International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Review On: Advanced and Robust Personalized Mobile Search Engine

Yogesh B. Jadhao[1], Shyam S. Gupta[2]

[1]Department of Computer Engineering, Siddhant College of Engineering, Sudumbare, Pune, Maharashtra, India 412109

[2]Professor, Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India 412109

Abstract: ***Internet is a fastest growing technology. On every second large amount of information is searched by different people. That's why the internet becomes the great source of information. But this information consists of many redundant and ambiguous data which is not useful for user. As the world has been digitized and all the search now takes place through handy devices and mobiles, so the query typing process of users has reduced and users try to minimize typing long queries and thereby increase search engine ambiguity. Sothe search results showing again and again same irrelevant results which leads to headache to the user. Many search engines give unnecessary results of user query which are not expected by the user. There is lots of research done to overcome this problem but exact solution is being difficult to find as the technology is continuously growing. To build a better solution on this problem we proposed a mobile search engine which will be useful to find the exact match of user query. Here we are using the click through log data given by user to match his exact expectation. We are also using the location based searching for retrieving a better search results.***

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

1. Introduction

1.1 Problem Definition

Now a day everyone is using the mobile phone for searching the information on the internet. Mobile search engine (MSE) accepts the user query and returns the relevant result to the user. People starts to use mobile phone for their official, educational, family or other normal life purposes. Now everyone wants all the information in the hand. The mobile phone has ability to provide the information to the user at any place and at any time through the internet.

The major problem arises when a user wants to search a particular query on mobile search engine. But it is giving ambiguous results and showing the long list of results which are not useful to the user. Then user suffered to find his exact relevant result among those all results. Many times user relevant results are not available then also the list of result is shown by the search engine. This problem may cause headache to the user.

The interaction between user and the mobile search engine is limited by the small factor of the mobile devices. To return the exact or user query relevant results the user profile must be take in consideration. Profiling of user interest and personalized data will help to produce the highly relevant search result.

Along with user profile we are going to use the location of the user to find relevancy in the search result. We also take the visited physical locations of users in the MSE. And this information can be easily obtained by using the mobile phones GPS devices, so it is referred to as GPS locations and it has an important role in mobile web search.

For example if the user is searching the information from "Africa". And he enters the query "lion" in the mobile search engine. Then according to his location the search engine will give the information about African lion. Because user is right now in Africa and the information of other regions lion is not useful for him. Search engine will also consider the user profile while returning the results.

2. Literature Survey

Many existing personalized web search systems are based click through data and clustering to determine users' preferences.

2.1 Concept Based Clustering

This approach consists of the following four major steps. First, when a user submits a query, concepts (i.e., important terms or phrases in web-snippets) and their relations are mined online from web-snippets to build a concept relationship graph. Second, clickthrough are collected to predict user's conceptual preferences. Third, the concept relationship graph together with the user's conceptual preferences is used as input to a concept-based clustering algorithm that finds conceptually close queries.

The underlying idea of this technique is based on concepts and their relations extracted from the submitted user queries, the web-snippets,1 and the clickthrough data. Clickthrough data was exploited in the personalized clustering process to identify user preferences: A user clicks on a search result mainly because the web snippet contains a relevant topic that the user is interested in. Moreover, clickthrough data can be collected easily without imposing extra burden on users, thus providing a low-cost means to capture user's interest. Concept based clustering goes in depth for finding out the synonyms or actual meaning of the query typed. Moreover many words have more than one meanings, so the concept clustering tries to find out the actual meaning of the entered query and thereby optimize the results shown to the user. The cluster in takes place based on

Figure 1: Architecture of concept based clustering

2.2 Kokono Search

It explains the mechanism of location-based search engine. This system consists of three components. The robot gathers web documents from the Internet. After a document was gathered, the robot prioritizes URLs that were included in it. For example, the priority is high when a link label contains location information (ex. Address etc.) Then the robot gathers web documents that have high priority. A parser extracts location information from web documents and converts them into latitude-longitude pairs or polygons. The retrieval module converts location information specified by user to a latitude-longitude pair and it creates a search circle whose canter is this pair. By judging overlaps of this circle and the latitude-longitude pairs or polygons, it picks up documents that are written about locations within this circle. The engine returns URLs of the documents as results of the search. The module calculates the radius of the circle automatically that the overlaps contain appropriate number of results.

