Research on LBS-based Optimizing Personal Recommendation System

Enhui Li, Xin Chen, Taiyan Hao
International School
Beijing University of Posts and Telecommunications
Beijing, Chian
E-mail:10213140@bupt.edu.cn

Xiangling Fu
School of software engineering
Beijing University of Posts and Telecommunications
Beijing, China
fuxiangling@email.buptsse.cn

Abstract—**In recent years, more and more internet companies trend to fuse their products with personal recommendation functions which based on a popular technology called LBS (location based service).Although the idea of this business model satisfies the characteristic of Internet industry felicitously, scientists and engineers gradually find out that the effect of the personal recommendation is worse than they expected. Due to the current technology which is not suitable and mature to make a real combination between personal recommendation and LBS module to support the internet applications, there are plenty of problems such as the network environment of the user, the location filtering, user modeling, data sparse and cold start.**

In this paper, we proposed a novel optimized personal recommendation system, which based on LBS. The system can not only solve the major problems mentioned above but also provide a good recommendation services based on customer's geographical position. According some statics we evaluate our system in an analog experiment and come to an conclusion that the new one provide a better performance than the traditional ideas.*(Abstract)*

Keywords-LBS; collaborative filtering; personal recommendation; optimizing personal recommendation [1] *(key words)*

I. INTRODUCTION

In 2009, the birth of Foursquare in America, a new topic in Internet industry after Twitter, attracts lots of attention. As a completely new style of software applying LBS technology, Foursquare has become the indicator to the future of Internet industry. However, the majority of IT Companies fail to make profits like Foursquare in this new market. According to Q4 report from Baidu which show the latest trend of mobile network in the fourth quarter in 2012, LBS takes the fourth position in the ranking list of users' searching flow. Only 7.3% of the total user's searching flow [1](showed in Figure.1) can be attributed to LBS. These statics show a fact that the potential value of LBS market was overestimated in the past few years.

Among the whole developing process of LBS, only a small amount of companies successfully earn money from this new area. The reason for this phenomenon can be attributed to the characteristics of LBS service and what customers think about these new things.

It is a common sense that an LBS service is provided to customers for free with advertisements and business news. The income of LBS service depends on these advertisements. However, it seems that customers pay less attention to these advertisements and business recommendation. According to the statics showed in Whitehouse, there are four reasons that attract customers to choose LBS service. First, keeping connection with friends; Second, getting to know the place where friends are willing to go; Third, Recording the routes of trips; The last one is receiving some information about discounts and commercial activities.

Thus, it is difficult for Internet enterprise to earn from this service although LBS is widely used. Customers are more willing to treat LBS as a service for social communication rather than a guide on shopping or finding a place for entertainment. However, the basic business model for LBS is supported by advertisement, commercial tenant and its potential increasing value. All these models need customers' willingness to follow the recommendation and purchase goods they want. Without customer's commercial and entertainment spending on this service, no business feedback can be produced.

In order to solve these problems, we need a good personal recommendation system which can provide suitable advertisements and personal recommendation to the right person. A good personal recommendation service needs to provide suggestion based on the geographical information of users. In the following part of the paper, a new personal recommendation model or algorithm will be showed to the reader. The new model has a combination of traditional LBS service and personal recommendation aiming at solving some problems which lower the effect of personal recommendation such as: the problem of network environment, traditional personal recommendation algorithm cold start problems.

II. ARCHITECTURE DESIGN FOR THE RECOMMENDATION SYSTEM

The overall flowchart of the personal recommendation design is showed as follows (Figure2). From this diagram, the whole system is divided into two main parts. They are Client App and server-side. Client is responsible for interactive services between system and users. User's requests with latitude and longitude information are also submitted to the server via client. As for the server-side,

