International Journal of Advanced Engineering & Innovative Technology,
Volume 1, Issue 1, April-2014, 24-29 ISSN: 2348-7208

Personalization of Search Engines for Mobiles

Reshma Sawant, Akshaya Deshmane, Shweta Sawant

B.E Computer Engineering, Navsahyadri Education Society's Group of Institutions, Pune.

For correspondence: reshma.sawant.30@gmail.com, akshayadeshmane.777@gmail.com, shwetasawant12@gmail.com

Abstract- The interaction between user and search engine is the major problem in mobile search. So Personalized mobile search engine is used to capture user preferences. This user preferences are on the basis of two concepts that is content concept and location concept through users clickthrough data. Ontologies are maintain at client and server side for location and content concept. GPS is used to detect user's location on the basis of users' click through data to find out location concept.In this paper we have presented PMSE client-server architecture. PMSE server carry out heavy tasks such as reranking, concept extraction and RSVM training. Clients are responsible for storing the click though data and forward the request to the PMSE server.

Index Terms— Ontology, content concept, location concept, training, reranking, click through, concept extraction

1. Introduction

The interaction between user and search engine is the major problem in mobile search engine. When we search, we get results that is concepts of same types. User profile is also not maintain which maintain privacy at user side and helps to return relevant results and personalize the search engine.In existing system it is very difficult to detect location as well as to maintain privacy level. So observing the need for different type of concepts PMSE classifies concepts into content concept and location concept in different ontologies[1].GPS is used to detect the location of user by considering users' physical location. DBPedia Mobile is one such location aware application for the iPhone that uses the GPS location of the client to display a map of user's location [2].For eg. If user is in Pune, then his/her physical location can be used for personalizing the result to search nearby hotels and get information about them. In this example Pune is the location concept while hotel is the content concept.

In PMSE, backend search engines like Google, Yahoo etc. are used. Client accept the request from user and forward it to PMSE server .PMSE server is used for performing heavy tasks like reranking, content extraction and training. For maintaining privacy, user profile is maintain at client side. The same content or location concept may have different degrees of importance to different users and different queries [1].Based on users' query, content and location entropies are used for measuring information.This increase the

personalization effectiveness for particular query of user[1].Analyzing clickthrough data is a useful means to understand users' taget preference in the returned search results, since it conveys partial relative relevance judgements on the links that a user has browsed through [3].Clickthrough data was usedto optimize the ranking in search engines [4].

PMSE classifies users' query into content concept and location concept and provides coherent strategy using a client-server architecture to integrate them into a uniform solution in a mobile environment [1].By maintaining user profile at client side, search results are personalized by considering the two concepts into account.GPS is used for finding users' physical location. PMSE client-server architecture, PMSE server is used for carrying out training and reranking the search results and send it to client on Google Android Platform [1].Two parameters, *2xpRatio* and *minmax* are used for personalizing the search result and maintain the privacy level.These two parameters are privacy parameters.

2 SYSTEM ARCHITECTURE

Fig . [1] show the detailed architecutre of PMSE i.e personalized mobile search engine. The architecutre show detailed design PMSE.The architecutre divides into two parts as client side and server side.At client side user exists and at server side te actual working of PMSE exists.

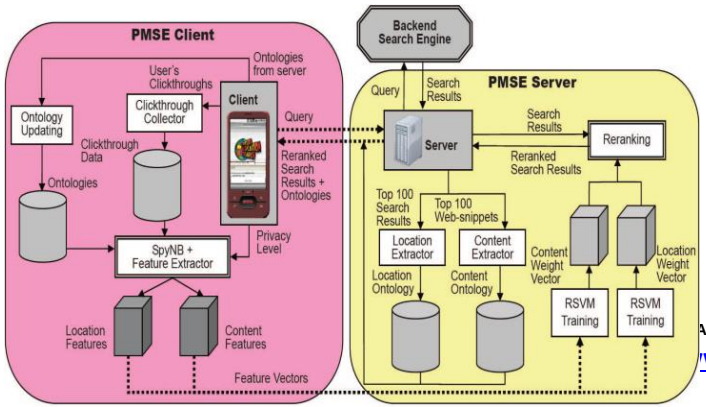


Fig.1 : System Architecture

To store click throghs ontology are used. Two types ontologies are used;one for storing contents and other is for location.Once ontogy is created then it automically update.Reranking is used for rerank the links that user visited and to show links as per user preferences.It matches the results which collected from backend search engine like Goggle,Yahoo,etc. and user preferences and rerank the result is send to s PMSE server.SpyNB is prediction algorithm used for checking the query weather it is conent or location concept.

