JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Mobile Computing: Aggregated Human Mobility Patterns

S. John Grasias

Det. of Computer Applications, AJK College of Arts and Science, Bharathiar University, Coimbatore, India

Available online at: www.ijcseonline.org

Accepted: 20/Jul/2018, Published: 31/Jul/2018

Abstract- Location-based mobile applications are steadily gaining popularity across the world. These applications require information about user's current location to access different kind of services. This paper discover aggregated mobility and place visiting patterns of people in developing countries using one CDR (Call Detail Records) dataset collected in Ivory Coast and two finegrained location information datasets collected in India and Switzerland. We have compared these mobility patterns with existing studies for developed countries and found several differences. One of the difference is that people in developing countries are less likely to travel long distance on weekends as compared to developed countries. This paper tries to fill that gap and provide practical and promising solutions to enable location-based services on both feature phones and smartphones using low energy location interfaces.

Keywords:[Aggregated, Call Detail Records, Clustering, WIFI]

I. Introduction

With the availability of massive cellular data and mobile phone users, studying large scale mobility patterns have become easier and recently, there has been lot of research work on using cellular data to characterize human mobility. There are primarily two sources of location data collection using mobile phones, either using an application running on the mobile device or from the cellular network. On the mobile phone end, there are various location interfaces such as GPS, WiFi and GSM (Cell ID). These location interfaces provide different level of accuracy and availability. For instance, GPS provides fine-grained location of a person but do not work in indoor environments. In case of the cellular network, identifier of a cell tower (Cell ID) is collected as part of Call Detail Records (CDRs) when a phone connects to the network to make or receive a phone call, send or receive a SMS or use data connection. Both the primary sources of location data have their own trade-offs. CDRs collected from cellular network provide an opportunity to perform analysis to find large scale mobility patterns. Such large scale analysis can not be performed when data is collected from individual's mobile phones. However, there exist significant amount of flexibility while collecting location from individual mobile phones such as location sampling interval could be set to be high and there is less chance of missing a place unlike CDRs. Similarly, high sampling of location information helps in unveiling the place visiting patterns as well as accurately estimating place stay duration.To the best of our knowledge, there is no study which finds movement and place visiting patterns of mobile users in developing countries. We hypothesize that human mobility patterns in the developing countries could be different from those in the developed world due to many reasons such as quality of transportation, socio-economic status, and population density etc. In this chapter, we have performed a detailed characterization of human mobility in a developing country using a CDR dataset.

A mobile user is expected to visit several places in a given duration. Researchers have worked on finding important places in an individual's life using CDR data as well as fine-grained location data collected using individuals' mobile phones. Build algorithms to identify important places in a person's mobility history using CDR data. Using ground truth derived from few volunteers, they have found that a person's location such as "Home" and "Work" location can be estimated with an error of about 1 mile. Additionally, there have been significant research work on finding places from different location interfaces i.e. GPS, WiFi, and GSM (Cell ID) and combining them with other sensors such as accelerometer.

II. Proposed Work

2.1 Aggregated Movement Patterns

Analyze aggregated movement patterns of participants for all the three datasets. For fair comparison, we compare movement patterns of dataset with the earlier studies performed with CDR data and compare movement patterns from datasets with each other. One of the metric to measure human movement is daily range which represents the maximum distance traveled by a person in day. For instance, if a person visits locations {C1,C2,C3,....Ck} in a day then the daily range will be the maximum pair wise distance between these locations. Median daily range for a person represent the most frequent (regular) travel that she takes most of the days i.e. home to workplace or vice-versa, where as maximum daily range represents the occasional large distance trips that she undertakes. For instance, a study done in US found that a person is more likely to do a long distance travel on weekends. Different percentile values of median and maximum daily ranges on weekdays and weekends across all the participants of dataset. We observed a significant variance in people's mobility on weekends. Majority of users preferred to stay at home during the weekends, while a small percentage of users chose to travel large distances. Some of the main observations from our analysis are as follows: On weekdays, 50th percentile of median daily range was observed to be zero miles, which represents that more than half of participants were staying at the same place (Cell ID) mostly. Similarly, 25th percentile of maximum daily range is zero mile which represents that one fourth of total participants did not moved in the whole duration of data collection (i.e. 2 weeks). However on weekends, more than half of participants did not travel. Further, 75th percentile of maximum daily range is 27.43 miles on weekdays and 6.96 on weekends which shows that mobility range of many people remains limited on weekends as compared to weekdays. 2. Median daily range values remains same up to 75th percentile on weekdays and weekends, signifying that most of people's movement pattern remains the same across weekdays and weekends. Higher value of 95th percentile in weekend median daily range suggests that some people travel farther distances on weekends as compared to weekdays.A person visits different places in a day and it is feasible to automatically discover most of these places with the location information captured by mobile phones. While number of visited places are different for every person, we are interested in analyzing aggregated place visiting patterns across all three datasets. We have different modalities of location data i.e. CDR, GSM, WiFi, and GPS in our datasets. For each of these modalities, the place discovery algorithm should cluster Cell IDs or WiFi APs according to physical places. However, clustering approach for each modality is different and expected to have varying level of accuracy. Following is a brief description of clustering algorithms which are used to discover places:

2.1.1 CDR-based Clustering : The task of a place discovery algorithm is to cluster Cell IDs observed by a person according to different physical places. CDR data is very sparse and it is possible that many places, which a person visits may not get discovered using the CDRs data. Previous studies have shown that a user's phone may connect to different cell towers even if she stays at the same place. We implement the algorithm presentedfor clustering Cell IDs which uses Hartigans leader algorithm to cluster nearby Cell IDs with the help of a threshold distance (td). This algorithm takes into account all the Cell IDs observed by a person in the given time period.

2.1.2 WiFi-based clustering : To discover places using WiFi data, we need to cluster WiFi APs according to physical places visited by a person. The basic assumption in WiFi-based clustering is that a person will observe different set of WiFi APs in different places. We have used UIM clustering algorithm presented in which can cluster WiFi APs into a set of distinct places. UIM clustering algorithm do not take signal strength into account to avoid signal fluctuation problem and uses regularity in human movement to solve partial scan problem.

2.1.3 GSM-based clustering: Unlike CDRs data, Cell IDs are logged at a finegrained interval (nearly 1 minute) in a mobile phone and it is less likely to miss a place. However, clustering of Cell IDs. physical places has challenges such as Cell ID may change even when a user stays at the same place due to various reasons such as network load, small time signal fading, and internetwork (2G to 3G or vice versa) handoff. The change in Cell ID at the same place is called as "oscillating effect". To solve above challenges, we use Graph-based clustering algorithm (GCA) described. GCA models the oscillating effect among Cell IDs using an undirected weighted graph (movement graph) and then performs clustering with the help of heuristics such as edge weights and node degree. The evaluation results, on two diverse datasets show that GCA was able to correctly discover about 80% of places corrected as compared to the ground truth generated using GPS/WiFi.In case of place visiting patterns among participants in India and Switzerland, we have found some similarities such as similar median stay duration and Saturday being the most preferred day of visiting infrequent places. Except Saturday, people in India were more likely to visit infrequent places on Sunday where as in Switzerland, it was Friday. There was a significant difference in arrival time pattern for infrequent places i.e. most of visits to infrequent places were performed during afternoon in Switzerland where as most of the visits were performed during evening. The place visiting patterns are important for many applications such as advertising, recommendations, and pollution exposure estimation.

III. Experimental Results

Our experiments with real-data, we highlighted the tradeoff of using different kind of mobility data for finding places. For instance, GSM-based approaches merges many places which are nearby but were able to discover places, where WiFi is not available at a large scale. In CDRs data, number of regular places visited are also smaller in developing countries as compared to developed countries.

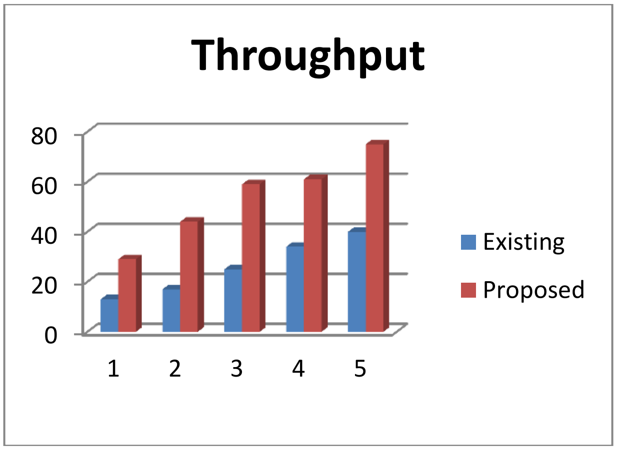

Figure 1: Throughput

Figure 1 represented into throughput values of compare with existing and proposed. Their proposed values higher than existing values. So their proposed throughput values are better result of this process.

