International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 9, September- 2014

An Overview on Mobile Data Mining

Use the data generated from your mobile phone to obtain useful knowledge

Neelmay Desai
Computer Department
D.J.Sanghvi College of Engineering
Mumbai, India

Prof. Lynette D'Mello
Computer Department
D.J.Sanghvi College of Engineering
Mumbai, India

***Abstract* :- During initial days, mobiles were only used as a medium for communication. This view of mobile phones as communication devices changed with the advent of a new category of mobile devices known as smart phones. These smart phones along with communication are capable of doing things that a computer can do. With the advancement of technology, these smart phones are becoming more and more powerful in terms of storage and computing features. The smart phone can provide us with data about various aspects such as movement of the user, behavior of the user, communication, etc. Today, since, most of the communication takes place through mobiles, it is very important for the analyists to analyze the data generated by mobile phones. A large amount of data is generated by mobile phones as the use of mobile phones is increasing continuously. In this paper, an overview of mobile data mining is provided and its techniques and applications are discussed.**

Keywords— Data Mining, Data Analyzing

I. INTRODUCTION

Data Mining is the approach that is applied on large, clean data sets to gain useful information and knowledge form it. Data mining help us in identifying patterns in this large data. It is applied to extract information where the data is large and knowledge is less.

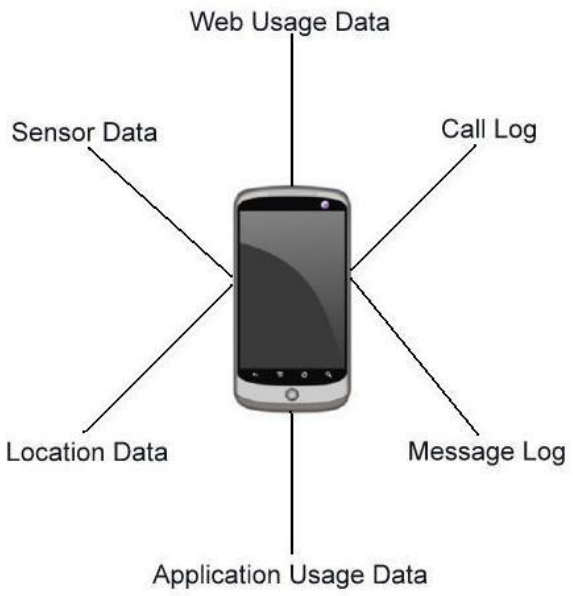

Figure 1: Types of Data Generated

Today, mobile phones have integrated a lot of features like multimedia, internet browsing, playing games, applications, etc. Thus, the use of a moblie phone is no longer limited to communication. As, the use of moblie phones has increased, it provides with large amount of data to extract knowledge from. Mobile data mining is an aproach, which analyzes the data collected from mobile phones and extracts useful knowledge from it. There are many ways in which data from mobile can be mined. This paper discusses some of such techniques.

Data can be collected from mobile phones in various forms. The data collected, can be from call logs, location, web usage, application usage, sensor data, etc. Figure 1 describes the different types of data, which can be generated from mobile phones. Using this data , various data mining activities can be performed. Some of these activities include Location based service and behvorial analysis.

II. MOBILE DATA MINING

Aquiring data from mobile phones and processing it is not an easy task. It undergoes various steps. These steps are as follows:

A. Gathering of Data

First, the data on which data mining is to be performed has to be gathered. Data can be gathered from different sources like call logs, message logs, application data, location data, etc. This data then undergoes ETL (Extraction, Transformation, and Loading) processing.

B. ETL Processing

The goal of the ETL tool is to extract the data, transform it into clean data and load it into Data Warehouse, where data mining activities are performed. The ETL consists of three steps.

In Extraction, the data from various sources is extracted and brought into a common format. In transformation, the data is transformed into a clean format by removing errors and ambiguities from the data. In loading, the transformed data is loaded into the Data Warehouse where the mining activites can take place.

C. Mobile Data Mining

After the data is loaded into the Data Warehouse, the mining activites like Behavorial analysis, location based services, etc can be performed.