At first user has to login to the PMSE server.After successful login to server only after he can acess to server.Once the user logged in then he can search anything.When user type any keyword this keyword is collected to clickthrogh data and the links which user visits is collected to the ontology for user checking user preferences.Spy+NB algorithm is used to predict the query into content and location.After the prediction the result is send to the reranking module.If the query is for first time typed for searching then this query is directly send to the server.No need for reranking.After the query is send to server then the server send the query to backend search engine like Google,yahoo,etc.The search results is send to reranking module.As the at the client side ontology maintains the user visited links it is also send to reranking module for rerank the links as per user perferences.At

reranking module it match the search result achieved from backend search engine like Google,Yahoo,etc. and the results achieved from ontology and Spy+NB .and the result is send to server.The final result is send to user.So by this way heavy tasks are done on server side due to low comutation power of mobile.

3 CONTENT AND LOCATION ONTOLOGIES

In content ontology, firstly the keywords are extracted from the users' query q from web-snippets. Web-snippets store the URL of webpages returned by the search results [1]. In case any query occurs frequently than it is stored in web-snippets at highest documents,so that it can referred for future preferences.Frequent item sets in datamining [5] is calculated by using following formula to measure the importance of keyword c_i with respect to query q [1]:

$$Support(c_i)=\frac{sf(c_i).|c_i|}{n} \quad [1]$$

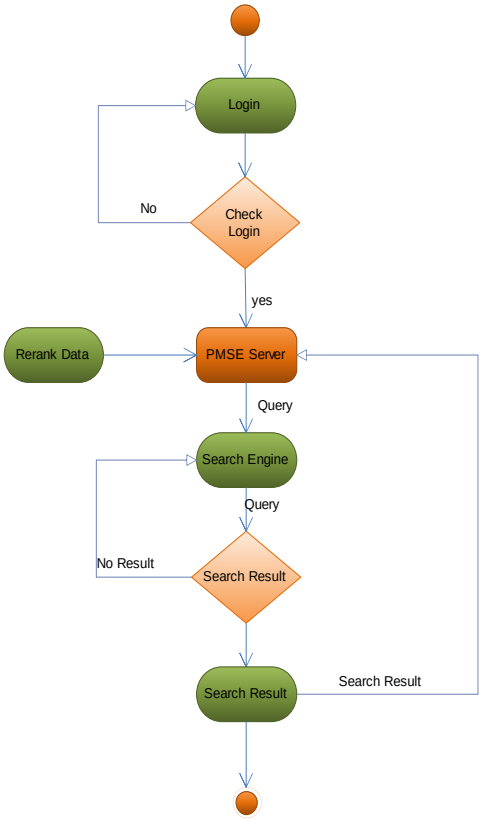
here, $sf(c_i)$ is snippet frequency of keyword [1].

$|c_i|$ is the number of terms in keyword [1].

In location ontology, users' location is detected through GPS with the help of users' clickthrough data use to create feature vectors containing user location preferences [1], which helps for reranking the search results.Location ontology store this information for future reference.

CONTROL FLOW OF SYSTEM

First user has to login to the server sucessfully.If the user name and password is wrong then login fails because only those user can acess the server who registered to server.After successful registeration user can fier query for searching data.When user fier query to server then server send this query to the backend search engines like Google,Yahoo,etc.And the reranked module matches the results and send to server.Then final result are send to users.



5 MATHEMATICAL MODEL

Let Q is a Input is given as user query to PMSE server

Q = {HI, PS, NS}

Where,

HI is History which is maintained in History of search.

PS is previous search information which maintain click through data.

NS is new search result which come by using previous search + New Result.

S is set of search engine.

S = {s1,s2,s3,s4…..sn}

Identify the searching on servers.

Q= {q1, q2, q3….qn}

Where 'Q' is main set of searching query on server q1, q2,q3…qn

q = L+ C

L is set of Location entropy,

C is set of Contents entropy.

$q(c) = -\sum_{i=1}^{k} P(c_i) log P(c_i)$

Where k is the number of content concepts C = {c1,c2,...,ck} extracted, $|c_i|$ is the number of search results containing the content concept C_i

$$|C| = |c_1| + |c_2| + \cdots + |c_k| ,$$

$P(c_i) = \frac{|ci|}{|C|}$

$q(l) = -\sum_{i=1}^{m} P(l_i) log P(l_i)$

m is the number of location concepts L ={l_1,l_2,...,l_m} extracted, $|l_i|$ is the number of search results containing the location concept l_i,

$$|L| = |l_1| + |l_2| + \cdots + |l_m| ,$$

$P(l_i) = \frac{|li|}{|L|}$

Where,

q = {HI, CTD, NS, RSVM…}

HI= History which is maintained in History of search.