database, location filter and CF / Predict algorithm module are in it. When the server receives a request, the user interest model in the database will be executed to get the history interest model of users. Then, the position filter functional model will take the latitude and longitude information and check the recommendation results according to requirements of the location filter. In this way the scope of the candidate recommendation set can be minimized. After that, the data will be delivered to the CF /Predict module (Collaborative Filtering Recommendation Algorithms) and checked by collaborative filtering algorithms. The systems mainly is based on the Item-based algorithm from Memory-based classification model to get the candidate set address as well as the personal Top-N recommendation set for a specific user; In the end, the recommendation result will be send to client. Besides, user's feedback about the personal recommendation may have an influence on their specific model, and the next information push will be improved based on these influences.
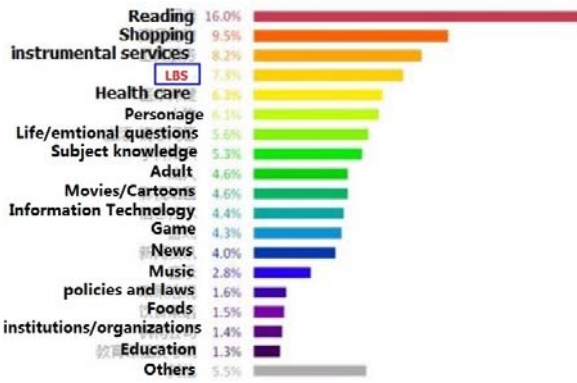

Figure 1.   The report of www.baidu.com Q4

III.   CLIENT APP

As a recommendation service based on mobile terminal, the system needs a client which has an independent interface used to communicate with the server. In order to finish the post-filtering, location information (mainly latitude and longitude) can be transmitted to the server-side via the independent port of client App. With the help of offline function component, users can still get their personal recommendation result when they are away from mobile internet. Besides, the part of the system can also make the reaction of the system more faster for the reason that client can make a quick mapping between the local offline map package and the local database without spending time on waiting for the response from server-side.

IV.   SERVER

The server side is the core part of the system which is used to process the user's request and show the results of interest
to the user as it is showed in the flow chart.
1) Screening module
Due to a huge number of relevant location information existed in the database the submitted by users, the system is
required to find out the suitable candidate set for the current user and submit them to the next stage.
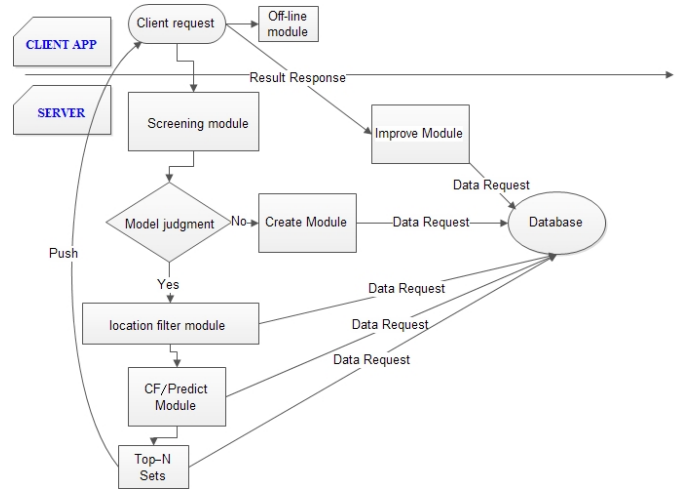

Figure 2.   The flowchart of the whole system

2) Model judgment module
After a preliminary screening of the result set, the system will automatically detect whether a historical model exist in the database. In this part, functional module is concentrate on the analysis towards user's short-term interest because of the characteristic of LBS system such as rapid change and position-sensitive features. However, this module fails to deal with user's interest shifted phenomenon. To address this issue, the scholars have proposed a method called "Long-term interest model and short-term interest model"[2].

In order to solve the disadvantage of forgotten the model. The scientists used a method of the classification processing. The user model is divided into two categories: one is the long-term interest in the model, and the other is the short-term interest model. The two models work at the same time, long-term interest model is responsible for interests that changes slowly, and the short-term interest model is obtained by analysis of special interest which changes frequently. According to a large number of studies, the method has a perfect performance. [3].

3) Location filter module
After the construction of user's interest, the system will automatically submit the information to the location filter function module. In this module, the server will check the candidate set and choose the adjacent one due to the position of user. The position information (geographic coordinates or geodetic coordinates) comes from the public interface towards telecom operators (such as the GSM network, CDMA network) or other positioning systems (eg, GPS). This paper takes reference of the clear and concise algorithm in the domestic literature[4].

