To Improve Location Based Services using Weka Tool

Miss. Chandni Kumbhar[1] Miss. Ashwini Waghmode[2] Miss. Shital Ranmode[3]
Miss. Komal Kokane[4] Prof. Gavali A. B.[5]
[5]Assistant Professor
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]S. B. Patil Collage of Engineering, Indapur, Dist - Pune. Savitaribai Phule Pune University

Abstract— Now days, number of peoples are using smart phones. With the rise of mobile technology new service, Location Based Service (LBS), has been developed in smart phone. Reliable LBS services as well as applications mostly depend on the integrity and accuracy of location data. As well as services depend on updated attributes. Identifying and organizing the location information is a very time consuming process. Human error is also another fact that is unable to avoid. We have to face, such types of problems. How to improve accuracy and quality of location data through the data de-duplication process using web mining techniques was not possible. Our system is an efficient and cost-effective way to integrate the location information. Which collects data from multiple sources. The most important advantage of our system uses various data mining techniques that automates de-duplication process. Different algorithms used for effective and efficient handling the location data integration task. In our system data mining tool used, that is the WEKA tool. It consists of in build algorithms.

Key words: Data mining, Location based Services(LBS), GIS, WEKA

I. INTRODUCTION

In the recent years, mostly smart phone is used. In large quantity, growth have experienced after 2008. From the paper [1], we could understand global smart phone shipments grew 43 percent in 2012.The combination of smart phone and high-speed mobile internet fulfills all user requirements. As well as number of customers attracts with different companies. Frequently used mobile application and services consist of the different concepts: social networking, discovering of places of interest, and social engagement based on the current location. Different companies are providing various Location Based Services and applications. Not only accuracy but also timely updated location information are the most important concepts to make the use of location based applications and services more effective. We used the concept of data mining techniques [2]. Due to the data mining concept, we could analyze the previous values and could predict the new values.

The application service provider and the map service provider have to take efforts to collect, analyze, update, and integrate the location information on a regular basis. In this way location information consists of different attributes. The integration process for the location information is more complicated step in case of accuracy.

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

Remaining paper consists different section that are follows. Section 2 contains related work of given system Section 3 gives detailed description of our system. Section 4 describes about architecture and advantages of our system. Section 5 describes the algorithm used in system. 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. Locations as well as community and time specific information are required information, according to the authors opinion. They proposed a new service pattern, after completion of study on the current applications, named A-LBS (Advanced Location Based Service). That was useful to improve works related to mobile social network service. The A-LBS depends on location information attributes. In which GIS technology and 3D platform are included . It 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 is useful to search person's location or activity.

In the paper [6] Jisnsoo Ahn; Suyoung Lim; Jungil Heo; Wooshik Kim, Jan a researched on the feasibility and usability for the futures of healthcare system. It is useful to identify patient's data as well location information. New concept introduced LBS and Location Determination Technology (LDT).That is very important in case of LBS services and application. When two methods compared, it could understand that network-based methods have low error rates. Stations are required to achieve such performance results for particular application. With the help of comparative study, previous methods and the LBS method suggest that the method makes great contribution for the emergency healthcare system with real-time location of the patients.

In the paper [7], it could understand that Receiver Operating Characteristic is more difficult measurement to understand at the time of calculating experimental results as well as 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. It gives the definition and the calculation of different mathematical formulae. The concept of variable also explained, parameter that controls the balance between P and R. As well as F-measure evaluation also explained. Different evaluation methods are involved in the data mining techniques. That consists 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 introduced the concept of constructing spatial index to improve search efficiency. Location based services and applications provide us the different way. Due to 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. Due to some drawback, traditional key-word search engine have their own deficiency on certain areas. That takes geographical factors into account.

III. INTRODUCTION TO OUR EXISTING SYSTEM

A. *Problem Statement:*

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

1) Identifying and organizing the location information can be very time consuming process.
2) As well as Human error is the inevitable fact.
3) In our existing system we used 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.

