Scientific Journal of Impact Factor(SJIF): 4.14
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue for ICPCDECT 2016,Volume 3 Issue 1

Improve Location Based Services Using Data Mining Techniques

Miss.Chandni Kumbhar[1],Miss.Ashwini Waghmode[2],Miss.Shital Ranmode[3],Miss Komal Kokane[4],ProfGavali A. B.[4]

Department of Computer Engineering
S. B. Patil Collage Of Engineering, Indapur, Dist. - Pune.
SavitaribaiPhule Pune University.

Abstract- *In the recent years, with the rise of mobile technology new service, Location Based Service, has developed. Reliable LBS applications or services mostly depend on the accuracy as well as integrity of location data .The most important thing is timely updated attributes. Identifying and organizing the location information is a very time consuming .Also human error is another fact that is unable to avoid. How to improve the accuracy and quality of location data through the data de-duplication process using web mining techniques till up to now, not many previous studies have been done. Our proposed system is an efficient and cost-effective way to integrate the location information .Which is collected from multiple data sources. The main advantage is that in our proposed system by using various data mining techniques de-duplication process is that automatically performed. Different algorithms have been used for effective and efficient handling the location data integration task. In this experiment data mining tool is used that is WEKA tool, it consists of in build algorithms.*

Keywords*- Data mining; Location based Services(LBS); GIS; Classification.*

I. INTRODUCTION

Now a day, all are becoming smart phone users. In large quantity, growth have experienced from 2008. From the paper [1], we could confirm global smart phone shipments grew 43 percent in 2012.The combination of smart phone and high-speed mobile internet fulfills our all requirements and important tasks. It also enables to attract the number of customers with the different companies. The most commonly used mobile application and services consist of the following: social networking, discovering of places of interest, recommendation of social events in local area and social engagement based on the current location. No of companies are providing various Location Based Services. Accurate and timely updated location information are the most important concepts to make the use of location based applications and services more effective. Here in this paper we use the concept of data mining [2]. Due to the data mining concept, we can analyze the previous values and can predict the new values.

The application service provider and the map service provider have to take great efforts to collect, analyze, update, and integrate the location information on a regular basis. In this way location information consists of different attributes and may not be well-formed from one source to another. The integration process for the location information is a critical step in case of accuracy.

The aim of this paper is to improve the Location Based Services by using data mining techniques. In this paper different types of algorithms are applied on the input set. Nomao dataset is used as input set [3]. On Nomao dataset data mining WEKA tool is used which contains in build algorithms [4].

Remaining paper consists different sections that are follows. Section 2 contains related work of given system Section 3 gives detailed description of our proposed system. Section 4 describes about architecture and advantages of proposed system. Section 5 describes the algorithm used in paper. Section 6 describes the comparative study of both system and Section 7 gives the conclusion of this paper.

II. RELATED WORK

In the paper [5] Dong and Guo, one application studied that is MSNSs. According to their opinion, locations as well as community and time specific information are required information. After completion of study on the current applications, they proposed a new service pattern, named A-LBS (Advanced Location Based Service). That's useful to improve works related to mobile social network service. The A-LBS is a location dependent service pattern. In which of 3D platform and GIS technology are included .This paper gives a real-life Campus-MSNS application for A-LBS pattern. The proposed pattern design is useful for mapping the real-world activity onto a computerized 3D platform, that's useful to search friends' location or activities.

In the paper [6] JinsooAhn;Suyoung Lim; JungilHeo; Wooshik Kim, Jan a researched on the usability and feasibility for the futures of healthcare system. This is useful to identify patient's location data. As well as new concept introduced LBS and Location Determination Technology (LDT).which is very important in case of LBS implementation. By comparing two methods, it could understand that network-based methods have low error rates. However, additional stations are required to achieve such performance results. From comparative study, previous methods and the LBS method suggest that the proposed method can make great contribution to the emergency healthcare domain with real-time location of the patients.

In the paper [7], it could understand that Receiver Operating Characteristic (ROC) is a very difficult measurement to understand at the time of calculating experimental results and data mining algorithms. According to the Powers opinion, using classification accuracy rate, recall, precision, and F-measure will not produce the best result. Having this limitations, on specific dataset or problem domains, Powers further introduced how to use ROC, and to use ROC with multiple other measurements to obtain more accurate evaluation results.