In position the filter module, according to the location information in the user request (mainly latitude and longitude), the geographic distance between position A (lat1, long1) and position B (lat2, long2) is calculated according to the formula (1).If the results satisfy the condition that "dist (user, item) ≤ Distmax(the maximum distance between

Candidate items)", the system will submit position information to the classification / clustering module. The main function of this module is to filter out unnecessary items. Thus the number of candidate items will be reduced. This method can also increase the efficiency by minimizing the candidate set.

$$dist = 2\sin^{-1}(\sin^2 + \frac{a}{2} + \cos(lat1)\cos(lat2)\sin 2\frac{b}{2})^{\frac{1}{2}} ER \quad (1)$$

a(a = lat1-lat2) is the latitude difference of two points A and B; b (b= long1-long2) is the expression of longitude difference between two points A, B; ER means Earth radius in km.

4) CF/Predict module

The result item set which is selected by the location filter module will be submitted to the core module of the system-----CF / Predict. This module, as described in the literal meaning, is divided into two parts---- CF and Predict. In this paper, we take the reference from a traditional paper in this areas named "Item-Based Collaborative Filtering Recommendation Algorithms"[5]:

CF means collaborative filtering which has lots of categories. In this paper, CF refers to item-based collaborative filtering (Item-based CF).The working principle of CF is to find similar items based on user's preference. According to the user's historical preferences and recommend similar items to them. From the calculated point of view, the system treats user's preference to a specific article as a vector to calculate the degree of similarity between them. After obtaining the similarity, the system will predicted which items may attract the customer's interest according to the user's historical recording of preference. A sorted list of items will be selected as the personal recommendation. In this section, the recommendation is mainly based on user's history. For CF, the method for similarity measurement plays an important role in the module. There are several ways for similarity measurement such as cosine similarity method, relevant similarity method (Pearson), and similarity calculation based on the cloud model. These measurement methods deal with its similarity verification by using the cosine similarity algorithm to [6][7].

$$Sim(I_a, I_b) = \cos(R_a, R_b) = \frac{\sum_{k=1}^{i} R_{a,k} R_{b,k}}{\sqrt{\sum_{k=1}^{t}(R_{a,k})^2}\sqrt{\sum_{k=1}^{t}(R_{b,k})^2}} \quad (2)$$

$R_{a,k}$ represents a k-th property value of item-a, $R_{b,k}$ represents a k-th property value of item-a. After the Similarity calculation according to this formula, the system will get the Top-N items similar to the items set based on the good threshold set before. After that, those items waiting to be recommended will be scored according to a set of similar items. The prediction formula is as follow:

$$r_{u'i} = \overline{r_u} + k\sum_{n\in \hat{N}} Sim(i, \overline{n})(r_{\overline{n},i} - \overline{r_u}) \quad (3)$$

K is a normalization factor,

$$k = 1/\sum_{\overline{n}\in \hat{N}} |Sim(u'\overline{n)}|$$

usually ; $Sim(i, j)$ indicates the similarity between the items i and j; $\overline{r_u}$ represents the user u,

and is defined as $\overline{r_u} = (1/|s_u|)\sum_{s\in S_u} r_{u,i}$ where $S_u = \{i \in S \mid r_{u'i} \neq 0\}$; $\hat{N}$ is the Top-N nearest neighbor set to show the similar items with n.

Predict part is mainly based on neighbor's evaluation for the item to predict the current user preference. Among a group of neighbors, the preference of them is very close to each other and these kind of neighbors are grouped according to the historical preference recordings. The scoring $r_{c's}$ of user C to user S can be calculated by the following formula:

$$r_{c's} = \overline{r_c} + k\sum_{n\in \hat{C}} Sim(c, \overline{c})(r_{\overline{c'},s} - \overline{r_c}) \quad (4)$$

$$k = 1/\sum_{\hat{c}\in C} |Sim(c'\overline{c)}|$$

K is defined as ; $Sim(i, j)$ represents the similarity between the user i and user j; $\overline{r_c}$ indicates that the user c' is defined as $\overline{r_c} = (1/|s_c|)\sum_{s\in S_c} R_{c's}$, where $S_c = \{s \in S \mid r_{c's} \neq 0\}$; $\hat{c}$ is Top-N similarity item set for the user c neighbors' set.

After the calculation of $r_{c's}$, the Top-N recommendation set will be pushed to users.

