A Review on Personalized Mobile Search Engine Using Location Concept and Content Concept

Namrata G Kharate
ME-Computer-II
MCOERC, Nasik-India

Prof. S. A. Bhavsar
Assistant Prof.Computer Deptt.
MCOERC, Nasik-India

***Abstract*— Now a days there is A major problem in mobile search is that 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 users' interests and personalize the search results according to the user's profiles.**

A personalized mobile search engine (PMSE) that captures the users' preferences in the form of concepts by mining their click through data. Due to the importance of location information in mobile search, In this paper PMSE classifies these concepts into content concepts and location concepts. The user preferences are organized in an ontology-based, multifacet user profile, which are used to adapt a personalized ranking function for rank adaptation of future search results. To characterize the diversity of the concepts associated with a query and their relevances to the user's need, four entropies are introduced to balance the weights between the content and location facets.In this system, 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 PMSE server. The privacy issues are address by restricting the information in the user profile exposed to the PMSE server with two privacy parameters

Keywords: **Clickthrough data, concept, location search, mobile search engine, ontology, personalization, user profiling**

1. INTRODUCTION

Social Network is a social structure made of individuals called nodes, which are connected by one or more specific types of interdependency, such as friendship, kinship, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige [1]. Social Network's link represents not only the flow between personal information, but the relation status through quantitative expression. The overall graph model of Social Network is composed of many nodes and the links that connect them, and each node's direct/indirect connection forms the entire network.

However, the current Personalized Systems based on Social Network were designed and constructed under the PC and it didn't provide the step by step transferring methods from PC to Smartphone. To solve these problems, this research actively analyzes an individual's characteristic based on the Social Network environment and develops a Personalized Information Retrieval System which can search for what a user wants accurately. Personalized Information Retrieval System for efficient personalized information provision proposed in this study differs from existing ones in methodology as follow:

Firstly, as the system is built on the basis of NFC (Near field communication), it attempts to provide its own custom service fast and easily using its information stored in NFC. Once SNS and NFC Smartphone are associated with each other, payment is made by touching a NFC tag when visiting well known restaurants, and the information recorded in SNS is supposed to provide search results customized to individual's tastes and preferences when carrying out a search in individualized search system. That is, typing the same search keyword may bring different search results on NFC Smartphone as individuals have different preferences.

Secondly, the existing Personalized Information Retrieval System fails to analyze the search system using Smartphone in Social Network environment. With an increasing number of web users using Smartphone and its individualized service under research, Smartphone environment does not provide user's search rankings suited to personal preferences. For example, when a user who wants to come by a pasta restaurant offering pasta for about 10$ and listens to rock music asks for information search via Smartphone, search results should also be prioritized and provided in favor of user's personalization taste. But, the existing systems do not show search rankings in consideration of individual's tastes and tastes. Therefore, in this study differentiated search results are provided on the basis of personalization information in User Profile Registry when a user carries out a search using Smartphone in Social Network environment.

Finally this research attempts to correct uncertain or vague relation between users on the existing Social Network environment and promote a more accurate and personalized information feeding, by suggesting a Personalized Information Retrieval System using Social Network's quantitative model.

Most of the current web information is developed based on HTML. Semantic assigns a meaning to a document and the Semantic Web is a high end automated intellectual technology that allows not only humans but machines to understand information. In order for a machine to take information from web and work, there has to be simple semantics for the machine to process. Semantic Web exists to express such semantics in a standardized method. Semantic

Web's basic data model consists of resource, property type, and property value. Each resource is equivalent to a single object, has various property types and values, and each object forms an organic link through property types and values. Studying the link allows the relationship between nodes to be inferred and analyzed. Unlike the existing Semantic Web, Social Network has a structure of object, sub object, property and relation. A node (person), which is an object, is composed of concept and role in a form of sub object that has various characteristics. Existing Social Network model has a limit on expressing quantitative information; hence it is not a solution for its uncertain or vague issues of Social Network. Exclusion of comments and links leads to only incomplete relations to be formed. In summary Social Network's basic structure system has fragmentary correlations between objects and does not support sub system for each object, which cripples an object from forming organic relations with others, making it difficult to comprehensively reflect personal information. On the other hand, Semantic Web has certain properties as sub system, which materializes objects. It creates a relationship regulation among them, enabling extensive relation inference, and furthermore the quantitative model would upgrade Social Network's basic structure for more various relationship analysis or inference.

