International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 02 | Feb-2018 www.irjet.net p-ISSN: 2395-0072

Mining Human Behavioural Patterns using Smart Phones – Procedures and Limitations

Shali Sara Abraham[1], Supriya L.P.[2]

[1]*Student, M.Tech Computer Science & Engineering.*
[2] *Assistant Professor Computer Science & Engineering.*
[1,2] *Sree Buddha College of Engineering, Ayathil, Elavumthitta, Pathanamthitta, Kerala, India.*

Abstract- *This paper introduces a group of ascendable algorithms to spot patterns of human daily behaviors. These patterns are extracted from the variable temporal knowledge that is collected from good phones. With the assistance of sensors that are out there on these devices, and have known frequent behavioral patterns with a temporal graininess, that has been galvanized by the method people section time into events. These patterns are useful to each end-users and third parties World Health Organization offer services supported this info. The quantifiability makes the analysis on resource unnatural and tiny devices like good watches possible. By analyzing the information on the device, the user has the management over the information, i.e., privacy, and also the network prices will be removed.*

Key Words: **frequent pattern mining, multivariate temporal data, human centric data.**

1. INTRODUCTION

1.1. SCALABLE HUMAN BEHAVIOURAL PATTERNS

The problem solving and systems administration skills of versatile and wearable gadgets create correct devices for deed and gathering knowledge concerning shopper exercises' (portable detecting). This has prompted a major extension of possibilities to review human conduct going from open transport route to prosperity. In addition, the approach of versatile and wearable gadgets empowers scientists to inconspicuously distinguish human conduct to a degree that wasn't beforehand conceivable. In any case, there's yet associate degree absence of wide acknowledgment of moveable police investigation applications in true settings.

There are various functions behind this crisscross amongst capability and acknowledgment. To start with, the constraint of assets associate degree an absence of accuracy within the gathered relevant info significantly may be a check regarding the battery life. Besides, the small size of sensors that are managing radio repeat, i.e., Bluetooth, Wi-Fi, and GPS, influences the character of their info (the littler the convenience, the less actual the information). The subsequent reason is that the locality of the telephone to the shopper on the grounds that these gadgets don't seem to be usually sent by their proprietors. In any case, good watches and wearable are body-mounted and later, the locality issue is a smaller amount of testing. Lastly, the operating framework confinements of cell phones, that expel foundation administrations once the hardware is below a considerable, load (with a selected finish goal to save lots of the battery life). Consequently, there's good info accumulation approach which will discover and record people's knowledge of misfortune or instability.

Existing works that bolster versatile info mining have offered exceptionally encouraging outcomes. Be that because it might, these reviews utilize specific instrumentation, that is understood for info quality among shoppers or break down info offline outside the convention. There's associate degree absence of elastic info mining techniques which will subsume the vulnerability. During this work, it gifts versatile algorithms that use associate degree assortment of sensors, e.g., Wi-Fi, area, and then forth that are accessible for the convenience. By utilizing gathered variable momentary info the calculations will acknowledge visit human behavioral examples (FBP) with an amount estimation (transient granularity), just like the human read of your time. The algorithms are tested, and their skilfulness, on 2certifiable datasets, and 2 very little devices, i.e., a Smartphone and smartwatch.

Identification of normal examples of human conduct has applications in a few areas, that modification from proposal frameworks to healthful services and transportation streamlining. As an example, a social welfare application will screen a client's physical movement routine. However, if there's associate degree adjustment in their schedules, that isn't perceived or detected by the shopper, (for example, unhappiness connected practices), then the framework will understand this and advise parental figures concerning the modification. Another utilization case is often transportation improvement. With a selected finish goal to the touch base at the prepare station on time, a framework will soak up the quality employee samples of a shopper and inform of the right time for departure the station. Then again, the flexibility (as way as plus efficiency) empowers on the convenience and online examination and on these lines evaluates each the system value and protection dangers of exchanging individual info to the cloud.

