile Search Engine

Malleshappa M R
Department Of Computer Science & Engineering
Visvesvaraya Technological University, Belgaum
KVGCE, SULLIA, DK

Savitha C K
Assoc. Professor
Department Of Computer Science & Engineering
Visvesvaraya Technological University, Belgaum
KVGCE, SULLIA, DK

Dr Antony P J
Director, PG Studies
Visvesvaraya Technological University, Belgaum
KVGCE, SULLIA, DK

***Abstract* - Mobile search engine is a metasearch engine that captures the user's preference in the form of concepts by mining their clickthrough data. It extracts the importance of location information in mobile search and classifies these concepts into content concepts and location concepts. By positioning by GPS, user's locations are used to supplement the location concepts in search engine. The user preferences are organize in an ontology-based, multifacet user profile, which are used to adapt a Personalizing ranking function for rank adaptation of search results. Mobile Search Engine characterizes the diversity of the concepts associated with a query and their relevances to the user's need. It associated with four entropies are introduced to balance the weights between the content and location facets. Based on the client-server model, it consists of a detailed architecture and design for implementation. In this design, the client collects and stores locally the clickthrough data to protect privacy, whereas heavy tasks such as concept extraction, training, and reranking are performed at the server. It addresses the privacy issue by restricting the information in the user profile exposed to the server with two privacy parameters are minDistance and expRatio. It prototypes search engine on the Google Android platform. It is an innovative approach for personalizing web search results. By mining content and location concepts for user profiling, it utilizes both the content and location preferences to personalize search results for a user. Mobile Search Engine incorporates a user's physical locations in the personalization process. It is using a GPS location helps to improve retrieval effectiveness for location queries.**

Index Terms – Spy NB, Location ontology, Content ontology, RSVM, Mobile search engine.

I. INTRODUCTION

Mobile search is the interactions between the users and search engines are limited by the small form factors of the mobile devices. As a result, mobile users tend to submit shorter, hence, more ambiguous queries compared to their web search counterparts. In order to return highly relevant results to the users, mobile search engines must be able to profile the user's interests and personalize the search results according to the user's profiles.

Personalization is analyzing the user's clickthrough data. However, most of the previous work assumed that all concepts are of the same type. Observing the need for different types of concepts, in this paper present an analysis of personalized queries in mobile search engine which represents different of concepts in different ontologies. In particular, it recognizing the importance of location information in mobile search and it separate concepts into location concepts and content concepts. For example, a user who is planning to visit Canada may issue the query "hotel," and click on the search results about hotels in Canada. From the clickthroughs of the query "hotel," PMSE can learn the user's content preference like "room rate" and "facilities" and location preferences like "Japan". Accordingly, Search Engine will favor results that are concerned with hotel information in Canada for future queries on "hotel." It introduces a location preferences offers search engine an additional dimension for capturing a user's interest and an opportunity to enhance search quality for users.

Search engine is incorporates context information revealed by user mobility and it also take into account the visited physical locations of users. Since this information can be conveniently obtained by GPS devices, it is hence referred to as GPS locations. GPS locations play an important role in mobile web search. For example, if the user, who is searching for hotel information, is currently located in "London, England," his/her position can be used to personalize the search results to favor information about nearby hotels. Here, he/she can see that the GPS locations as "London, England", help reinforcing the user's location preferences like "Japan" derived from a user's search activities to provide the most relevant results. This proposed framework is capable of combining a user's GPS locations and location preferences into the personalization process. This paper 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.

The client is responsible for receiving the user's requests, submitting the requests to the Search engine server, displaying the returned results, and collecting his/her clickthroughs in order to derive his/her personal preferences. The server is responsible for handling heavy tasks such as forwarding the requests to a commercial search engine, as well as training and reranking of search results before they are returned to the client. The user profiles for specific users are stored on the Search engine clients, thus preserving privacy to the users. It

ISSN: 2278 – 7798

has been prototyped with clients on the Google Android platform and the Search engine server on a PC server to validate the proposed ideas.