Sasaki [8] wrote a research paper in 2007 on F-measure. This gives the definition and the calculation of different mathematical formulae. The concept of variable also explained which is the parameter that controls the balance between P and R. And F-measure evaluation also explained. Different evaluation methods are in the data mining .Which includes classification tasks, such as classification accuracy rate, Type I error rate and Type II error rate.

Zhang et al. [9] proposed the architecture for LBS system and the thought of constructing spatial index to improve search efficiency. Location based applications and services provide us the different way that we retrieve information and improve our personal lives. There are many LBS enabled applications developed recently which helped people within many sectors in a great extent. Because of having some drawback, traditional key-word search engine have their own deficiency On certain areas. This takes geographical factors into account.

III. INTRODUCTION TO OUR PROPOSED SYSTEM

A. Problem Statement

From above related work we have addressed following problems in existing system:

- Identifying and organizing the location information can be very time consuming process.
- As well as Human error is the inevitable fact.

In our proposed system we use WEKA tool .We know that WEKA tool consists of different classification data mining algorithms and data mining is useful to predict the values by using previous values, which are already available i.e. Prediction technique.

B. Proposed System as a solution for Existing System

We propose an efficient and cost effective method to integrate the location information. This is collected from multiple data sources and the De-duplication process which is automatically performed by using Data mining techniques. We are going to use data mining tool on Nomao dataset, which is WEKA tool. This consists of in build data mining algorithms. Due to this concept, required time will be reduced for identification and organizing the location information.

IV. BLOCK DIAGRAM / ARCHITECTURE OF PRPOSED SYSTEM

Following diagram shows the flow of our proposed system.

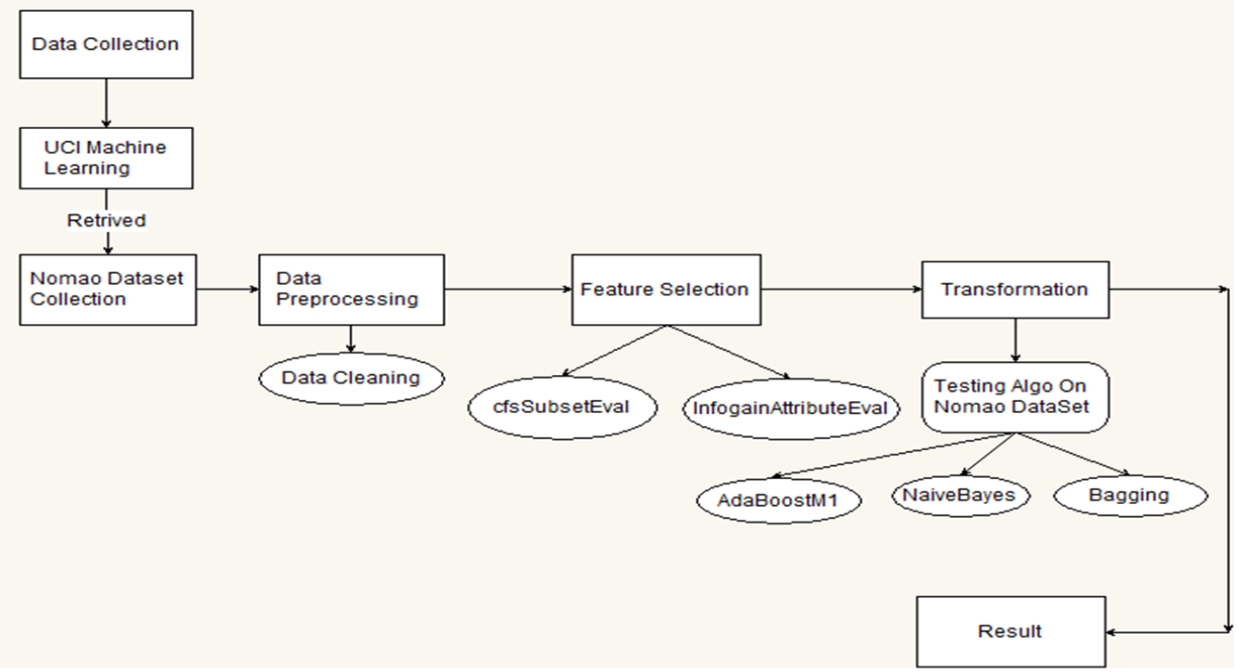

Figure 1: Block diagram of proposed System.

