

Ontology Based Personalized Search Filtering on Smart phone (OBPSF): A Survey

Devayani Phadke

PG Student, Department of Computer Engineering,
Sinhgad Technical Education Society's, Smt. Kashibai Navale college of Engineering
Pune, Maharashtra, India

Jyoti Nandimath

Asst. Prof. Department of Computer Engineering,
Sinhgad Technical Education Society's, Smt. Kashibai Navale college of Engineering
Pune, Maharashtra, India

Abstract - Nowadays in common people, mobile technology and internet are becoming an integral part of daily life. In mobile search the interactions between the users and search engines are limited by the small form factors of the mobile devices. In this system user is enter query on smartphone application, whatever may be the user clickthrough, gets store in .owl file by using ontology web language. This ontology and query will send to server ,at server side reranking is takes place according to ontology .By using ontology web serach personalization also takes palce. By using spyc4.5 algorithm for classification getting good results. Basically Ontology Based Personalized Search Filtering is personalize mobile search engine.

1. INTRODUCTION

Due to advances in mobile technologies and internet access users life become very comfortable and convenient to do the work very intelligently. User's enter queries, click some of the links in the results, click on ads, spend time on website pages, reconstruct their queries, and perform many actions. These interactions can serve as a significant source of information for improving web search result ranking.

In mobile search[1] the interactions between the users and search engines are limited due smartphone devices. As a result, mobile users tend to submit shorter and ambiguous queries compared to their web search counterparts. Also search engine does not gives personalized result, it gives the results globally which are same for all users.

This system adopts the meta search approach for which backend is on one of the commercial search engines, such as Google to perform the main search. The client is responsible for entering the user's requests, submitting the requests to the server, displaying the returned results, and collecting user clickthrough in order to derive user personal preferences.

- In this system[1] client server architecture is used.
- In this system unique characteristics of content and location concepts, and provides a consistent strategy using a client-server architecture to integrate them into a uniform solution for the mobile environment.
- System incorporates a user's physical locations in the personalization process. The influence of a user's GPS locations in personalization. The GPS locations help improve retrieval effectiveness for location queries (i.e., queries that retrieve lots of location information).
- The proposed system is an new approach for personalizing web search results. By mining content and location concepts for different user profiling, it utilizes both the content and location preferences to personalize search results for a user.
- This system proposes a innovative and realistic system design, this design adopts the server-client model in which user queries are forwarded to a server for processing the training and reranking quickly. System implement a working prototype of the clients on the Google Android platform, and the server on a PC to validate the proposed ideas.
- Privacy preservation is a challenging issue in this system, where users send their user profiles along with queries to the server to obtain personalized search results. System addresses the privacy issue by allowing users to control their privacy levels with two privacy parameters, minDistance and expRatio. Proposed system facilitates good ranking quality and smooth privacy preserving control.
- Proposed system show that the ontology-based user profiles can successfully capture users' content and location preferences and utilize the preferences to produce relevant results for the users. It significantly outperforms existing approaches which use either content or location preference only.

2. COMPARATIVE WORK

Clickthrough data have been used in determining the users preferences on their search results. Many existing personalized web search systems [3], evaluating user preferences of web search results is crucial for search engine development, deployment, and maintenance. A real-world study of modelling the behavior of web search users to predict web search result preferences. Accurate interpretation and modelling of user behavior has important applications to ranking, web search personalization, click spam detection and other tasks. Key insight of this work to improving robustness of interpreting implicit feedback is to model query-dependent deviations from the expected noisy user behavior. In this work shows that model of clickthrough interpretation improves prediction accuracy over state-of-the-art clickthrough methods. Generalize this approach to model user behavior beyond clickthrough, which results in higher preference guess accuracy than models based on clickthrough information alone.

In [4], Geographic web search engines allow users to constrain and order search results in an intuitive manner by focusing a query on a particular geographic area. Geographic search technology, also called local search, has recently received key interest from major search engine companies. Academic research in this part has focused primarily on techniques for extracting geographic knowledge from the web. In this paper, study of problem of efficient query processing in scalable geographic search engines. Query processing is a major bottleneck in standard web search engines, and the most important cause for the thousands of machines used by the major engines. Geographic search engine query processing is different in that it requires a combination of text and spatial data processing techniques. They propose several algorithms for efficient query processing in geographic search engines, combine them into an existing web search query processor, and estimate them on large sets of real data and query traces.