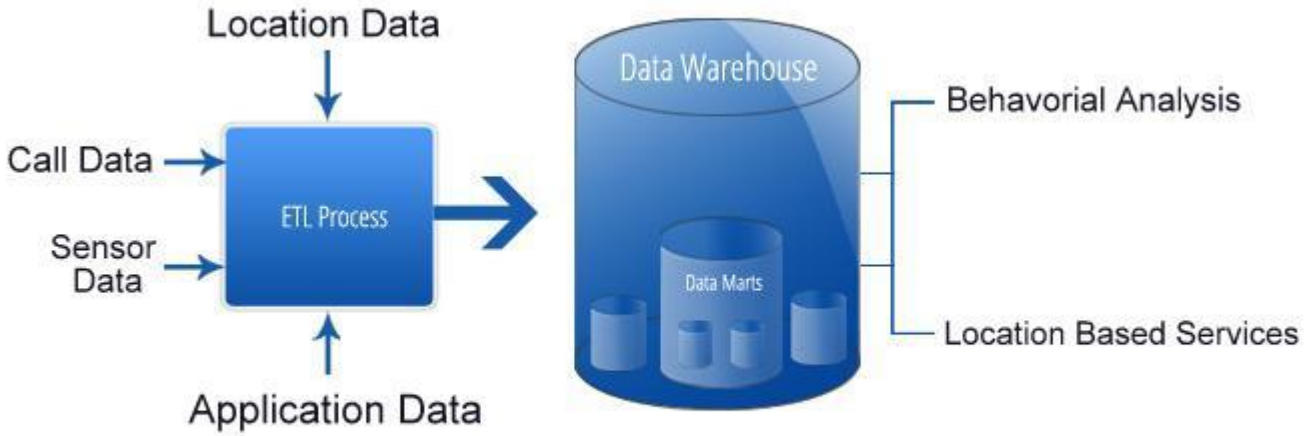

Figure 2: Mobile Data Mining Process

III. MOBILE DATA MINING SCOPE

Once the data is loaded into the Data Warehouse, it is ready for various mining activites to be performed. Behavorial Analysis, Location Analysis, etc can be performed. These analysis are now discussed in details. Figure 2 depicts the mobile data mining process.

A. Behavorial Analysis

Behavior analysis [1], analyze the behavior of a mobile user by considering various activities carried out by the user on the mobile phone and its social impact. Behavior analysis makes the mobile phone tailored for a person based on the interaction made by the person. Behavior of a mobile user may be the social interactions in a community or it may be the personality of the user or it may be an organizational behavior.

In behavior analysis, the user interaction with the mobile phones gives an overview of what a user is doing at a particular context, such as waiting in the mall, doing shopping or waiting for bus. The habits of mobile users are reflected by the context-awsreness of the user. Therefore, it is implied that the associations between user interaction records and the corresponding contexts can be used to characterize user habits. For example, the following associations can characterize the habits of a person:

- Given the context that a person is waiting for a taxi in the traffic and the surrounding environment is noisy, the person usually plays games.
- Given the context that a person is going to college in bus, the person usually listens to a particular song.
- Given the context that a person is on a vacation, the person usually shoots pictures.
- Given the context that a person's exam is coming, the person usually studies.

B. Location Analysis

Location prediction [2] provides a means of identifying the next location that a mobile user will be visiting based on the movement history of the mobile user. When the user's next location is predicted in advance it will lead to effective resource allocation with regards to the mobile operators. Location prediction also helps in environment where location based queries tends to come from the user of the mobile phone. If the next location is predicted the location based queries can be effectively satisfied.

Three main steps are taken in predicting the location of a mobile phone. In the first step the history of the mobile user movements are obtained and are organized in an effective manner. In the second step patterns are generated using an appropriate method from the mobile movement data. In the third step real time location information of a user is provided and the next location the user will be visiting is obtained based on the pattern generated in the second step.

Location Based Service (LBS) [3] are application specific services that are rendered by an application on the user's mobile phone based on the current location in which the user is active. LBS increase the revenue of an application vendor by customizing the application based on the usage of that particular application in a specific location. Some of the major LBS provided by a mobile application are Traffic Information, Restaurant Information, nearest Hospital and nearest Bank. The general architecture of the Service request and LBS Reply is depicted in Figure 3.