A. Flow of the project

The flow of our proposed system is explained in this section. The data mining process consist of different steps that are,

- Data collection
- Data preprocessing
- Feature selection
- Data transformation
- Modeling
- Evolution

In our proposed system Nomao dataset is selected. A different data algorithms and feature selection methods will be applied on Nomao dataset which is retrieved from UCI M/C learning repository. This Nomao dataset contains 34465 instances and each instance is represented by 120 attributes. After collection of Nomao dataset, data cleaning process is performed. After that data model is built. To remove the missing data in the Nomao dataset the Replace Missing values filter function in Weka data mining tool is used.
NaiveBayes was selected because NaiveBayes is proven to be a simple, efficient and effective algorithm that can produce reasonable performance on most of the data types. The rest of tree algorithms were selected based on the experience that these algorithms can usually produce better results. The experiment is conducted in three stages. The first stage is to run the six chosen algorithms against the Nomao dataset using cfsSubsetEval feature selection method. The second stage is to run the same six algorithms against the Nomao dataset using the informationGain feature selection method. The last stage is to run the six algorithms against the Nomao dataset using cfsSubsetEval but with two fewer attributes.

The different algorithms are applied on Nomao dataset, which are listed as:

- AdaBoostM1
- Bagging

- Naïve Bayes

The 10-fold cross validation on the complete dataset is used for each stage of the experiment using 10-fold cross validation every instance in the complete dataset will be used as both the training data and testing data. After each stage of the experiment results are recorded.

B. Advantages of proposed system

- We can obtain the desired destination within less amount of time by using data mining technique. Because of that we can save our searching time.
- As well as one more advantage is that it's cost effective.
- Due to automatic De-duplication process, it removes redundant data and saves the memory.

V. USED ALGORITHMS

In our proposed system we are using various algorithms. Which are given below **.**

A. AdaBoostM1 [10]

INPUT

Sequence of m example {(x1, y1) ----- (xm, ym)with
Labels Yi€ Y={1----k} Weak learning algorithm Weak learn
Integer T specifying number of iterations Initialize D1(i)=1/m for all i
Do for t=1,2,----T

1] Call Weak Learn, Providing it with the distribution Dt.

2] Get back a hypothesis h_t:X→Y

3] Calculate the error of h_t: $€_t=\Sigma Dt(i)_{.i:ht(xi)\neq yi}$If $€t>1/2$, then set T=t-1 & abort loop.

4] Set $Bt=€_t/(1-€_t)$.

5] Update distribution Dt.

$D_{t+1}(i) = {}_{(i)}X\{Btifht(xi)=yi\ 1\ Otherwise$
Where Zt is a normalization constant (chosen to that D_{t+1})will be distribution).
S -The final hypothesis: $h_{fin}(x) = \arg\max \Sigma \log 1/_{:h()=y}$
S - The final hypothesis: $h_{fin}(x)=\arg\max \Sigma \log 1/Bt_{t:h}$

B. Bagging [11]

TRAINING PHASE

1] Initialize the parameters
a) D= {}, the ensemble.
b) L, the number of classifiers to train.

2] For K=1, -------, L
a) Take a bootstrap sample SK from Z,
b) Build a classifier Dk using SK as the training
Set,

c) Add the classifier to the current ensemble D=D
U Dk.

3] Return D.

CLASSIFICATION PHASE

4] Run D1, -----DL on the input X,

5] The class with the maximum number of votes is choosen as the label for X.

C. NaiveBayes

INPUT

1] Given features X_1, X2,…,X_nPredict a label Y

2] Parameters for modeling P(X_1,…,X_n|Y): 2(2_n-1)

3] Parameter for modeling P(X_1|Y),…,P(X_n|Y): 2n.