The consequences of the calculations are an arrangement of identified FBPs, which is a blend of time-stamped characteristic/esteem (sensor/information) with a confidence level. For example, {confidence: 60 percent; 15:00-16:00; call: #951603XXXX; SMS: #951603XXXX} is a client profile that incorporates one FBP. This case indicates two rehashed practices, which are (i) making or getting a call and (ii) sending or accepting an instant message to 951603XXXX. These two behaviours have been occurred 60 percent of the time, between 15:00 16:00 every day.

2. LITERATURE SURVEY

2.1. CLUSTERING OF PHYSICAL ACTIVITIES

2.1.1. DATA PRE-PROCESSING

The measuring system knowledge has been normalized employing a second-order forward-backward digital low-pass Butterworth filter, with a cut-off frequency of three cps. A slippery average window of 512 samples, with a five-hundredth overlap (256 samples), has additionally been applied to the information so the records are often reduced, while not losing info [2].

2.1.2. FEATURE EXTRACTION

Features have then been extracted from the data. This has been undertaken in 2 completely different modes, time and frequency. These 2 modes disagree as time domain analysis measures the signal over the amount of the recording. In distinction, frequency domain analysis depicts however the signal's energy is distributed over a variety of frequencies. As such, frequency domain techniques are extensively accustomed capture the repetitive nature of a detector signal. This repetition usually correlates to the periodic nature of a selected activity like walking or running. The advantage of frequency-related parameters is that these are less at risk of signal quality variations. Utilizing the mathematical operators quick Fourier rework (FFT) and Power Spectral Density (PSD) the raw signal has been born-again between these 2 modes. From the time domain, the mean, median, variance, root mean sq. (RMS), variance and correlation are calculated for the measuring system signals. From the center rate monitor, the mean of this signal has additionally been determined. From the frequency domain, energy, entropy, peak frequency and median frequency are calculated for the measuring system signals.

2.1.3. FEATURE SELECTION

Whilst a variety of options are generated, some is also redundant. During this instance, spatiality reduction, utilizing Principal element Analysis (PCA), has been performed to search out a segment of the foremost necessary options. The options are analyzed in terms of "blocks". As an example, all of the median feature vectors (ankle_median, chest_median, hand_median) have used PCA. Throughout successive iteration, all of the foundation mean sq.vectors is processed (ankle_rms, chest_rms, hand_rms).

This method has been perennial till all of the options are processed, and also the prime 2 elements with the most effective discriminated capabilities are chosen. Every feature within the bi-plot is diagrammatical as associate degree eigenvector and also the direction and length of the vector indicates, however, every variable contributes to the principal elements of the plot. The feature highest to the horizontal axis, of Table one, shows that toil of physical activities from the information assortment.

2.1.4. CLUSTERING

Clustering strategies are often divided into 2 main groups: hierarchal and partitioning. The hierarchal approach constructs the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion, whereas partitioning relocates instances by moving from one cluster to a different, ranging from associate degree initial partitioning. Following associate degree analysis of the literature, the techniques that are eligible for the analysis embody collective (hierarchical) and k-means (partitioning) algorithms.

Level of Physical exertion	Activity
Light	Lying Sitting Standing Ironing
Moderate	Descending Stairs Vacuum Cleaning Normal Walking
High	Running Ascending Stairs

Table 1. Exertion of Physical activities from the data collection.

In this instance, conniving the silhouette averages are often accustomed overcome this issue. This worth is employed as a measure of the standard of the ensuing clusters. The worth of k that has the biggest SA indicates the foremost applicable value to use. Fig. one describes the procedures of knowledge assortment strategies and agglomeration of physical activities.

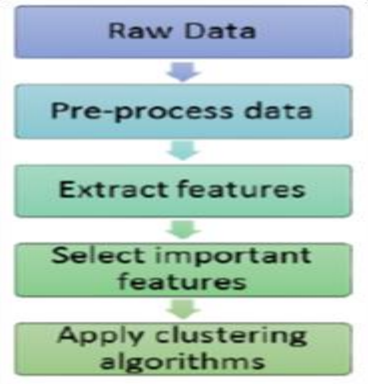

Fig 1. Data collection and clustering of physical activities.