Using data mining techniques the application that provide LBS promote the appropriate information that a user needs at a particular location in advance. To provide effective LBS, the application vendor has to collect all the data regarding the application user such as the location of the user, information requested by the user etc. The request patterns are identified using an appropriate data mining method. When there is a

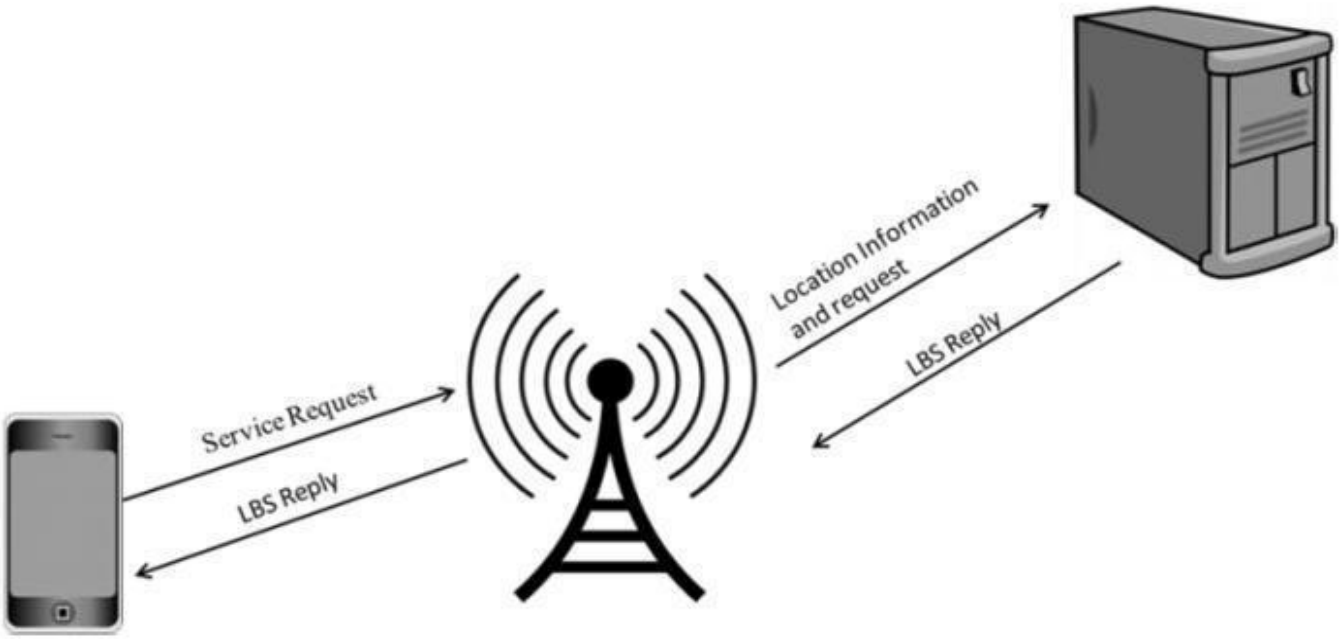

Figure 3: LBS Request and Reply

request by the application user the application server sends the appropriate LBS based on the pattern identified.

IV. MOBILE DATA MINING ACTIVITIES

A. Behavorial Analysis

Both context and interaction data are used to analyze the behavior of the user. Context logs collect the history context data and interaction records of mobile users, and thus can be used as data sources for mining behavior patterns. Specific surveys can also be conducted to obtain data if the context logs are limited.

Given a contextual feature set F= { f_1, f_2, ...,f_i}, a context C_i is a group of contextual feature-value pairs, i.e., C_i = {(a_1 : b_1), (b_2 : b_2), ..., (a_i : b_i)}, where $a_i \in F$ and b_n is the value for $a_n (1 \leq n)$. A context with contextual feature-value pairs is called a p-context.

Contexts may have different granularity with respect to the numbers of contextual feature-value pairs they contain. For example, {(Is a holiday?: Yes),(Time range: AM 8:00-10:00)} is a context with two contextual feature-value pairs, and {(Is a holiday?: Yes),(Time range: AM8:00-10:00),(Transportation: Train)} is a context with three contextual feature-value pairs so it expresses richer context information than the previous one.

An interaction record is an item in the interaction set I = {I_1, I_2...I_n}, where I_n ($1 \leq n$) denotes a kind of user interaction.