OUTPUT

1] Estimate P(Y=v) as the fraction of records with Y=v
P(Y = v) = $(Y=v)\#\ records$

2] Estimate P(X_i=u|Y=v) as the fraction of records
with Y=v for which X_i=u
P(Xi = u | Y = v) = $Count(Xi=u \wedge Y=v)Count(Y=v)$

This corresponds to Maximum Likelihood estimation of model parameters

• The *Naïve Bayes Assumption*:
• Assume that all features are independent given the class label Y
Equationally speaking:
P(X_1, . . . ,Xn | Y) = $\Pi p(Xi_{ni=1}$ | Y).

VI. COMPARATIVE STUDY

->	**Time**	**Memory**	**Cost**	**De-duplication**
Existing System (Web Mining)	More Time	Wastage of Memory	More Costly	Dependent Process
Proposed System(Data Mining)	Less Time	Saving of Memory	Cost Effective	Automatic Process

Table 1. Comparative study

As we compare existing system with our proposed system, we can state that our proposed system is more beneficial than existing system. Because, existing system requires more time for searching locations whereas proposed system requires less amount of time.

In existing system because of redundant data, there is wastage of memory. That's why it is more costly. On the other hand, our proposed system removes redundant data, so that it can save the memory. Having this reason our proposed system is cost effective. Data de-duplication is a dependent process in existing system, whereas in our proposed system Data de-duplication process is automatically performed.

VII. CONCLUSION

In this paper, we have proposed a model which will be important for the improvement of accuracy and integrity of the location information for Tourist application. Which depends on timely updated information. As well as theoretical analysis shows us how the updated information affects on user's expected output in desired time.

REFERENCES

[1] "Strategy Analytics. Global smartphone shipments reach a Record 700million units in 2012. January 25, 2013.
[2] Ling Liu, Zijiang Yang and YounesBensliman, "Using Data Mining Techniques to Improve Location Based Services".
[3] UCI Machine Learning Repository. www.ics.uci.edu/~mlearn.
[4] Weka. www.cs.waikato.ac.nz/ml/weka.
[5] Dong Fengyu; Gou Xuerong, "An advanced location based service (ALBS) on mobile social network," 2009. IC- BNMT '09. 2nd IEEE International Conference on , vol., no., pp.740,743, 18- 20 Oct. 2009.
[6] JinsooAhn; JungilHeo; Suyoung Lim; JunghoSeo; Wooshik Kim, Consumer Electronics, "A Study of Healthcare System for Patient Location Data Based on LBS," 2008. ICCE 2008. Digest of Technical Papers. International Conference on , vol., no., pp.1, 2, 9-13 Jan. 2008.
[7] Powers, D.M.W.,"The problem of Area Under the Curve," Information Science and Technology (ICIST), 2012 International Conference on ,vol., no., pp.567,573, 23-25 March 2012.
[8] Sasaki, Y. (2007). The truth of the F-measure. Teach Tutor mater, 1-5.
[9] Zhang Leilei; Kunhui, L.; Ye Weichuan, "Location-Based Mobile Search: And the Application of Searching for Hotels," Instrumentation,Measurement, Computer, Communication and Control (IMCCC), 2012 Second ternational Conference on , vol., no., pp.644,647, 8-10 Dec ,2012.
[10] Yoav Freund Robert E. Schapire "Experiments with a New Boosting Algorithm", AT & T Research 60 Mountain Avenue.
[11] AmitSrinet ,DaveSnyder.,"Bagging and Boosting Algorithm"
[12] Jonathan Huang and I. H. Witten's , E. Frank's and Jeremy "Naive Bayes Classification",and "Data Mining"
[13] Dr. Sudhir B. Jagtap, Dr. Kodge B. G., "Census Data Mining and Data Analysis using WEKA", (ICETSTM – 2013) Singapore.
[14] "Location-Based Services System (LBSS) Stage Description 3GPP2S.R00 19.
[15] "Data Mining Concepts and Techniques" Han, J.Camber, 3rd edition Morgan Kaufmann.

