International Advance Journal of Engineering Research (IAJER)
Volume 2, Issue 7 (July- 2019), PP 01-04
ISSN: 2360-819X
www.iajer.com

Research Paper **Open Access**

Use of Personalisation in Mobile Data Mining

R. R. Shelke
H.V.P.M. COET, Amravati.

ABSTRACT:- In today's world internet technology growing rapidly and people used internet on their mobile phones, desktops or laptops etc. for various purposes. Various search engines are available for extracting required information. The goal of mobile data mining is to provide advanced techniques for searching required data from mobile devices. Mobile catches user location and gives information associated to that location. Location is one of the important factors in mobile data mining and the information related to location categorize as content and location concepts. The user preferences are prepared by ontology based, wide-ranging user profiles, which are used to adapt a personalized ranking function for rank adaptation for search results.

Keywords:- Data mining, personalisation ,mobile data mining, ranking.

I. INTRODUCTION

The 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 users' profiles. The small form factors of the mobile device limits the interaction between the mobile user and the search engines. In order to get a highly relevant result, user profiling is recommended. User profiling is nothing but the mobile data mining capturing the user interest to personalize. Mobile data mining 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, miner classifies these concepts into content concepts and location concepts. In addition, users" locations (positioned by GPS) are used to supplement the location concepts. The user preferences are organized in an ontology-based, multi-facet 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 relevance to the users need, four entropies are introduced to balance the weights between the content and location facets.

II. LITERATURE REVIEW

Many researchers worked on data mining techniques[1-2]. Some researchers developed data mining technique for mobile devices[3-4]. Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee [6] suggest a practical approach for Personalized Mobile Search Engine by assuming the metasearch method which is capable to reply on the commercial search engines, accomplish a indisputable search. In personalized mobile search engine process the user request are handled by the client who is submitted to the personalized mobile search engine server. Server replies with the results and users clickthrough data for obtaining the user's unique preferences. The server of personalized mobile search engine is trustworthy for management of the heavy tasks .Server preparing and reranking the search results according to the users content and location preferences sooner than they come back to client side. The personalized mobile search engine clients keep records of the specific user's profile and maintain user's privacy. For validating the proposed scheme the personalized mobile search engine client prototyped with android and server on personal computer [7]. For properly distinguishing variety of the concepts linked with query and their relevancy of the user the idea of content and location randomness's for computing the amount of information related to location and content linked with query as well as interest of the user in content or location information.

The initial ranking techniques PageRank [5] and Hypertext Induced Topic Search (HITS) belongs to the category of web structure mining. These algorithms have undergone several modifications and are still in use today by the standard search engines. Link structure and popularity are the basis of ranking for algorithms in this category. Based on the fact the more incoming links a document has the more popular and thus the more relevant it is. Such documents are ranked at higher ranks irrespective of their relevancy with the users' search

goal. Web content mining category includes ranking techniques proposed in [5]. In addition to the link structure of the documents these strategies also rely upon the similarity between the submitted query and the content of available on the web document. It proves to be more efficient in satisfying the users' information needs when unambiguous query is explicitly submitted by the user. As it solely relies on similarity with respect to the query, it fails to address users' search goals for ambiguous queries.

Pooja Devi, Ashlesha Gupta, Ashutosh Dixit Compared HITS and PageRank Link based Ranking Algorithms. On the basis of this study we conclude that both page rank and HITS algorithm are different link analysis algorithms that employ different models to calculate web page rank. The PageRank and HITS algorithm give importance to links rather than the content of the pages. According to PageRank algorithm, rank score of a web page is divided evenly over the pages to which it links whereas HITS algorithm rank pages according to authority and hubness of a page. Page Rank is a more popular algorithm used as the basis for the very popular Google search engine. This popularity is due to the features like efficiency, feasibility, less query time cost, less susceptibility to localized links etc. which are absent in HITS algorithm. However though the HITS algorithm itself has not been very popular, different extensions of the same have been employed in a number of different web sites. Results demonstrate that HITS calculates authority nodes and hubness correctly. HITS may also be combined with other information retrieval based rankings. After going through exhaustive analysis of PageRank and HITS algorithms for ranking of web pages against the various parameters such as methodology, input parameters, relevancy of results and importance of the outcome, it is concluded that these techniques have limitations particularly in terms of time response, accuracy of results, importance of the outcome and relevancy of results. An efficient web page ranking algorithm should meet out these challenges efficiently with compatibility with global principles of web technology.

Efficient Generalized Forensics Framework for Mobile Devices has been studied by Rizwan Ahmed and Dr. Rajiv V. Dharaskar [8]. According to them ,Mobile devices are becoming more and more similar to desktop computers as smartphones continue to gain more capabilities. Mobile devices normally stay connected online all the time because of their default characteristics and user behavior. As a consequence of the integration of mobile networks into the Internet, security threats on one network will affect the other network. This makes smartphones the attractive targets to hackers. We need to be well prepared for the promise attack, we, as defenders, should study mobile botnets attacking techniques.