2. RELATED WORK

Clickthrough data have been used in determining the users' preferences on their search results. Table 1, showing an example clickthrough data for the query "hotel," composes of the search results and the ones that the user clicked on (bolded search results in Table 1). As shown, cis are the content concepts and lis are the location concepts extracted from the corresponding results

TABLE 1
Clickthrough for the Query "Hotel"

Doc	Search Results	c_i	l_i
d_1	Hotels.com	room rate	*international*
d_2	**JapanHotel.net**	**reservation, room rate**	**Japan**
d_3	Hotel Wiki	accommodation	*international*
d_4	US Hotel Guides	map, room rate	USA, California
d_5	Booking.com	online reservation	USA
d_6	**JAL Hotels**	**meeting room**	**Japan**
d_7	**Shinjuku Prince**	**facility**	**Japan, Shinjuku**
d_8	Discount Hotels	discount rate	*international*

Many existing personalized web search systems [2], [3], [4], [5] are based clickthrough data to determine users' preferences. Joachims [3] proposed to mine document preferences from clickthrough data. Later, Ng et al. [4] proposed to combine a spying technique together with a novel voting procedure to determine user preferences. Search queries can be classified as content (i.e., non-geo) or location (i.e., geo) queries. Examples of location queries are "hong kong hotels," "museums in london," and "virginia historical sites." In [6], Gan et al. developed a classifier 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 [8] 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.

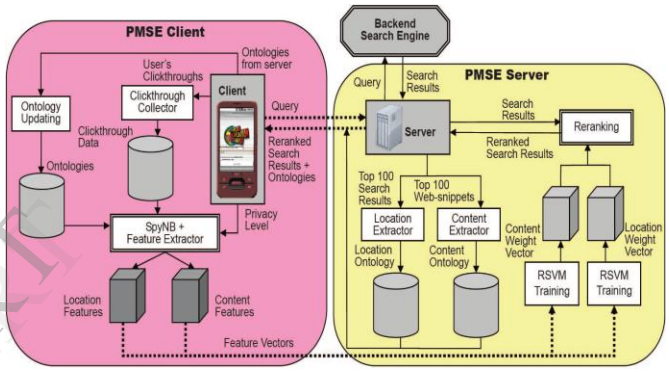

Fig. 1. The general process flow of PMSE

More recently, Li et al. [9] 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 locationsensitive topics. It recognizes the geographical influence distributions of topics, and models it using probabilistic Gaussian Process classifiers.

3. Proposed system

However, most of the previous work assumed that all concepts are of the same type. We separate concepts into location concepts and content concepts to recognize information importance So far there have been many papers written & researched on search engines. There is tremendous evolvement in this field. But there is only one such paper written so far on Personalised Mobile Search Engine [PMSE]. In this paper, we propose a realistic design for PMSE by adopting the metaearch approach which replies on

one of the commercial search engines, such as Google, Yahoo, or Bing, to perform an actual search. The client is responsible for receiving the user's requests, submitting the requests to the PMSE server, displaying the returned results, and collecting his/her clickthrough in order to derive his/her personal preferences. The PMSE server, on the other hand, 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 PMSE clients, thus preserving privacy to the users. PMSE has been prototyped with PMSE clients on the Google Android platform and the PMSE server on a PC server to validate the proposed ideas Studies the unique characteristics of content and location concepts, and provides a coherent strategy using client-server architecture to integrate them into a uniform solution for the mobile environment.