It also recognizes that the same content or location concept may have different degrees of importance to different users and different queries. To formally characterize the diversity of the concepts associated with a query and their relevance's to the user's need, this paper introduce the notion of content and location entropies to measure the amount of content and location information associated with a query. Similarly, to measure how much the user is interested in the content and location information in the results and proposes click content and location entropies. Based on these entropies, it develops a method to estimate the personalization effectiveness for a particular query of a given user, which is then used to strike a balanced combination between the content and location preferences. The results are reranked according to the user's content and location preferences before returning to the client.

II. RELATED WORK

Clickthrough data have been used in determining the user's preferences on their search results. Many existing personalized web search systems are based click-through data to determine user's preferences and mine document preferences from clickthrough data. Later, Ng et al proposed to combine a spying technique together with a novel voting procedure to determine user preferences and introduced an effective approach to predict user's conceptual preferences from clickthrough data for personalized query suggestions.

Search queries can be classified as content or location queries. Gan et al. developed a classifier [4] to classify geo and non-geo queries. It was found that a significant number of queries were location queries focusing on location information. In order to handle the queries that focus on location information, a number of location-based search systems designed for location queries have been proposed. Yokoji [17] proposed a location-based search system for web documents. Location information was extracted from the web documents, which was converted into latitude-longitude pairs. When a user submits a query together with a latitude-longitude pair, the system creates a search circle centered at the specified latitude-longitude pair and retrieves documents containing location information within the search circle. Later on, Chen et al studied the problem of efficient query processing in location-based search systems. A query is assigned with a query footprint that specifies the geographical area of interest to the user. Several algorithms are employed to rank the search results as a combination of a textual and a geographic score. More recently, Li et al [7] proposed a probabilistic topic-based framework for location-sensitive domain information retrieval. Instead of modeling locations in latitude-longitude pairs, the model assumes that users can be interested in a set of location-sensitive topics.

III. PROPOSED METHOD

Mobile Search Engine profiles both 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. It proposes and implements a new and realistic design for Mobile Search Engine. To train the user profiles quickly and efficiently and forwards user requests to the server to handle the training and reranking processes. PMSE addresses this issue by controlling the amount of information in the client's user profile being exposed to the Mobile Search Engine server using two privacy parameters, which can control privacy smoothly, while maintaining good ranking quality.

Mobile Search Engine using client-server architecture (Fig. 1), which meets three important requirements. First, computation-intensive tasks, such as RSVM training, should be handled by the PMSE server due to the limited computational power on mobile devices. Second, data transmission between client and server should be minimized to ensure fast and efficient processing of the search. Third, clickthrough data, representing precise user preferences on the search results, should be stored on the PMSE clients in order to preserve user privacy.

In the PMSE's client-server architecture, PMSE clients are responsible for storing the user clickthroughs and the ontologies derived from the PMSE server. Simple tasks, such as updating clickthoughs and ontologies, creating feature vectors, and displaying reranked search results are handled by the PMSE clients with limited computational power. On the other hand, heavy tasks, such as RSVM training and reranking of search results, are handled by the PMSE server. Moreover, in order to minimize the data transmission between client and server, the PMSE client would only need to submit a query with the feature vectors to the PMSE server, and the server would automatically return a set of reranked search results according to the preferences stated in the feature vectors. The data transmission cost is minimized, because only the essential data are transmitted between client and server during the personalization process.

PMSE's design consists of some issues as limited computational power on mobile devices and data transmission minimization. Since the ontologies can be derived online at the PMSE server, an alternative system design is for the user to pass only the clickthrough data to the PMSE server. Mobile Search Engine using concepts to model the interests and preferences of a user. Since location information is important in mobile search, the concepts are further classified into two different types, namely, content concepts and location concepts. The concepts are modeled as ontologies, in order to capture the relationships between the concepts. Only can represent concepts but also capture the relationships between concepts.

A. Content Ontology

In content concept extraction method first extracts all the

ISSN: 2278 – 7798

keywords and phrases from the web-snippets arising from q. If a keyword/phrase exists frequently in the web-snippets arising from the query, treat it as an important concept related to the query, as it coexists in close proximity with the query in the top documents.

