# **Data mining algorithms for mobile phones**

With the ubiquity of mobile phones in many countries, call data records (CDR) are becoming increasingly popular source to be used to discover the spatiotemporal behaviour of people in order to create easier ways of transportation for those who need it or to create certain facilities such as shops, schools or even hospitals. In some cases, CDR is also used help track criminals on the run or even track suspected criminals in order to prevent any future crimes. The common reason CDR is used in most of these cases instead of censuses or surveys is because of the great range covered, cost effectiveness (much cheaper), time effectiveness (much quicker) and reliability (some people may not be aware of how many places they travel to or how long they spend at each location). In this summary I will cover the various ways CDR is used in mobile mining for discovery and evaluation purposes, how they are implemented and the overall effectiveness of CDR and the methods used.

## Uses of mobile mining

There are two ways to collect human movement behavior through mobile mining [1]: the location system (i.e. radio frequency signals such as GPS, cellular networks and Wi-Fi) and the motion system (built in phone sensors such as the accelerometer, compass and magnetic sensors). These both have their uses in mining as collecting data from these two systems allows one to know the position of the user (via location systems), altitude and velocity (via motion systems) making gathered data more detailed and accurate.

Call data records are obtained through telecommunications, who themselves obtain the call data records by recording the cellular tower through which a call, text or multimedia text message was last sent or received for certain mobile number. From this the call length duration, longitude, latitude, the number interacted with, time, date and other information that can be obtained easily when required. The research conducted determines the appropriate information to take from these records before the experiment is carried out. A good example of this is in paper [2] where the name of the subscriber of the phone, the parties on the phone call, time and date of calls and call duration are used to prevent crime by tracking those who are suspicious of committing any future crimes. This was done as it reduces risk of human error and reduces the long term financial as well as time cost. Although Assisted Global Positioning Systems (AGPS) could be used for tracking, it isn´t always the best way to keep an eye on someone, as pointed out by paper [3]. This is due to many people having their AGPS turned off as it consumes a lot of power, on top of not many phones actually having AGPS. This paper aimed to Identify who were tourists but decided to do this through CDR mining, in which they pinpointed the location of a phone user by taking the longitude, latitude, caller id (derived from phone number) and a “tourist integer” calculated through an algorithm of theirs.

## Implementation of mobile mining into algorithms

Depending on what one wants to research, the CDR is handled differently. Algorithms that include CDR are used with varying quantities used from the database itself. When tracking suspect criminals, all of the provided information from the data is used in the algorithm as there are varying aspects that can define a criminal based various aspects set by the Anti-Crime team (such as call duration, last location, IMEI number, etc.) When the higher authority needs suspect information, they give a process and authority request to access CDR of suspects. From the global CDR, requested numbers are extracted from CDR. A file is then generated which contains requested CDR mobile numbers in a log format, with the generated file being sent to the respective department of anti-crime. The CDR file is then parsed and processed. The CDRs are then entered into database and analysis of the produced database is produced based on the aforementioned various aspects to find the most probable criminal. A result is then produced. This is just one algorithm involving the use of CDRs thoroughly as well in an effective way due to being able to produce a likely criminal based from their unique CDR.

## Evaluation

From previously read papers [4] ~ [17] mining the CDRs were useful in the respective experiments and theses, they were used in due to being able to obtain a large amount of information of a large number of mobile phone users (and effectively test subjects) in order to produce a reasonable conclusion from their behaviours. When comparing new systems to older systems, CDR methods are proven to be more:

* Technically efficient – time to produce output, calculated by space and time for system produce output;
* Economic efficient – takes into account the cost required to implement system and inefficiency is minimised given resources.
* Management efficient – how easily the new system can be maintained or handled.

Using my previous example of the anti-crime algorithm, the results produced compared to the traditional paper system, and system used in papers [18] and [19]. The proposed system proved to be more technically efficient, economic efficient and management efficient than the other three systems. This is also the case with the tourist experiment in which finding which phone users were locals of an area and those users that were tourists themselves. This helps countries in being able to efficiently manage their tourism departments, fiscal resources poured into tourism such as helping non-government tourism organisations/companies. (The algorithm used to produce the “Tourist integer” involved mining CDR for longitude and latitude floats in order to determine if they visited any pre-determined tourist locations.)

## Conclusion

In closing, the efficiency and effectiveness of mobile mining and CDRs are useful in discovering a variety of human related behaviours that can allow for proper management of planning and resources. The usage of mining in modern research can help drastically reduce the cost of these projects while also increasing the accuracy and scope of their results. These methods can be transferred over to other areas of mining as well, not just in mobile phones, but possibly in other areas of data mining such as in the world wide web. Although web mining is used in ways such as displaying appropriate advertisements, on a broader scope similar to mobile mining it could also be used to appropriately set up servers in locations that may not have been considered for programs that require low latency to function effectively (as a very basic example).

## References:

[1] Zhenzhen Wang et al, “Applying mobile phone data to travel behaviour research: A literature review”

[2] Er. Saiqa Khan et al, “Criminal Investigation Using Call Data Records (CDR) through Big Data Technology”

[3] Ratul Sikder et al, “An Efficient Approach of Identifying Tourist by Call Detail Record Analysis”

[4] Marcos R. Vieira et al, “Characterizing Dense Urban Areas from Mobile Phone-Call Data: Discovery and Social Dynamics”

[5] Peiyu Yang et al, “Identifying Significant Places Using Multi-day Call Detail Records”

[6] Mrs. Suja C Nai et al, “Impact of CDR Data Analysis Using Big Data Technologies for the Public: An Analysis”

[7] Feng Liu et al, “Annotating mobile phone location data with activity purposes using machine learning algorithms”

[8] Weijian Sun et al, “Characterizing User Mobility from the View of 4G Cellular Network”

[9] Biplav Srivastava et al, “Case Studies in Managing Traffic in a Developing Country with Privacy-Preserving Simulation as a Service”

[10] Feng Ling et al, “Mining Travel Behaviors of Tourists with Mobile Phone Data: A Case Study in Hainan”

[11] Xuzhao Wang et al, “Travel Distance Characteristics Analysis Using Call Detail Record Data”

[12] Yihong Yuan et al, “Correlating mobile phone usage and travel behavior – A case study of Harbin, China”

[13] Olle Järv et al, “Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records”

[14] Michel Daoud Yacoub et al, “Wireless Communications Track in Brazil”

[15] Qining LIN et al, “Mobile Customer Clustering Based On Call Detail Records For Marketing Campaigns”

[16] Mirco Musolesi, “Big Mobile Data Mining: Good or Evil?”

[17] Andrew Shields et al, “Application of multiple change point detection methods to large urban telecommunication networks”

[18] Raj Kumar Vishwakarma; Ravi Shankar, "Modeling Brain and Behavior of a Terrorist through Fuzzy logic and Ontology"

[19] Huiqi Zhang; Ram Dantu, "Predicting social ties in mobile phone networks"