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

The differences between existing works and ours are:

- Most existing location-based search systems require users to manually define their location preferences or to manually prepare a set of location sensitive topics. PMSE 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.
- We propose and implement a new and realistic design for PMSE. To train the user profiles quickly and efficiently.
- Existing works on personalization do not address the issues of privacy preservation. PMSE addresses this issue by controlling the amount of information in the client's user profile being exposed to the PMSE server using two privacy parameters, which can control privacy smoothly, while maintaining good ranking quality.

4. SYSTEM DESIGN

Fig. 2 shows PMSE's client-server architecture, 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 clickthrough 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 together 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.

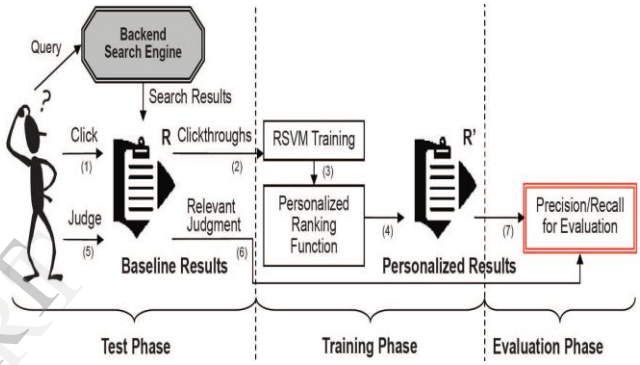

Figure 2: Flow of the evaluation process

The data transmission cost is minimized, because only the essential data (i.e., query, feature vectors, ontologies and search results) are transmitted between client and server during the personalization process. PMSE's design addressed the issues: 1) limited computational power on mobile devices, and 2) data transmission minimization.

PMSE consists of two major activities:

1. Reranking the search results at PMSE server. When a user submits a query on the PMSE client, the query together with the feature vectors containing the user's content and location preferences (i.e., filtered ontologies according to the user's privacy setting) are forwarded to the PMSE server, which in turn obtains the search results from the back-end search engine (i.e., Google). The content and location concepts are extracted from the search results and organized into ontologies to capture the relationships between the concepts. The server is used to perform ontology extraction for its speed. The feature vectors from the client are then used in

RSVM training to obtain a content weight vector and a location weight vector, representing the user interests based on the user's content and location preferences for the reranking. Again, the training process is performed on the server for its speed. The search results are then reranked according to the weight vectors obtained from the RSVM training. Finally, the reranked results and the extracted ontologies for the personalization of future queries are returned to the client.

2. Ontology update and clickthrough collection at PMSE client. The ontologies returned from the PMSE server contain the concept space that models the relationships between the concepts extracted from the search results. They are stored in the ontology database on the client.1 when the user clicks on a search result, the clickthrough data together with the associated content and location concepts are stored in the clickthrough database on the client. The clickthroughs are stored on the PMSE clients, so the PMSE server does not know the exact
set of documents that the user has clicked on. This design allows user privacy to be preserved in certain degree. Two privacy parameters, minDistance and expRatio, are proposed to control the amount of personal preferences exposed to the PMSE server. If the user is concerned with his/her own privacy, the privacy level can be set to high so that only limited personal information will be included in the feature vectors and passed along to the PMSE server for the personalization. On the other hand, if a user wants more accurate results according to his/her preferences, the privacy level can be set to low so that the PMSE server can use the full feature vectors to maximize the personalization effect.

5. CONCLUSIONS

The proposed personalized mobile search engine 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. The possible outcome will improve retrieval effectiveness for location queries (i.e., queries that retrieve lots of location information). For future work, we will investigate methods to exploit regular travel patterns and query patterns from the GPS and clickthrough data to further enhance the personalization effectiveness of PMSE.

REFERENCES

[1] Meishan, H.; Aixin, S.; Ee-Peng, L. Comments- Oriented Document Summarization: Understanding Documents with Readers' Feedback. *SIGIR'08*, 2008,291-298.
[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning UserInteraction Models for Predicting Web Search Result Preferences," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.
[3] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc.Int'l Conf. World Wide Web (WWW), 2001.
[9] 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.