CTD = Maintain Click through Data on client side and send to RSVM.

RSVM = Receive the Extracted Data on RVSM Server which are send through android mobile.

NS = new search result which come by using previous search + New Result.

SpyNB Method is the learns user behavior models from preferences extracted from click through data

SpyNB = {Po, U, PN}

Let Po be thepositive set,

U the unlabeled set, and

PN the predictednegative set

We get (PN $\subset$ U) from SpyNB.

Po = {po_1,po_2….po_n}

U={u_1,u_2,….u_n}

PN = {pn_1,pn_2,…pn_n}

Identify the processes as P.

P= {Set of processes}

P = {P1, P2, P3,P4……}

If (History found about CTD) then

P1 = {e1, e2, e3, e4}

Where

{e1=i|i is to search data on selected Search Engine}

{e2=j|j is to retrieve information on search engine}

{e3=k|k is to Send CTD to RSVM for reranking as user preference

{e4=l|l is to Check GPRS Connection on android mobile}

If (No History found about the downloading of related data) then

P1 = {e1, e2, e3}

Where

{e1=i|i is to search data on selected search engine }

{e2=j|j is to retrieve information on search engine}

{e3=l|l is to Check GPRS Connection on android mobile}

A= {S, HI, P,PS,NS,Q }

Initial conditions as I_0

(a) Android device should be activate the gprs connection

(b) Admin have good internet connection.

To control the amount of personal information exposed out of users' mobile devices, PMSE filters the ontologies according to the user's privacy level setting, which are specified with two privacy parameters, minDistance and expRatio.

MinDistance: PMSE employs distance to filter the concepts in the ontology.

Let C_{i+1} is a child of another concept c_i in our ontology-based user profile, then c_i and c_{i+1} are connected with an edge whose distance is defined by

$$D(ci,ci+1) = \frac{1}{Pr(ci+1|ci)}$$

We aim at filtering the concepts that are minDistance close to the leaf concepts, and the concept c_i will be pruned when the following condition is satisfied:

$$\frac{D(ci-1,ck)}{D(root,ci-1)+D(ci-1,ck)} < minDistance$$

Where c_{i-1} is the direct parent of c_i, and c_k is the leaf concept, which is furthest away from

$c_i(\text{argmax}_{c_k} D(c_{i+1}, c_k))$, in the ontology.

If user profile is Uq,0, while the protected user profile for the query q with minDistance p is Uq,p. Thus, the concept entropy HC (Uq,p) of the user profiles can be computed using the following equation:

$$H_C(U_{q,p}) = -\sum_{c_i \in U_{q,p}} pr(c_i) \log pr(c_i),$$

Where c_i is any concept that exists in the user profile Uq;p for the query q.

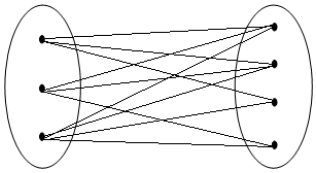
expRatio:

Given HC (Uq,0) and HC (Uq,p), the exposed privacy expRatioq,p can be computed as:

$$expRatio_{q,p} = \frac{H_C(U_{q,p})}{H_C(U_{q,0})}.$$

Users PMSE Server

P1,P2,P3,P4



Many to many relationship

6 CONCLUSION

In this paper we studied how serch results showing type of results can be classified into content and location concept.With the help of users clickthrough data we classify query into content and location concept.PMSE server handles tasks like rereanking and training.reranking of the data is done with the help of feature weight vectors, so that it help for future references for user.

7 REFERENCES

[1] Kenneth Wai-Ting Leung,Dik Lun Lee, and Wang Chien Lee, "PMSE:Personlized Mobile Search Engine",*IEEE Transaction on Knowledge and Data engineering*, 2013.

[2] Kapil Goenka," Mobile Web Search Personalizion using ontological user profile".

[3] Joachims,T.,"Evaluating retrieval performance using clickthrough data," *ACM SIGIR Workshop on Mathe- matical/Formal Methods in Information Retrieval, Tampere, Finland*, 2002.

[4] Boyan,J.,Freitag,D.,Joachims,T., "A Machine learning architecture for optimizing web search engines", *AAAI workshop on Internet-Based Information Systems,Portland,*