In [7], an approach to automatically optimizing the retrieval quality of search engines using clickthrough data. Naturally, a superior information retrieval system should present relevant documents high in the ranking, with less related documents following below. While previous strategies to learning retrieval functions from examples be present, they typically require training data generated from relevance decisions by experts. This makes them hard and expensive to apply. The goal of this work is to develop a method that utilizes

clickthrough data for training, namely the query-log of the search engine in connection with the log of links the users clicked on in the presented ranking. Such clickthrough data is available in large quantity and can be recorded at very low cost. Taking a Support Vector Machine (SVM) approach, this work presents a method for learning retrieval functions. From a theoretical point of view, this method is shown to be well-founded in a risk minimization framework.

Furthermore, it is shown to be feasible even for large sets of queries and features. The theoretical results are confirmed in a controlled experiment. It shows that the method can successfully adapt the retrieval function of a meta-search engine to a particular group of users, outperforming Google in terms of retrieval quality after only a couple of hundred training examples.

In [12], the paper addresses search engine personalization. They present a new approach to mining a user's preferences on the search results from clickthrough data and using the discovered preferences to adapt the search engine's ranking function for improving search quality. They develop a new preference mining technique called SpyNB, which is founded on the practical supposition that the search results clicked on by the user reset the user's preferences, but it does not draw any conclusions about the results that the user did not click on. As such, SpyNB is still applicable even if the user does not follow any order in reading the search results or does not click on all relevant results. Their extensive online experiments demonstrate that SpyNB discovers many more accurate preferences than existing algorithms do. The interactive online experimentation further confirms that SpyNB and our personalization approach are effective in practice. They also show that the efficiency of SpyNB is comparable to existing simple preference mining algorithms.

3. PROPOSED ARCHITECTURE

Providing good personalized search, proposed architecture uses Client-server model .

For providing good personalized search, proposed architecture uses Client-server model .

Figure 1. shows system architecture.

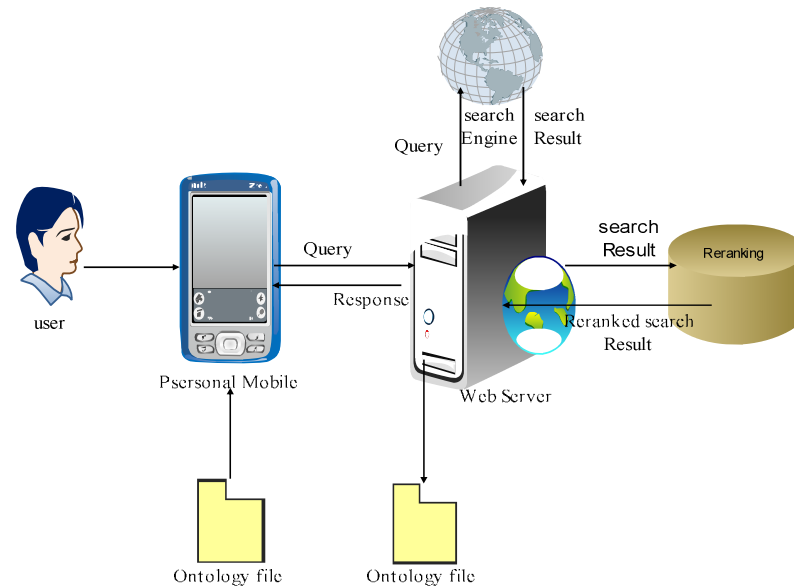


Fig.1: System Architecture

As shown in Fig. 1, proposed system architecture is providing

- 1) An application is on android smart phone where user is going to do login and enter a query.
- 2) Server will rerank the results. Backend for this server is global search engine and sends response to the application on an mobile.
- 3) RSVM is used for reranking.

