

Predictive Analysis of Big Data- Prediction of Mobile User's Next Location



^{#1}Rohan Tigadi, ^{#2}Arundhati Naikavde

¹rohan5393@gmail.com
²arundhati0002@gmail.com

^{#12}Department of Information Technology,
Sinhagad Institute of Technology and Science, Narhe
Savitribai Phule Pune University, Pune
India

ABSTRACT

Generally analysis of RDBMS and other conventional data sources is provided by several architectures and module application. In today's world, mobile phones have become a necessity in the field of communication. This has resulted in creation of a Big Data set in various mobile network providers. Analysis of Big Data can be used for several purposes and one of the major significance lies in the predictions made on the big data analytics.

Keywords: Analysis of Big Data, Predictive Analysis, Mobile Phone Users.

ARTICLE INFO

Article History

Received: 11th September 2016

Received in revised form :

11th September 2016

Accepted: 13th September 2016

Published online :

16th September 2016

I. INTRODUCTION

One of the interesting fields of prediction is about predicting the location of people based on their mobile phone usages. However this has become quite a tedious task due to two main reasons: the usage data is very large and in an unstructured format also there are several parameters for specifying the location and the time. This makes it much more complicated to determine common user behaviour patterns.

However, the previous approaches for the same could be summarized as not efficient in terms of big data sets, but in usual data sets where data set could be structured and in an organized form. This approach would overcome the conventional issues faced in big data sets and provide an efficient way to analyse the large data.

The key components of the proposed approach is identifying the patterns found out in the big data set, and using the same for futuristic predictions which here refers to the prediction of mobile user's next location.

II. MAIN IDEA

The primary objective of this analysis of big data could be formatted into two parts, namely:

A. Pre-processing Unstructured Data Set

The call detail records (CDR) of every mobile users would be maintained by every mobile network service provider. This would result in humungous data in a very much unstructured manner on which bare analysis would be very complex. In order that this data be in a much simpler manner, the first step would involve the conversion of the unstructured format to structured data by means of big data platforms.

Also this would include identifying the patterns in which a particular mobile user would have travelled. This would help in arranging the data set in a pattern-confidence type of data set which could be further used during the prediction phase.

B. Applying the Predictions to the Pre-processed Data

The already pre-processed data would be then provided to the prediction algorithm which actually predicts the futuristic cases. These futuristic predictions would be however based on the frequency of the patterns occurring in the CDR.

Sequential patterns which are organized in the pre-processed data are analysed and picked out with respect to the instantaneous input given to the prediction algorithm. The pattern with the highest confidence is finally concluded and the next mobile user's location would be given as an output of the predictive analysis.

III. APPROACHES AND ALGORITHM

A. Data Organization Step

Initially the CDR would be in the patterns of several repetitions of form of: (BSID,time). Also there would be some unnecessary attributes present like names of customers, addresses, email, contact info etc. All of these unnecessary attributes would be neglected during the formation of the pre-processed data. The series of such patterns would be analysed by the big data analytics platforms and finally the pre-processed data would be comprising of data of the patterns of BSIDs (Base Station IDs).

Further to enhance the processed data, the frequency of every individual pattern (support) is recorded and appended at the end of each pattern. This would be used in determining the confidence/support of a particular pattern required at the time of application of prediction algorithm.

Fig.3.1 shows the ideal scenario of pre-processing the data.

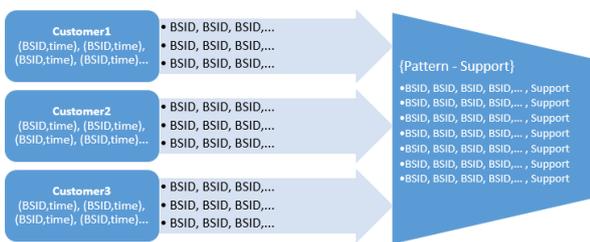


Fig.3.1 Pre-processing Data

B. Prediction Step

The support appended at the end of each pattern would be used in for the prediction stage. The actual prediction algorithm would be on the basis of Apriori Algorithm.

A set of BSIDs would be given as input to the Apriori algorithm as a key. The key consists of patterns of BSID. All the instances of this pattern occurring in the pre-processed data is fetched. The pattern with the maximum support is chosen from amongst the matching patterns. Accordingly the next location of the mobile user is predicted based on the output of the Apriori algorithm.

```

Algorithm 1 Prediction Algorithm


---


Input: key
Output: prediction
1: maximumSupport ← 0
2: for all pattern ∈ frequentPatterns do
3:   if key ≡ pattern[1 : k - 1] then
4:     if maximumSupport < pattern.support then
5:       prediction ← pattern[k]
6:       maximumSupport ← pattern.support
7:     end if
8:   end if
9: end for
10: return prediction


---


    
```

Fig.3.2 Apriori Algorithm for Prediction

Fig.3.2 shows the working of the prediction algorithm applied.

IV. EXPERIMENTAL RESULTS

The system calculates the predictions based on the analysis of Big Data. The results vary according to the amount of data provided in for the predictive analysis.

Prediction Accuracy \propto Size of Big Data

It was observed that the accuracy of the predictions were directly proportional to the size of the data under analysis. Also the predictions were limited and a flat scale was observed after a certain size of data which stated that the prediction accuracy was stagnant after a considerable amount of data.

The following graph Fig.4.1 states overview of the experiments carried out:

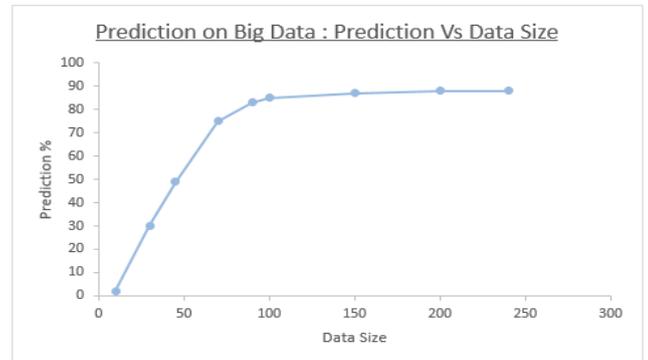


Fig.4.1 Prediction (accuracy) vs Size of Big Data

Based on the experimental observations, it would be concluded that the approach maximizes the performance of the prediction algorithm despite huge data sets.

V. CONCLUSION

Unlike the relational analysis which would be much of a tedious task to compute and predict the futuristic changes, the analysis of big data plays a vital role. Predictions on unstructured data that can be an aggregated source of information can thus be applied the newly big data analytics framework. Based on the outcomes the mentioned prediction algorithm would help in further innovative processes for instance here, the prediction of the mobile user's next location.

ACKNOWLEDGEMENT

This research was supported by project guide Prof.S.R.Gote under the project title "Predictive Analysis of Big Data : Prediction of Mobile User Location" 2015. Thanks and appreciation to the Department of Information Technology, Sinhgad Institute of Technology and Science, Narhe.

REFERENCES

[1] Mahdi Esmaeili1 and Fazekas Gabor, "Finding Sequential Patterns from Large Sequence Data" (IJCSI, January 2010).

[2] Mining Big Data: Current Status, and Forecast to the Future Wei Fan Huawei Noah's Ark Lab Hong Kong Science Park Shatin, Hong Kong Albert Bifet Yahoo! Research Barcelona Av. Diagonal 177 Barcelona, Catalonia, Spain.

[3] Intel "Big Data" Science and Technology Center Vision and Execution Plan Michael Stonebraker, Sam Madden, Pradeep Dubey stonebraker@csail.mit.edu,