2.2. LIMITATIONS OF MOBILE SENSING APPLICATIONS:

The rate of modification in mobile phones has been staggering [3]. Every new good phone unharnessed offers advances in sensing, computation, and communications. As a result, good phones represent the primary actually mobile present computing machine.

2.2.1. BEWELL: A MOBILE HEALTH APP

The BeWell app (for Androids) unceasingly tracks user behaviors on 3 distinct health dimensions while not requiring any user input—the user merely downloads the app and uses the phone as was common (see https://play.google.com/store/apps/details?id=org.bewellapp).

Classification algorithms run directly on the phone to mechanically infer the user's sleep length, physical activity, and social interaction. Additionally, to classifying activities that influence health, BeWell additionally computes a weighted score between zero and one hundred for every dimension. A score of one hundred indicates that the user is matching or prodigious suggested pointers (averaging eight hours of sleep per day, for example).

BeWell will run a complete mode on the phone or will interwork with the cloud to store longitudinal knowledge patterns. BeWell promotes improved behavioral patterns via persuasive feedback as a part of associate degree associate degree imaged aquatic system rendered as a close show on the good phone's wallpaper screen. The speed of the big orange clownfish mirrors the user's activity, whereas the quantity of little blue fish reflects the user's level of social interaction with others. Finally, the ocean's close lighting conditions indicate the user's sleep length the previous night. Users will passively look at the image of health dimensions and replicate on however they're doing.

Many of the challenges of building BeWell associated with developing low-energy sensing capabilities, feature engineering, and also the correct classification of health dimensions while not limiting the phone's battery period or usability.

These psychological feature eventualities align with several of the open challenges in AI. an everlasting problem AI researcher's face is determining a way to create systems additional versatile, adaptable, and protractile.

Similarly, psychological feature phones can request showing intelligence mix info from completely different sources, not by generic knowledge pooling however by investment famed relationships between human behavior at the cluster and individual levels. The phone would need a reasoning framework that considers multiple objectives and makes differing types of selections supported user desires like whether or not to intervene (in the case of a patient relapse), supply a suggestion (perhaps reorganizing the user's calendar supported measured stressors), or taking action (such as ordering and paying for a cafe latte in advance).

2.3. BATTERY MANAGEMENT

Polling a device's state will cut back battery life [4]. The robot API is event driven; thus gathering the information had a negligible impact on regular battery life. By programming a Broadcast Receiver connected to associate degree robot Service running in the background, whenever the robot OS broadcasts ACTION_BATTERY_CHANGED, the subsequent battery info was recorded: battery level, battery scale (maximum level value), battery share, battery technology (i.e. Li-ion), health rating of the battery, whether or not the phone was blocked to AC/USB, whether or not the phone is charging, temperature, voltage, time period and usage time period, battery standing (charging, discharging, full and not charging) and phone events associated with battery (fully charged and user simply unplugged, charging, finished charging, running on battery, unplugged once not totally charged). Table a pair of describes the distribution of battery life in several mobile applications.

Platform	Distribution
HTC	44.6%
Sony Ericsson	29.8%
Motorola	14.8%
Samsung	7.5%

Table 2. Distribution of battery life in different mobile applications

Battery management needs user intervention in 2 respects: to stay track of the battery out there so users will decide a way to prioritize amongst the tasks the device will perform; and to physically plug the device to the charger and surrender its quality. Demonstrate systematic however from time to time erratic charging behavior (mostly thanks to the actual fact that charging takes place once the phones are connected to a PC); principally opt to interrupt phones' charging cycle so reducing battery life. Aim to stay battery levels higher than half-hour thanks to associate degree automatic close notification, and systematically overcharge phones (especially throughout the night). Table three describes quality and API level of various Platforms in Android.

Platform	API Level	Popularity (Source: Google)	Popularity (Source: Study)
Android 1.5	3	12.0%	-
Android 1.6	4	17.5%	36%
Android 2.1	7	41.7%	33%
Android 2.2	8	28.7%	31%

Table 3 Popularity and API level of different Platforms in Android.

