ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

Personalized Mobile Search Engine With Multiple Preference

D.Sudha[1], S.Vijayalakshmi[2], V.Komathi[3], K.Amala[4]

Information Technology, Christ College of Engineering and Technology, Puducherry[1,2,3,4]

Abstract: Data mining is a system employing for more computer learning technique to automatically analyze and extracting knowledge from data stored in the database. The goal of data mining is to extract hidden predictive information from database. This paper make use of data mining concept for collecting user's multiple preference from clickthrough data. The collecting user preference is based on the content and the location concepts. In existing system, RSVM algorithm do not perform re-ranking for multiple preference. To overcome this drawback, the proposed work is based on PRRA algorithm. This algorithm is used to find the shortest path which help us to get a better result. PMSE concentrate more about privacy which based on user as well as location by leveraging the amount of content. To characterize the diversity of the concepts associated with a query and their relevance's to the user's need, four entropies are introduced to balance the weights between the content and location facets [11].

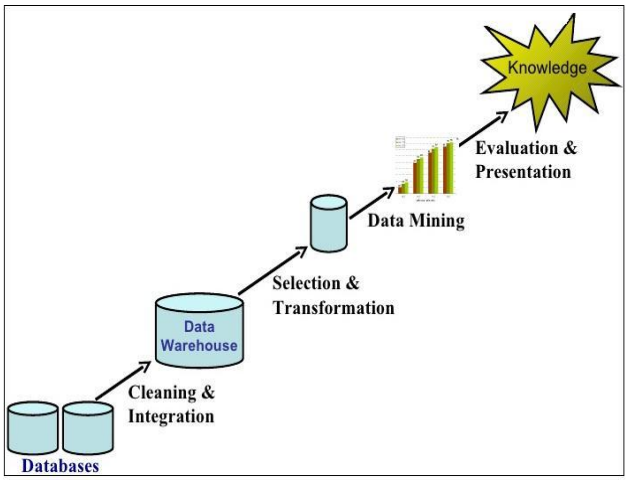
Index Terms- PMSE, Click through, Multiple preference, Search engine.

I. INTRODUCTION

A. Data Mining

Data Mining, also known as Knowledge-Discovery in Databases (KDD), is the process of automatically searching large volumes of data for patterns [19].

Data Mining applies many older computational techniques from statistics, machine learning and pattern recognition. Steps involved in KDD process:



Knowledge discovery as a process consists of an iterative sequence of the following steps:

1. Data cleaning (to remove noise or irrelevant data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)[12].
5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns) [13].
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures) [13].

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns [20]. Some of the data mining task are classification, regression, clustering, summarization, dependency modeling and deviation detection.

The five major elements in data mining are Extract, transform, and load transaction data onto the data warehouse system. Store and manage the data in a multidimensional database system. Provide data access to business analysts and information technology professionals. Analyze the data by application software. Present the data in a useful format, such as a graph or table [14].

B. Techniques

There are many techniques of data mining. The most common techniques used in the field of data mining are followings.

Artificial neural networks - Non-linear predictive models that learn through training and resemble biological neural networks in structure. This predictive model uses neural networks and finds the patterns from large databases.

Decision tree - Set of decisions are represented by Tree-shaped structures. These decisions generate rules for the classification of a dataset under the large databases.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Genetic algorithms - Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

Nearest neighbor method – [21] A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). This is sometimes called the k-nearest neighbor technique.

Rule induction - The extraction of useful if-then rules from data based on statistical significance between different records of database. Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data [21].

C. Scope

Automated prediction of trends and behaviors: Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

Automated discovery of previously unknown patterns: Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors [15][21].