Finally, some problems in the traditional LBS system can be solved such as the data sparse and cold start due to the use of CF / Predict algorithm. The basic idea for this algorithm is that the similarity between the articles is calculated in advance according to the historical preference data of all users. Then the recommendation created and may have some changes based on user's feedback. For the reason that the direct resemblance between items is stable, the calculation of the similarity between the different items can be pre-online. All the results will be list in the table. When the user needs recommendation, the possible scoring value will be calculated in the look-up table. Thus, the above two problems can be solved simultaneously.

5) Feedback module

The system will get some response after the receive the results of the Top-N Recommendation is received by users. The response can be a variety of type such as scoring, reviews and so on. These information will generate an impact to the improvement of the user model, so that the results are more accurate next recommended.

V. MODULE VERFICATION AND REVIES

Due to the reason that business information is difficult to obtain in real life, the validation of the model is based on user data in Foursquare to process an analog simulation, and review based on comprehensive analysis.

A. Data collection

Collecting and recording information about latitude and longitude of businesses around the Beijing Workers' Gymnasium:

TABLE I.   LATITUDE AND LONGITUDE

Neighborhood of Beijing Workers' Gymnasium		
Name	*Category*	*Details*
FishEye Cafe	Cafe	
Greyhound Cafe	Cafe	
Luga's Pho-Pho	Restaurant	Western cuisine
La Pizza	Restaurant	Pizza
Must Guette Hot Dogs& Bagels	Bagel Shop	Hotdog
Flamme-The Ultimate Steak	Steakhouse, BBQ Joint	Steak
Kocoon Spa	Spa	Massage , Gym pool
Coolth KTV	KTV	
Village North	Shop&Service	
Transit	Chinese Restaurant	

In addition:

We choose 50 users as a basic sample and make a survey among them. The distribution of its long-term and short-term interest distribution is as follow:

B. The verification of the model:

There are two metrics for accuracy predication. One is statistical accuracy metrics and the other is decision support accuracy metrics. Due to the reason that we are lacking of enough size of data, we take several statistical accuracy method to measure the error the rang between the recommended data and the actual evaluation data such as : Mean absolute error MAE (Mean Absolute Error), the root mean square error (standard error) RMSE (Root Mean squared Error) standard average absolute error NMAE (Normal Mean Absolute Error), etc.;
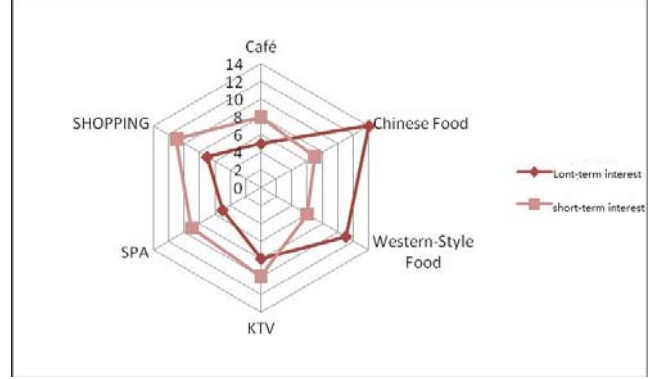

Figure 3.   The distribution of its long-term and short-term interest distribution

Common MAE standard is easy to understand and calculate the average absolute difference to calculate the predicted score and the actual score intuitive measure of the accuracy of prediction Recommended. MAE smaller the value, the higher the precision is recommended. Let the user project prediction score set {P1, P2, ... Pn}, corresponding to the actual score set MAE expressed as follows:

$$MAE = \frac{\sum_{i=1}^{n} |p_i - r_i|}{n} . \tag{5}$$

Different square deviation component of RMSE which is present as $(p_i - r_i)^2$ occupies different proportion. The larger square deviation component is, the more the sum of square deviation component will be. In this point, RMSE is the same as MAE. The lower it is, the higher precision the result of recommended will be. The calculation equation of RMSE is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - r_i)^2}{n}} \tag{6}$$

C. Combination between experimental data and the system model

Make an experiment according to data so as to verity the function of the system. The statistics of 50 corresponding user's short-term and long-term interest will be substituted into the system model. In the system, interest will be predicted first which means making an training through the collected data and predict based on the training results. Second, the predicted results will be submitted to the model and maintain the final recommendation with the help of collaborative filtering.