IV. BLOCK DIAGRAM / ARCHITECTURE OF EXISTING SYSTEM

A. *Working of Our Existing System:*

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

- Data collection
- Data pre-processing
- Data transformation
- Modeling
- Evolution

Following diagram shows the flow of our existing system with six algorithms.

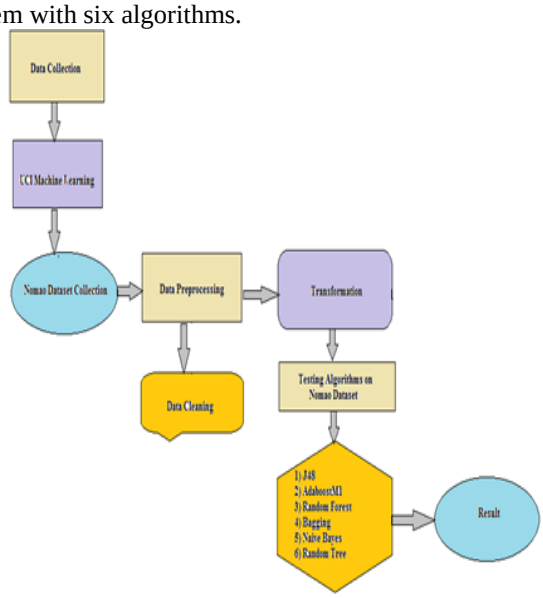

Fig. 1: Architecture of Existing System

In our system Nomaodataset used. A different data algorithms applied on Nomao dataset. Which retrieves dataset from UCI M/C learning repository. This Nomao dataset consists of 34465 instances and each instance represents 120 attributes. After collection of Nomao dataset, data cleaning process performed. After cleaning process data model built. To remove the missing data in the Nomao dataset the Replace Missing values filter function used. That is in build in WEKA tool.

Naïve Bayes selected because Naïve Bayes proved to be a simple, efficient and effective algorithm. It produces reasonable performance on most of the data types. The rest of tree algorithms selected based on the experience that these algorithms usually produce better results. The experiment conducted in three stages. In the first stage six algorithms applied against the Nomao dataset. In the second stage classification is applied. In the last stage we calculated and compared all the values of every algorithm. Finally, conclude that Naive Bayes is best algorithm for Pizza Owner service provider application.

The different algorithms applied on Nomao dataset that are following:

- AdaBoostM1
- Bagging
- Naive Bayes
- Random Forest
- Random Tree
- J48

The 5-fold cross validation on the complete dataset used for each stage of the experiment. Using 5-fold cross validation every instance in the complete dataset used as both the training data and testing data. After each stage of the experiment results recorded.

V. COMPARATIVE STUDY

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

Table 1: Comparative Study of Previous and Existing System

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

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

VI. RESULT AND ANALYSIS

This section contains the experimental result. This experiment conducted on Windows 7 operating system. We used WEKA tool, which has six inbuilt algorithms. By comparing all the values we could decide Naive Bayes algorithm is better for searching the results. In our system we developed Pizza Owner service provider application. We apply all six algorithms on dataset, we conclude that Naïve Bayes is the best algorithm for that application. Here we calculated all values, with the help of values we created graphs. Graph for correct and incorrect values, in that we calculated both values and concluded which graph gives highest correct value.

A. Result Comparison by Applying Different Algorithms:

1) Figure 1 Shows Correct and Incorrect Values Comparison:

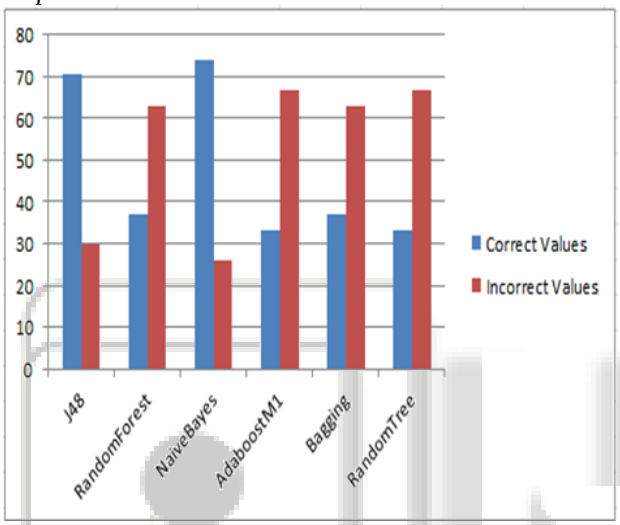

Fig. 1: Comparison of Correct and Incorrect Values

In experiment we calculated correct and incorrect values. With the help of both values we could decide which algorithm is best. From above graph we get correct value (70.374) and incorrect value (29.6296) for J48. Whereas, Random Forest gives correct value (37.037) and incorrect value (62.963). Naive Bayes gives the highest correct value (74.074) and lowest incorrect value (25.9259).AdaboostM1 and Random Forest both algorithm give the same correct value (33.3333) and incorrect value (66.6667). Bagging algorithm gives correct value (37.037) and incorrect value (62.963).By comparing all values, we understood which algorithm gives more accuracy.

2) Figure 2 Shows Comparison of Precision and Recall Values:

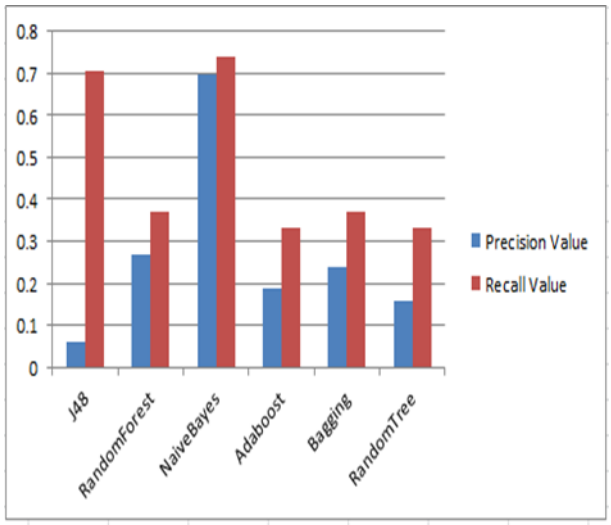

Fig. 2: Comparison of Precision and Recall Values

In experiment we also calculated Precision and Recall values. Precision value calculated using this formula ((Retrieved Value + Relevant Value) / (Relevant)). Whereas, Recall value calculated with the help of ((Retrieved Value + Relevant Value) / (Retrieved Value)).From above graph we get precision value (0.602) and recall value (0.704) for J48. Whereas, Random Forest gives precision value (0.267) and recall value (0.370). Naive Bayes gives the precision value (0.699) and recall value (0.741).AdaboostM1 gives precision value (0.186) and recall value (0.333). Bagging algorithm gives the precision value (0.240) and recall value (0.370). Random Tree gives precision value (0.157) and recall value (0.333).By comparing all values, we understood which algorithm gives highest precision value and lowest recall value.

VII. CONCLUSION

In this paper, we have discussed the importance of the accuracy as well as integrity of location information in today's mobile application. The experimental results show which algorithm is best for application.

We developed pizza Owner service provider application. From experimental result we conclude Naive Bayes algorithm is best.

REFERENCES

[1] "Strategy Analytics. Global smartphone Shipments reach a Record 700million Units in 2012.January 25, 2013.
[2] Ling Liu, Zijiang Yang and Younes Bensliman, "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] Jinsoo Ahn; Jungil Heo; Suyoung Lim; Jungho Seo; 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 UnderThe Curve," Information Science and Technology (ICIST), 2012 InternationalConference 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 Hotel 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] Amit Spinet, Dave Snyder., "Bagging and Boosting Algorithm"

[12] Jonathan Huang and I. 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.