D. Applications

- Fraud detection is used to suspect with marked differences between current usage and user history.
- Health care uses data mining to help uncover effective drug treatment for certain types of patients.
- Business and Finance it identify the usual features of customers who buy the same product from the company.
- DNA Analysis it provide similarity search and comparison among DNA sequences.
- Sports and Gaming to determine how much an individual customer should be spending while visiting their favorite casino

II. ONTOLOGY

The subject of ontology is the study of the categories of things that exist or may exist in some domain. [16] The product of such a study, called ontology, which is a catalog of the types of things that are assumed to exist in a domain of interest from the perspective of a person who uses a language for the purpose of talking about domain.
The relation between Ontologies and data mining in two manners [17][22]:

From ontologies to data mining, we are incorporating knowledge in the process through the use of ontologies, i.e. how the experts comprehend and carry out the analysis tasks. Representative applications are intelligent assistants for discover process, interpretation and validation of mined knowledge, Ontologies for resource and service description and knowledge Grids.

From data mining to Ontologies, we include domain knowledge in the input information or use the ontologies to represent the results. Therefore the analysis is done over these ontologies. [22]The most characteristic applications are in medicine, biology and spatial data, such as gene representation, taxonomies, applications in geosciences, medical applications and specially in evolving domains.
When we can represent and include knowledge in the process through ontologies, we can transform data mining into knowledge mining.

Domain Ontology – [22] The models on many scientists work to represent their work hypotheses are generally cause effect diagrams. Models make use of general laws or theories to predict or explain behavior in specific situations. Currently these cause effect diagrams can be without difficulty translated to ontologies, by means of conceptual maps which discriminate taxonomy organized as central concepts, main concept, secondary concepts, specific concepts.

Metadata Ontologies - As Spyns et al. affirm ontologies in current computer science language are computer-based resources that represent agreed domain semantics. Unlike data models, the fundamental asset of ontologies is their relative independence of particular applications, i.e., an ontology consists of relatively generic knowledge that can be reused by different kinds of applications/tasks.

Axiomatized ontology - distinguishes subtypes by axioms and definitions stated in a formal language, such as logic or some computer-oriented notation that can be translated to logic [16].

Prototype-based ontology - distinguishes subtypes by a comparison with a typical member or prototype for each subtype. Large ontologies often use a mixture of

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

definitional methods: formal axioms and definitions are used for the terms in mathematics, physics, and engineering; and prototypes are used for plants, animals, and common household items.

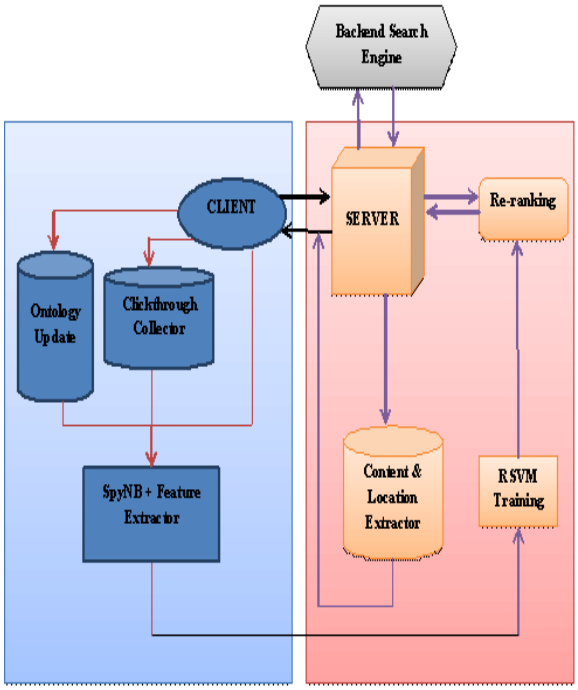
Advantage of an explicit ontology of data-mining techniques, which defines the various techniques and their properties. Main characteristics are systematic enumeration of valid DM processes, so they do not miss important, potentially fruitful options [17]. Effective rankings of these valid processes by different criteria, to help them choose between the options. An infrastructure for sharing data mining knowledge, which leads to what economists call network externalities.