III. SYSTEM DESIGN

Mobile data mining represents a very promising area for users and professionals that need to analyze data where users, resource and applications are mobile. The combined use of a data mining approach with mobile programming technologies could be used for the implementation of mobile knowledge discovery applications. An architecture for mobile data mining is given in fig 1

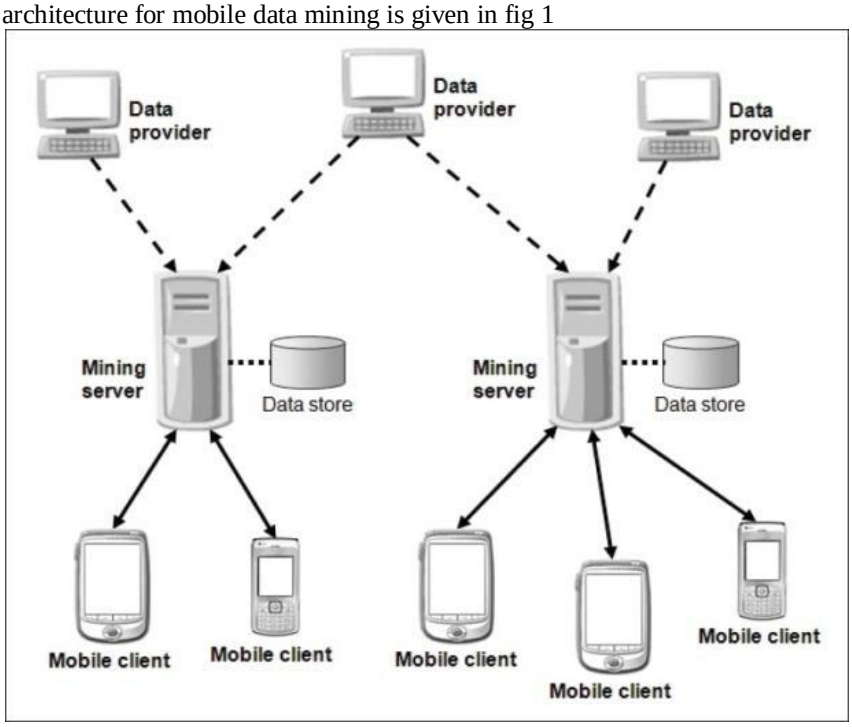

Fig 1. Architecture for mobile data mining.

The architecture used in system includes four types of components s follows

- Data providers: the applications that generate the data to be mined.
- Mobile clients: the applications that require the execution of data mining computations on remote data.
- Mining servers: server nodes used for storing the data generated by data providers and for executing the data mining tasks submitted by mobile clients. Data generated by data providers is collected by a set of mining servers that store it in a local data store. Depending on the application requirements, data coming from a given provider could be stored in more than one mining server. The main role of mining servers is allowing mobile clients to perform data mining on remote data by using a set of data mining algorithms. Once connected to a given server, the mobile client allows a user to select the remote data to be analyzed and the algorithm to be run. When the data mining task has been completed on the mining server, the results of the computation would be visualized on the user device either in textual or visual form.

This client-server architecture is also personalised mobile data mining, which meets three important requirements. First, computation-intensive tasks, such as reranking , should be handled by the mining 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 clients in order to preserve user privacy. In the mobile data mining (MDM) client-server architecture, clients are responsible for storing the user clickthroughs and the ontologies derived from the MDM server. Simple tasks, such as updating clickthoughs and ontologies, creating feature vectors, and displaying reranked search results are handled by the MDM clients with limited computational power. On the other hand, heavy tasks, such as reranking of search results, are handled by the MDM server. Moreover, in order to minimize the data transmission between client and server, the MDM client would only need to submit a query together with the feature vectors to the MDM 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.

The two major activities handled by systems are

1. Reranking the search results at mining 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 . 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 (The Reduced Support Vector Machine generating a non-linear separating surface for a large dataset which requires a small portion of the dataset for its characterization)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, there ranked 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. 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 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 in certain degree. If the user is concerned with his/her own privacy, the privacy level can be set so that only limited personal information will be included in the feature vectors and passed along to the server for the personalization.

IV. CONCLUSION

Data mining help users find useful information from huge amount of data. Therefore it is useful in search engine. When the same query is submitted by different users, typical search engines return the same result regardless of who submitted the query. A search engine infrastructure must be able to provide the same quality of service to all queries received during a day. To receive personalized web services, the user has to provide personal information and preferences, in addition to the query itself, to the web service. However, detailed personal information could identify the sender of sensitive queries, thus compromise user privacy. Long-term search history contains rich information about a user's search preferences, which can be used as search context to improve retrieval performance. The user profiles for particular users are stored on the clients, thus preserving privacy to the users. The design adopts the server-client model in which user queries are forwarded to a server for processing the training and re-ranking quickly.

REFERENCES

[1]. K. Meena, M. Durairaj and K.R. Subramanian, "Data Mining Algorithm Used in a Mobile Computing Environment" ,College Science in India ,ISSN 1939-2648 3 : 1 February 2009

[2]. Bharati M. Ramageri , " Data Mining Techniques And Applications " , Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305,ISSN : 0976-5166 , 2010.

[3]. Domenico Talia*y* and Paolo Trunfioz , " Mobile Data Mining on Small Devices." http://grid.deis.unical.it/papers/pdf/MI2010.pdf Viewed on 25-03-2011

[4]. R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare ," DATA MINING FOR MOBILE DEVICES USING WEB SERVICES ", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622.

[5]. Pooja Devi, Ashlesha Gupta, Ashutosh Dixit ," Comparative Study of HITS and PageRank Link based Ranking Algorithms", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014.

[6]. 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

[7]. Tushar Mohod, Prof. S. N. Gujar, "A Survey on a Personalized Mobile Search Engine", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 12, December 2014, pp 7347-7352.

[8]. Rizwan Ahmed, R. V. Dharaskar, "Study of Mobile Botnets: An Analysis from the Perspective of Efficient Generalized Forensics Framework for Mobile Devices", National Conference on Innovative Paradigms in Engineering & Technology (NCIPET-2012), Proceedings published by International Journal of Computer Applications (IJCA), pp. 5-8

[9]. V. Shanmuga Priya, S. Sakthivel, "An Implementation Of Web Personalization Using Web Mining Techniques", International Journal of Computer Science and Mobile Computing ISSN 2320–088X, *IJCSMC, Vol. 2, Issue. 6, June 2013, pg.145 – 150.*