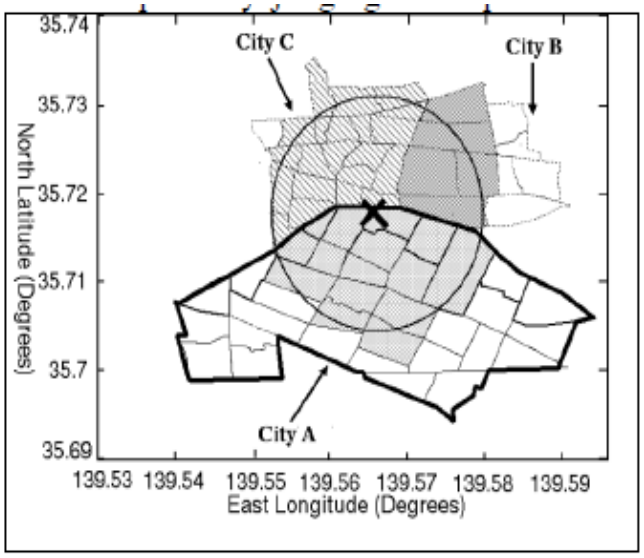

Figure 2: Kokono search.

2.3 Privacy Search

Personal data, i.e. personal documents, browsing history and emails might be helpful to identify a user's implicit intents. However, users have concerns about how their personal information is used. Privacy, as opposed to security or confidentiality, highly depends on the person involved and how that person may benefit from sharing personal information. The question here is whether a solution can be found where users themselves are able to set their own privacy levels for user profiles to improve the search quality. An algorithm is provided for the user to automatically build a hierarchical user profile that represents the user's implicit personal interests.

General interests are put on a higher level; specific interests are put on a lower level. Only portions of the user profile will be exposed to the search engine in accordance with a user's own privacy settings. A search engine wrapper is developed on the server side to incorporate a partial user profile with the results returned from a search engine. Rankings from both partial user profiles and search engine results are combined. The customized results are delivered to the user by the wrapper.

The solution has three parts: First, a scalable algorithm automatically builds a hierarchical user profile from available source data. Then, privacy parameters are offered to the user to determine the content and amount of personal information that will be revealed. Third, a search engine wrapper personalizes the search results with the help of the partial user profile.

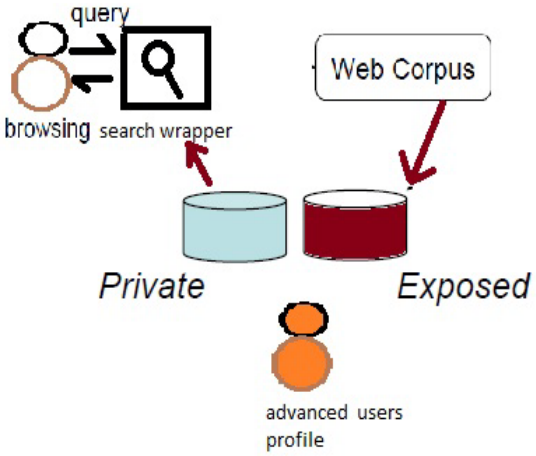

Figure 3: Privacy Search

2.4 Partially Supervised Classification

We investigate the following problem: Given a set of documents of a particular topic or class P, and a large set of mixed documents that contains documents from class P and other types of documents, identify the documents from class P in M The key feature of this problem is that there is no labelled non- P document, which makes traditional machine learning techniques inapplicable, as they all need labelled documents of both classes. We call this problem partially supervised classification.

The key feature of these problems is that there is no labelled negative document, which makes traditional classification methods inapplicable, as they all need labelled documents of every class. We call this problem partially supervised classification, as there are only positive documents (which can be considered as labelled with the class positive), but not labelled negative documents.

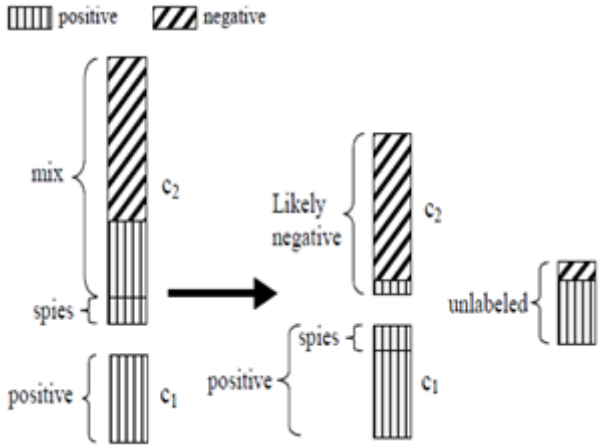

Figure 4: Partially supervise classification

3. Proposed System

To overcome all the limitations of the existing system we have proposed the new Personalized Mobile Search Engine (PMSE). Our approach creates a user profile based on the ontology and produces efficient search results. We are considering the clickthrough data along with the user profile and the user location. The location of the can be capture by using the mobile GPS locator. The user profile is created on the basis of the user interest and user clickthrough data. Finally by using the combination of all three strategies we can able to produce highly relevant search result.