III. RELATED WORKS

Personalized Mobile search engine have been used to provide search result according to the priority of the user preference. Some of the existing personalized web search systems are based on the clickthrough data to determine users' preferences one among them [5] where Joachims proposed to mine preferences from clickthrough data. Leung et al, [6] introduced an efficient approach to determine users' conceptual preferences from clickthrough data. Search engines [2] can often return better results to users by analyzing features such as user location or geographic terms in web pages and user queries. [10] proposed a two-step strategy to improve retrieval effectiveness. In the first step, the system automatically deduces, for each user, a small set of categories for each query submitted by the user, based on his/her search history. In the second step, the system uses the set of categories to augment the query to conduct the web search. [5] presented a Support Vector Machine (SVM) algorithm that leads to a convex program and that can be extended to non-linear ranking functions. Experiments show that the method can successfully learn a highly effective retrieval function for a meta-search engine.

IV. EXISTING SYSTEM

In existing 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, and to perform both feature extraction and RSVM training on the PMSE server to train the weight vectors for reranking. However, if all clickthroughs are exposed to the PMSE server, the server would know exactly what the user has clicked. To address privacy issues, clickthroughs are stored on the PMSE client, and the user could adjust the privacy parameters to control the amount of personal information to be included in the feature vectors, which are forwarded to the PMSE server for RSVM training to adapt personalized ranking functions for content and location preferences.the existing work they use Computer forensic analysis.



Block diagram of existing system

Existing framework is capable of combining a user's GPS locations and location preferences into the personalization process for the best of our knowledge. [18] A personalization framework that utilizes a user's content preferences and location preferences as well as the GPS locations in personalizing search results. a realistic design for PMSE by adopting the metasearch approach which replies on one of the commercial search engines, such as Google, Yahoo, or Bing, to perform an actual search using Re-ranking Support Vector Machine(RSVM). RSVMs critical disadvantage is unavoidable because the number of data pairs is quadratic to the number of data objects. Double order complexity problem discussed above happen in the RSVM design search engine. The Naive Bayes algorithm Scoring can be parallelized irrespective of the algorithm for efficient search, but the Naive Bayes cannot be used binary and multiclass classification problems.

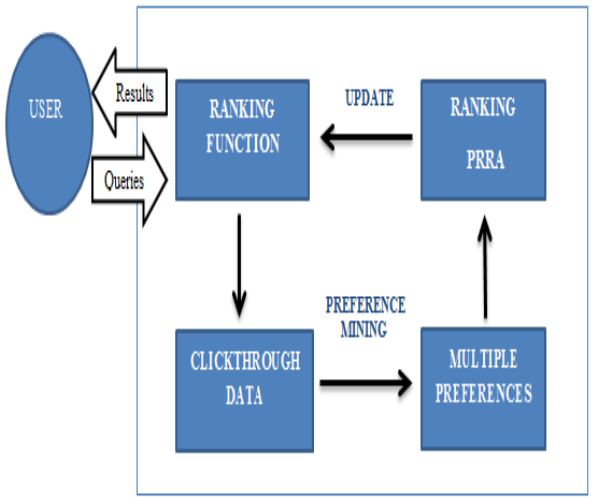
V. PROPOSED WORK

Page Rank Reviser Algorithm to find the shortest path this will help us to get a better result. A dedicated running I Spy ensures that user interaction or other software won't interrupt the processing tasks. Preference level is increased by twice for the accuracy of the result we are using phonetic. Page Rank Reviser Algorithm Reduce the Max Redraw Rate in settings to save CPU on resizing the live feeds. The two algorithms used for simultaneous access of multiple personalized Search. PMSE concentrate more about privacy. To concentrate the privacy based on user as well as location by leveraging the amount of content [11]. To characterize the diversity of the concepts associated with a query and their

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

relevance's to the user's need, four entropies are introduced to balance the weights between the content and location facets.



Block diagram of proposed system

MODULE 1: MOBILE CLIENT

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 clickthroughs and ontologies, creating feature vectors, and displaying re ranked search results are handled by the PMSE clients with limited computational power. 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 page Rank Reviser Algorithm 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.

MODULE 2: PMSE SERVER

Heavy tasks, such as Page Rank Reviser Algorithm training and re-ranking of search results, are handled by the PMSE server.

PMSE Server's design addressed the issues: Limited computational power on mobile devices and Data transmission minimization.