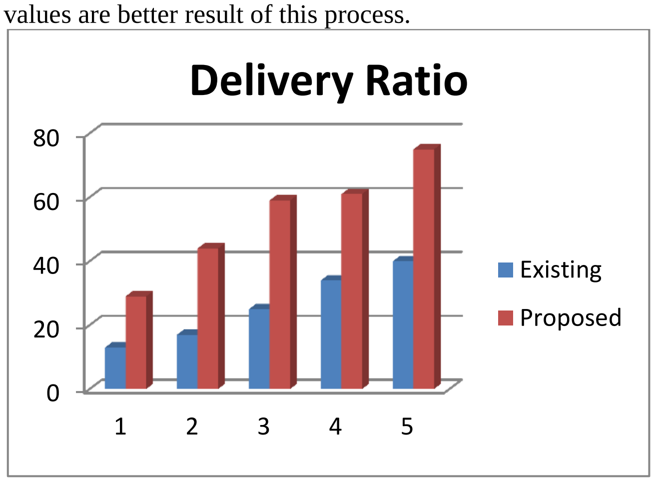

Figure 2: Delivery Ratio

Figure 2 represented into delivery ratio values of compare with existing and proposed. Their proposed values higher than existing values. So their proposed delivery ratio values are better result of this process.

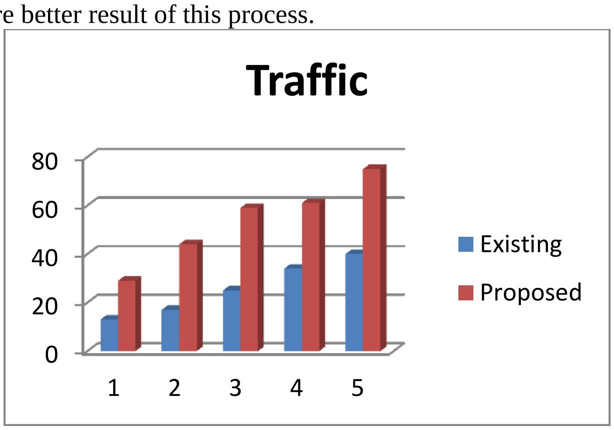

Figure 3: Traffic

Figure 3 represented into traffic values of compare with existing and proposed. Their proposed values higher than existing values. So their proposed traffic values are better result of this process.

IV. Conclusion

This research paper concludes proposed CBS-based approach that removes the necessity of war-driving or building a Cell ID database for GSM-based localization. This provides better results in throughput, PDR, Traffic. Evaluation using real-world traces showed that the proposed approach can provide reasonably good accuracy, which is sufficient for many location based services. We have developed several location-aware application using CBS-based localization technique and even built multi-modal techniques using Cell ID and GPS, which can minimize energy consumption on smart phones. Hence, CBS-based localization is a promising solution, especially for feature phones and provides mobile users in developing countries, an opportunity to access location based services without any extra infrastructure. Our system MobiShare facilitates local content sharing mechanisms by enabling scalable content search, encounter prediction, and notifying interested users, whenever they are in vicinity. Results from our deployment confirmed that use of MobiShare brings much more control, reliability, and resource-efficiency as compared to totally distributed architecture.

REFERENCES

[1] Latest Mobile Statistics. MobiThinking (online), 2011. http://mobithinking. com/mobile-marketing-tools/latest-mobile-stats.

[2] China 3G Mobile Subscribers Statistics, 2012. http://in.news.yahoo.com/ billion-mobile-phones-china-3g-penetration-low-042833846.html.

[3] Featurephones to smartphones ratio, 2012. http://mobithinking.com/ mobile-marketing-tools/latest-mobile-stats/.

[4] India 3G Mobile Subscribers Statistics, 2012. http://www.zdnet.com/in/ 3g-subscribers-form-2-percent-of-indias-mobile-users-7000004883/.

[5] MIT Technology Review Big Data Articlef, 2012. http: //www.technologyreview.com/featuredstory/513721/ big-data-from-cheap-phones/.

[6] Mobile and PC based Internet Users in China Statistics, 2012. http://www.reuters.com/article/2012/07/19/ us-china-internet-idUSBRE86I0FC20120719.

[7] Mobile only Internet User Statistics, 2012. http://mobithinking.com/ mobile-marketing-tools/latest-mobile-stats/b#mobile-only.

[8] Nokia Nearby Hyperlocal Search, 2012. http://www.youtube.com/watch?v= gsELVWr9v3E. [9] Opera Mobile Web Report, 2012. http://business.opera.com/smw/2012/ 09/.

[10] PlaceMap Mobile Application, 2012. https://play.google.com/store/ apps/details?id=com.iiitd.muc.placemap.

[11] WiFiShare, 2012. https://play.google.com/store/apps/details?id=com. iiitd.muc.wifishare. [12] WiFiShare Client, 2012. https://play.google.com/store/apps/details? id=com.iiitd.muc.wifi_shr.