D. Verify the results and evaluation

Adding the experimental data into the optimized personal recommendation system based on LBS, after several tests, and compared with the traditional and is being widely used

personal recommendation algorithm, the results are as follows in Figure 4 and Figure 5:

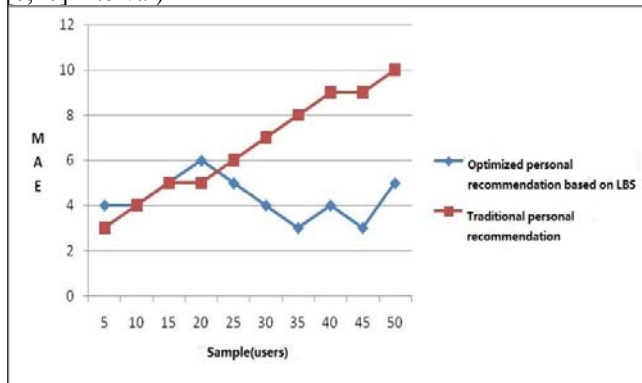(MAE, RMSE dimensionless tends to make the results [0,10] interval)

Figure 4.   MAE comparison:

RASE and MSE line graph comparison, we can clearly see that the use of optimized personal recommendation model with LBS significantly improve the accuracy of the recommended. And increases with the sample volume, and the system exhibited a good stability.
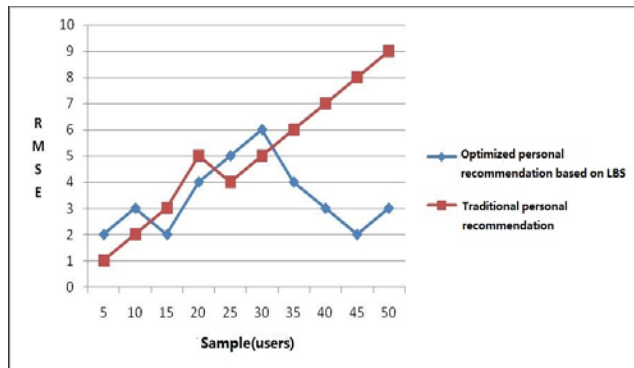

Figure 5.   RMSE comparison:

VI.   CONCLUSION

Based on the combination between LBS and personal recommendation algorithm, the recommendation system can well solved the problems mentioned at the beginning of this paper. After collecting some basic information actual data from applications, the system can successfully extract business value from those data and benefit both customers and merchants. After collecting some basic information actual data from applications, the system can successfully extract business value from those data and benefit both

customers and merchants. As for customers, they can get more useful recommendation and find out the exactly items or services they require. As for business, the new system may bring them more profits by generating more sales. With the optimizing recommendation system with LBS elements it can be more effective and reliable for users and problems such as instability of user's network environment, imprecise location information, lacking effective user model would be solved gradually.

After collecting some basic information actual data from applications, the system can successfully extract business value from those data and benefit both customers and merchants. As for customers, they can get more useful recommendation and find out the exactly items or services they require. As for business, the new system may bring them more profits by generating more sales. With the optimizing recommendation system with LBS elements it can be more effective and reliable for users and problems such as instability of user's network environment, imprecise location information, lacking effective user model would be solved gradually.

In the future, this research on LBS-based optimizing personal recommendation system can be commercialized with some business strategic partners. The big data can help our system produce a high-quality performance.

REFERENCES

[1]   Baidu   mobile   trends   report   for   the   fourth   quarter http://labs.chinamobile.com/yueke/show/?url=http://it.sohu.com/2013 0226/n367151085.shtml

[2]   .YangJie. Research and application for personal recommendation [D].University of science and technology of China.2009.

[3]   Kim H R&Chan PK.Learning Implicit User Interest Hierarchy for Context in Personalization[J].1n:Applied Intelligence, 2008:153 一 166.

[4]   Zhang chunyong Chen qun, a technology and project on mobile personal recommendation based on.2011

[5]   Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl，Item-Based Collaborative Filtering Recommendation Algorithms,

[6]   Liu jianguo, Zhou tao, Wang binghong, the process of personal recommendation research. The process of science, 2009;19(1):1—5

[7]   Huang chuangguang, Yin jian, Wang jing, etc.. collaborative filtering algorithm with unsure neighborhoods for personal recommendation. Computer science newspaper, 2010; 33(8) : 1369—1377