PMSE consists of two major activities: Re-ranking the search results at the PMSE server and Ontology update & clickthrough collection at a mobile client

MODULE 3: PAGE RANK REVISER ALGORITHM SEARCH RESULTS

Multiple search can be done at the page ranking algorithm. 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 backend search engine. The content and location concepts are separated 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 page Rank Reviser Algorithm is used to find a weight vector of the content and the location concepts . It represents the user interests based on the user's content and location preferences for the Page-ranking. Again, the training process is performed on the server for its speed. The search results are then re-ranked according to the weight vectors obtained from the PRRA training. Finally, the re-ranked results and the extracted ontologies for the personalization of future queries are returned to the client.

MODULE 4: CLICK THROUGH COLLECTION

PMSE server contains 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 selects on a result, the clickthrough data together with the combination of content and location concepts are saved in the clickthrough database on the client. This script is a ellaborate guide to how ISpy works, what it can do and how you can get the most out of it. New software can seem scaring off at first so we have to put a lot of cause in order to cover the most mutual scenarios associated with comprehensive coverage of all feature of the software. ISpy is capable of much more efficient than just motion detection. 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.

VI. IMPLEMENTATION AND RESULT

The user queries are stored as a clickthrough data collection in the client database. Using the clickthrough database user preference can be extracted through I-SPY technique. This preference can be analyzed with the result of backend search engine and provided re-ranked search results using PRRA algorithm. The Page Rank Reviser Algorithm(PRRA) is used to measure the counting of websites used in their user search results. Thus the PMSE will provide efficient search results by supporting the multiple preference of the particular user. PMSE maintaining good ranking quality and the data transmission between the user and the search engine should ensure fast and efficient processing of the search.

CONCLUSION

In existing system it provides result based on only a particular preference and also it need more time to produce the user preference. In this paper we have used the PRRA algorithm for re-ranking the backend search and ISpy technique to provide a multiple preference in order to

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

attain the efficient search results. Thus we illustrate that our method efficiently recognizes the user queries from the search engine.

REFERENCES

[1] 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.
[2] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis ofGeographic Queries in a Search Engine Log," Proc. First Int'lWorkshop Location and the Web (LocWeb), 2008.
[3] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. Int'l ACMSIGIR Conf. Research and Development in Information Retrieval(SIGIR), 2006.
[4] E. Agichtein , E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," Proc. 29thAnn. Int'l ACM SIGIR Conf. Research and Development in InformationRetrieval (SIGIR), 2006.
[5] T. Joachims, "Optimizing Search Engines Using ClickthroughData," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and DataMining, 2002.
[6] 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.
[7] 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.
[8] 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.
[9] S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc. Int'l Conf. World Wide Web (WWW), 2001.
[10] Fang Liu, Clement Yu, and Weiyi Meng "Personalized Web Search for Improving Retrieval Effectiveness" IEEE Trans. Knowledge and Data Eng., Vol. 16, No. 1, January 2004
[11] http://www.chennaisunday.com/IEEE%202013%20Java/ PMSE%20A%20Personalized%20Mobile%20Search%20Engine.doc
[12] http:/ /neuron.csie.ntust.edu.tw/homework/ 99/ NN/ homework3/ M9915908-hw3-1/ Approaches.htm
[13] http://www.indiastudychannel.com/resources/123676-Data-mining.aspx
[14] http://www.anderson.ucla.edu/faculty/jason.frand/teacher/ technologies/palace/datamining.htm
[15] http://www.thearling.com/text/dmwhite/dmwhite.htm
[16] http://www.jfsowa.com/ontology/
[17] http://www.irma-international.org/book/data-mining-ontologies/ 234/
[18] http://www.kresttechnology.com/krest-academicprojects/krest-mtech projects/ CSE/ abstracts/ data%20mining/ 21.doc
[19] http://en.wikipedia.org/wiki/Knowledge_discovery
[20] http://en.wikipedia.org/wiki/Data_mining
[21] http:/ /faisalsikder.wordpress.com/ tag/ scope-of-data-mining/
[22] http://www.igi-global.com/ book/ data-mining-ontologies/234