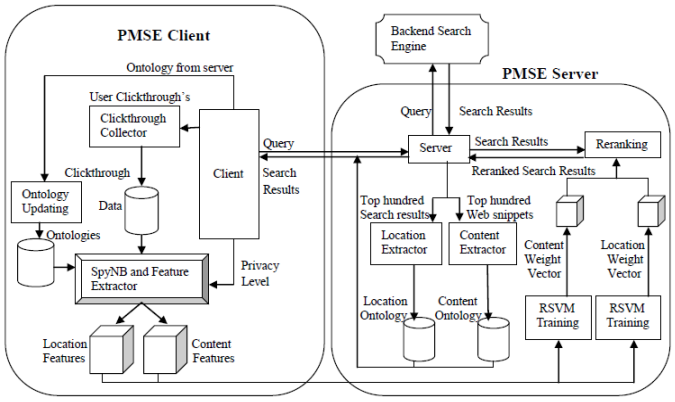

Fig. 1 The general process flow of Mobile Search Engine

Content ontology (Fig. 2) created for the query "hotel," where content concepts linked with a one-sided arrow are parent-child concepts, and concepts linked with a double-sided arrow are similar concepts. It shows the possible concept space determined for the query "hotel," while the clickthrough data determine the user preferences on the concept space. In general, the ontology covers more than what the user actually wants. The concept space for the query "hotel" consists of "map," "reservation," "room rate," etc. If the user is indeed interested in information about hotel rates and clicks on pages containing "room rate" and "special discount rate" concepts, the captured clickthrough favors the two clicked concepts. Feature vectors containing the concepts "room rate" and "special discount rate" as positive preferences will be created corresponding to the query "hotel." As indicated, when the query is issued again later, these feature vectors will be transmitted to the PMSE server and transformed into a content weight vector to rank the search results according to the user's content preferences.

B. *Location Ontology*

In this approach for extracting location concepts is different from that for extracting content concepts. It observes two important issues in location ontology formulation. First, a document usually embodies only a few location concepts, and thus only very few of them co-occur with the query terms in web-snippets. To alleviate this problem, we extract location concepts from the full documents. Second, the similarity and parent-child relationship cannot be accurately derived statistically because the limited number of location concepts embodied in documents. Furthermore, many geographical relationships among locations have already been captured as facts. Thus, we obtain about 17,000 city, province, region, country names and create predefined location ontology among these locations. It organizes all the cities as children under their provinces, all the provinces as children under their regions, and all the regions as children under their countries. The statistics of our location ontology are provided in Table 1.

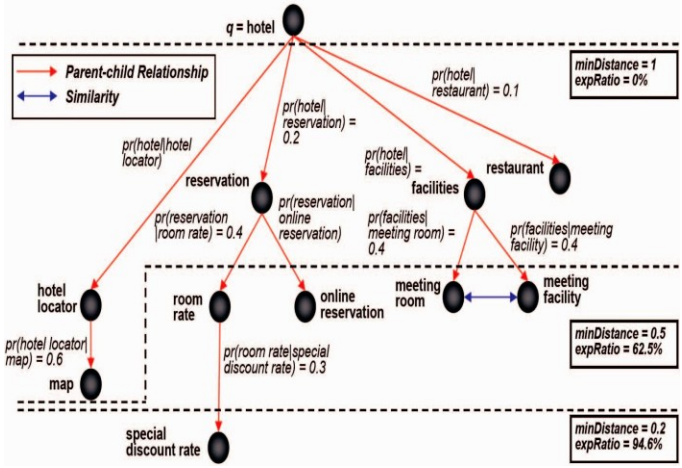

Fig. 2 Content Ontology of hotel queries

TABLE 1

STATISTICS OF THE LOCATION ONTOLOGY

No. of Countries	7	Total No. of Nodes	6899
No. of Regions	90	Country-Region Edges	90
No. of Provinces	699	Region-Province Edges	959
No. of Towns	0003	Province-City Edges	4897

C. *Ranking Support Vector Machine (RSVM)*

RSVM should be handled by the PMSE server due to the limited computational power on mobile devices. Ranking SVM is employed to learn a personalized ranking function for rank adaptation of the search results according to the user content and location preferences. For a given query, a set of content concepts and a set of location concepts are extracted from the search results as the document features.