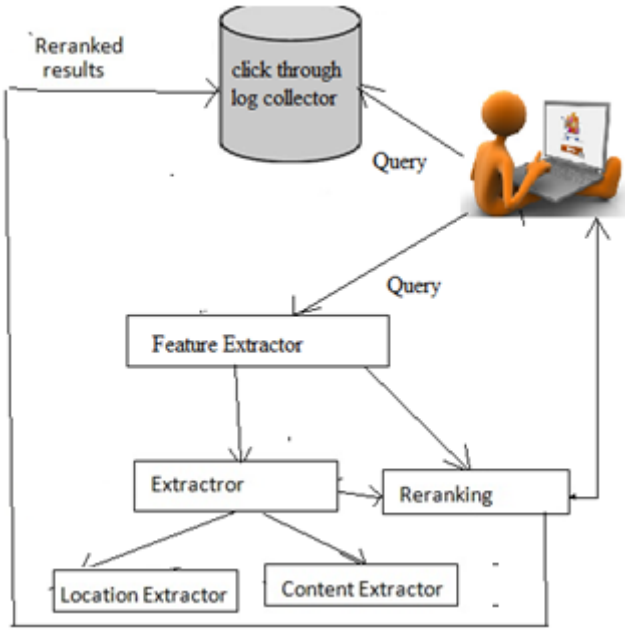

Figure 5: System flow

Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided.

In the above architecture, a client stores the user clickthroughs and the ontologies derived from the server. Other simple tasks are handled by the Extractor such as updating clickthoughs and ontologies, creating feature extractor, and obtaining reranked search results. And the other tasks are handle by the re-ranker, which includes RSVM training and reranking of search results., For reducing the data transmission, the user would only need to submit a query along with the feature extractor to the server, and the server will return a set of reranked search results according to the user profile stated in the feature extractor.

3.1 User Profiling

User Profile is generated on the basis of user interest. This is classified into two type's content profile and the location profile. The content profile is modeled on the basis of the ontology. The content profile is generated by the clickthrough collector and feature extractor. Second location profile which is generated on the basis of the user location. The user location can be found by using the GPS locator which is inbuilt inside the mobile device. The ontologies is nothing but a possible concept space arising from a queries entered by the user, which are collected along with the clickthrough data for future location adaptation.

We observe some issues in location profile formulation. First, an extractor usually considers only a few locations, and thus very few of them will match with the query terms in search box. For avoiding this problem, we extract location profile from the full data retrieved by engine. Second, the similarity relationship cannot be efficiently derived because the limited numbers of location concepts are stored in retrieved documents.

4. Conclusions

Thus we comes to conclude that our proposed system which is based on clickthrough data, user profile and user location will produce highly relevant results to the user search query. Here we have used the concept of ontology to make our searching more efficient. The Location based ontology will help to generate more accurate results from the bunch of ambiguous results.

References

[1] Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang Chien Lee "PMSE: A Personalized Mobile Search Engine" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013.
[2] Nat'l geospatial, http://earth-info.nga.mil/, 2012.
[3] svmlight, http://svmlight.joachims.org/, 2012.
[4] World gazetteer, http://www.world-gazetteer.com/, 2012.
[5] E. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," Proc. 29th Ann. Int'l ACM SIGIR Conf.

Research and Development in Information Retrieval (SIGIR), 2006.
[6] 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.
[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] 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.
[9] 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.
[10] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] 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.
[16] 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.
[17] C.E. Shannon, "Prediction and Entropy of Printed English," Bell Systems Technical J., vol. 30, pp. 50-64, 1951.
[18] 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.
[19] 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.
[20] E. Voorhees and D. Harman, TREC Experiment and Evaluation in Information Retrieval. MIT Press, 2005.
[21] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. Int'l Conf. World Wide Web (WWW), 2007.
[22] S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc. Int'l Conf. World Wide Web (WWW), 2001.

Author Profile

Yogesh Jadhao received B.E degrees in Computer Science And Engineering from Amravati University in 2013, respectively. During 2013-2014, he stayed in M.E (Computer Engineering) from Savitribai Phule Pune University, Maharashtra, India.

Prof. Shyam S. Gupta is Assistant lecture in Siddhanth College of Engineering, Sudumbare, Pune, Maharashtra, India.