In the client-server architecture, clients are storing the user clickthroughs and the ontologies derived from the server. Simple tasks, such as updating clickthroughs and ontologies, generating feature vectors, and showing reranked search results are handled by the clients with some degree of computational power. On the other hand, heavy tasks, such as RSVM training and reranking of search results, are handled by the server. Furthermore, in order to reduce the data transmission between client and server, the client would only need to submit a query together with the feature vectors to the 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 (i.e., query, feature vectors, ontologies and search results) are transmitted between client and server during the personalization process. Design addressed the issues: 1) limited computational power on mobile devices, and 2) data transmission minimization.

System consists of two major activities[1]:

1. Reranking the search results at server. When a user submits a query on the 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 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 do ontology extraction for its speed. The feature vectors from the client are then used in RSVM training to get 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 client. The ontologies returned from the 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. When the user clicks on a search result, the clickthrough data jointly with the associated content and location concepts are stored in the clickthrough database on the client. The clickthroughs are stored on the clients, so the server does not know the exact set of documents that the user has clicked on. This design allows user privacy to be preserved to some extent. Two privacy parameters, *minDistance* and *expRatio*, are

proposed to control the amount of personal preferences exposed to the server. If the user is concerned with 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 server for the personalization. On the other hand, if a user wants more accurate results according to preferences, the privacy level can be set to low so that the server can use the full feature vectors to maximize the personalization effect.

The proposed new approach to mining a user's preferences on the search results from clickthrough data and using the discovered preferences to adapt the search engine's ranking function for improving search quality. This system develop a new preference mining technique called SpyC4.5, which is based on the practical assumption that the search results clicked on by the user reset the user's preferences, but it does not draw any conclusions about the results that the user did not click on. As such, SpyC4.5 is still valid even if the user does not follow any order in reading the search results or does not click on all relevant results. The SpyC4.5 discovers many more accurate preferences than existing algorithms do. The SpyC4.5 personalization approach are effective in practice. They also show that the efficiency of SpyC4.5 is comparable to existing simple preference mining algorithms.

4. CONCLUSION

This Paper proposes system architecture, profile the users interests and personalize the search results according to the users profiles. The other global search engines are not giving the personalised result. For all the search result is same. System represents different types of concept in different ontologies to include context information revealed by user mobility system also takes into account the visited physical location of users. Main computation task is distributed to the server so that it gives effective performance SpyC4.5 classifier gives better performance.

5. REFERENCES

- [1] Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee, PMSE: A Personalized Mobile Search Engine, IEEE Trans. On Knowledge and Data Engineering, Vol. 25 No.4, April 2013
- [2] E. Agichtein, E. Brill, and S. Dumais, Improving Web Search Ranking by Incorporating User Behavior Information, Proc. 29th Ann. Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.
- [3] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, Learning User Interaction Models for Predicting Web Search Result Preferences, Proc. Ann. Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.
- [4] Y.-Y. Chen, T. Suel, and A. Markowetz, Efficient Query Processing in Geographic Web Search Engines, Proc. Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006
- [5] 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.
- [6] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, Analysis of Geographic Queries in a Search Engine Log, Proc. First Intl Workshop Location and the Web (LocWeb), 2008.
- [7] T. Joachims, Optimizing Search Engines Using Clickthrough Data, Proc. ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining, 2002.
- [8] K.W.-T. Leung, D.L. Lee, and W.-C. Lee, Personalized Web Search with Location Preferences, Proc. IEEE Intl Conf. Data Mining (ICDE), 2010.
- [9] 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.
- [10] H. Li, Z. Li, W.-C. Lee, and D.L. Lee, A Probabilistic Topic-Based Ranking Framework for Location-Sensitive Domain Information Retrieval, Proc. Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2009.
- [11] B. Liu, W.S. Lee, P.S. Yu, and X. Li, Partially supervised Classification of Text Documents, Proc. Intl Conf. Machine Learning (ICML), 2002.
- [12] 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.
- [13] 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.
- [14] C.E. Shannon, Prediction and Entropy of Printed English, Bell Systems Technical J., vol. 30, pp. 50-64, 1951.
- [15] Q. Tan, X. Chai, W. Ng, and D. Lee, Applying Co-Training to Clickthrough Data for Search Engine Adaptation, Proc. Intl Conf. Database Systems for Advanced Applications (DASFAA), 2004.