D. *SpyNB (Spy Naive Bayes)*

SpyNB (Spy Naive Bayes), to generate a set of preferences that are then fed into the RSVM (Ranking Support Vector Machine) for optimizing the ranking function for the user. SpyNB discovers the fragment preference pairs as constraints that are fed into the RSVM. SpyNB is an elective means to generate the positive and negative datasets, from which accurate preference fragment pairs can be derived for optimizing the ranking function.

ISSN: 2278 – 7798

IMPLEMENTATION

In search engine, user profiles need to create for a login a new user and send a query along with user identification to database. Server creates a database for particular proposed models are location and contents ontologies using GPS by longitude-latitude. Server retrieves all matching search results using RSVM training. By using Spy Nb algorithm, user retrieve the prefered locations and its contents by reranking function. In this paper mainly design by the ontologies and databases.

RESULTS

Mobile search engine is effictively used by the user to retrive the particular user prefered locations and its contents. It is mainly contributed by reranking function to retrive the user preffered good locations and its contents in the form of ontologies using GPS and lattitude and longitude pair. Basically mobile search engine is to retrieve user preferred or as best of the locations and its contents.

CONCLUSION

Mobile search engine is mainly using in android platform to extract and learn a user's content and location preferences based on the user's clickthrough. User mobility is adapted and incorporated the user's GPS locations in the personalization process. It observed that GPS locations help to improve retrieval effectiveness, especially for location queries and also proposed two privacy parameters, minDistance and expRatio, to address 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.

REFERENCES

[1] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.

[2] 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.

[3] K.W. Church, W. Gale, P. Hanks, and D. Hindle, "Using Statistics in Lexical Analysis," Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, Psychology Press, 1991.

[4] 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.

[5] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.

[6] K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences," Proc. IEEE Int'l Conf. Data Mining (ICDE), 2010.

[7] K.W.-T. Leung, W. Ng, and D.L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 11, pp. 1505-1518, Nov. 2008.

[8] H. Li, Z. Li, W.-C. Lee, and D.L. Lee, "A Probabilistic Topic-Based Ranking Framework for Location-Sensitive Domain Information Retrieval," Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2009.

[9] B. Liu, W.S. Lee, P.S. Yu, and X. Li, "Partially Supervised Classification of Text Documents," Proc. Int'l Conf. Machine Learning (ICML), 2002.

[10] W. Ng, L. Deng, and D.L. Lee, "Mining User Preference Using Spy Voting for Search Engine Personalization," ACM Trans. Internet Technology, vol. 7, no. 4, article 19, 2007.

[11] J.Y.-H. Pong, R.C.-W. Kwok, R.Y.-K. Lau, J.-X. Hao, and P.C.-C. Wong, "A Comparative Study of Two Automatic Document Classification Methods in a Library Setting," J. Information Science, vol. 34, no. 2, pp. 213-230, 2008.

[12] C.E. Shannon, "Prediction and Entropy of Printed English," Bell Systems Technical J., vol. 30, pp. 50-64, 1951.

[13] Q. Tan, X. Chai, W. Ng, and D. Lee, "Applying Co-Training to Clickthrough Data for Search Engine Adaptation," Proc. Int'l Conf. Database Systems for Advanced Applications (DASFAA), 2004.

[14] J. Teevan, M.R. Morris, and S. Bush, "Discovering and Using Groups to Improve Personalized Search," Proc. ACM Int'l Conf. Web Search and Data Mining (WSDM), 2009.

[15] E. Voorhees and D. Harman, TREC Experiment and Evaluation in Information Retrieval. MIT Press, 2005.

[16] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. Int'l Conf. World Wide Web (WWW), 2007.

[17] S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc. Int'l Conf. World Wide Web (WWW), 2001.

ISSN: 2278 – 7798