Interaction records capture the occurrences of user interactions with mobile devices, such as playing games, message session or clicking photos. Abstract user interactions are also possible to define, such as entertainment interactions, business interactions, to capture the high-level semantic information of user behaviors.

Association Rule Mining is used by behavorial analysis for finding interesting patterns from the context logs.

B. Location Based Service

Location Prediction and Location Based Services needs to use the spatio-temporal data [3] collected from the mobile phones. The spatio-temporal data contains the information of location and time as a pair. The spatio-temporal data is hard to mine as the search space for finding the knowledge is very high. The basic approach for mining this kind of data is to identify the frequent patterns in the data.

A Frequent Pattern [4] is defined over a database of sequences D, where each element of each sequence is a time-stamped set of items i.e., an *itemset*. Time-stamps determine the order of elements in the sequence.

The spatio-temporal sequence and mobile sequence transaction are mined using Association rule mining [5]. An Association Rule Mining is done on any transaction database. A transaction database (TDB) is a set of transactions, where each transaction, denoted as a tuple < T_{id}, X >, contains a set

of items (i.e., X) and is associated with a unique transaction identity (T_{id}). A transaction $<T_{id}, X>$ is said to contain itemset Y if $Y \in X$. The number of transactions in TDB containing itemset is called the support of itemset, denoted as S(Y). Given a minimum support threshold *min_sup* and a minimum confidence threshold *min_conf*, $A \Rightarrow B$ is an association rule if $S(A \cup B) \geq min\_sup$ and $(S(A \cup B)/S(A)) \geq min\_conf$, where *A*, *B* denote two non-overlapped itemsets, *A* is called the antecedent, and *B* is called the consequent.

For instance,

If,
A=Position at home,
S(A)=1,
B=Position at office,
S(B)=2,
Min_sup=2,
Min_conf=2,

Then,

$S(A \cup B)=1+2=3$,

Therefore,

$S(A \cup B)/S(A)=3$.

Since, $(S(A \cup B)/S(A)) \geq min\_conf$ and $S(A \cup B) \geq min\_sup$, $A \Rightarrow B$ is the association rule. Thus, the location is predicted to be 'B' i.e. office, under the given conditions.

Most of the traditional association rule mining algorithms divide the mining procedure into two stages. All frequent itemsets are found from the transaction database in the first stage. The rules are generated from the frequent itemsets and their confidences are calculated in the second stage

V. FUTURE WORK

In future, more types of mobile data mining will be explored. Main focus will be on different ways of gathering the data. We are developing ways of providing user specific functions without compromising the privacy of the user. Furthur, the mining activities like Intrusion Detection, etc. are to be incorporated.

VI. CONCLUSION

Mobile data mining is a fast growing area of data mining which gives importance to the mobile phone user. The data collected from the mobile phones give invaluable knowledge representing various aspects of the user thus enabling the vendor to customize mobile usage as per the user needs. In spite of the customization provided to the user, the privacy of the user is compromised. We have explored areas where Mobile Data Mining is applicable and have demonstrated ways how data can be mined in those areas. We look forward to explore more and more areas and succeed in the above-mentioned future works.

REFERENCES

[1] GokulChittaranjan, Jan Blom, DanielGatica-Perez. 2011.Mining large-scale smartphone data for personality studies. In Personal and Ubiquitous Computing.

[2] Le-Hung Tran, Michele Catasta, Luke K. McDowell, Karl Aberer. 2012. Next Place Prediction using Mobile Data. In Mobile Data Challenge 2012 (by Nokia), Newcastle.

[3] M Erwig, RH Gu, M Schneider, M Vazirgiannis, "Spatio-temporal data types: An approach to modeling and querying moving objects in databases", 1999, Springer

[4] Jingjing Wang, BhaskarPrabhala. 2012. Periodicity Based Next Place Prediction. In: Mobile Data Challenge 2012 (by Nokia), Newcastle.

[5] J Hipp, U Güntzer, G Nakhaeizadeh, "Algorithms for association rule mining—a general survey and comparison", 2000 - dl.acm.org

[6] Fosca Giannotti, Mirco Nanni, Fabio Pinelli , Dino Pedreschi. 2007. Trajectory pattern mining. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.

[7] RakeshAgrawal, RamakrishnanSrikant. 1995. Mining Sequential Patterns. In ICIDE, 3-14.