2.4 PREDICT USERS' PHONE PROXIMITY

2.4.1. ARM + ROOM METHOD

Predict users' phone proximity victimisation info that's already out there on the phone, instead of victimisation the additional Bluetooth tag [5]. for every subject, it use the Bluetooth tag proximity info as ground truth, and plan to predict whether or not the phone was inside arm's reach, or inside arm + space. Victimisation the discourse info it collected with the AWARE framework. If the predictions are correct, application developers will use the prediction models to work out once it will use the phone to gather discourse info from phone homeowners (arm's reach) or to gather discourse info concerning the owner's atmosphere and deliver info to (arm + room).

It created models that might classify phone proximity. It used a call tree classifier victimization the ID3 formula thus it may interpret the ensuing trees and verify that options were most significant to the classification task. options close to the foundation of call trees sometimes have high prognosticative power and may be treated as necessary options.

2.4.2. MOBILE PHONE PROXIMITY LOGS

One of the key concepts is to take advantage of the actual fact that trendy phones use each a short-range RF network (e.g., Bluetooth) and a long-range RF network (e.g., GSM), which the 2 networks will augment one another for location and activity logical thinking. the thought of work cell tower ID to work out approximate location are acquainted with readers, however, the thought of work Bluetooth devices is comparatively recent and provides different types of data.

Bluetooth may be a wireless protocol within a pair of.40–2.48 rate varies, developed by Ericsson in 1994 and free in 1998 as a serial-cable replacement to attach completely different devices. Though market adoption has been at the start slow, in step with trade analysis estimates by 2006 ninetieth of PDAs, eightieth of laptops, and seventy-fifths of mobile phones are shipped with Bluetooth. Each Bluetooth device is capable of "device-discovery," that permits to gather info on alternative Bluetooth devices inside 5–10 m. This info includes the Bluetooth Macintosh address (BTID), device name, and device kind. The BTID may be a 12-digit hex variety distinctive to the actual device. The device name is often set at the user's discretion; e.g., "Tony's Nokia." Finally, the device kind may be a set of 3integers that correspond to the device discovered; e.g., Nokia itinerant or IBM laptop computer.

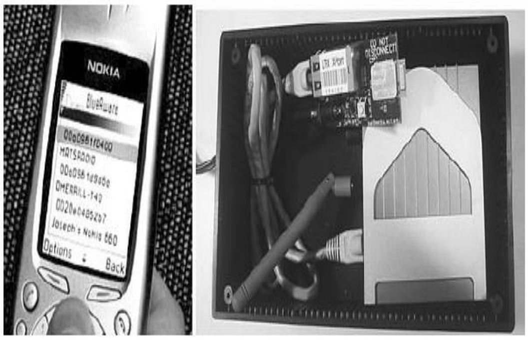

Fig 2. Bluedar, a Bluetooth detecting device.

2.5. DATA COLLECTION METHOD IN SMART PHONE

The data is collected using smartphones equipped with special software aiming to make the data collection invisible to the participants while optimizing the ratio between data collected and power consumed [7]. The database can then be accessed to perform analysis on the (anonymized) data and to visualize it to the participants. The different data modalities collected with the client can be categorized as follows:

2.5.1. SOCIAL INTERACTION DATA

Social interaction is inferred from call logs, short message logs and Bluetooth scanning results. In addition, it can use information from acoustic environment samples to detect the devices sharing the same acoustic space at any given time. Together these parameters can reveal the events where the persons in question are interacting with each other.

2.5.2. LOCATION DATA

Location is determined based on GPS (when available), cellular network information, and WLAN access point information (when available).

2.5.3. MEDIA CREATION AND USAGE DATA

Information is captured concerning locations where images have been captured, video shot or music played.

2.5.4. BEHAVIORAL DATA

Information is received concerning application usage, activity detection based on acceleration sensor, and regular device usage statistics based on call and short message logs. The locations visited and transportation means can be derived from location data. This data can be complemented with the help of questionnaires administered for the participants.

3. CONCLUSION

In this work, it planned an ascendable approach for daily behavioural pattern mining from multiple detector info. This work has been benefited from artificial datasets and users World Health Organization use completely different Smartphone brands. It uses a completely unique temporal graininess transformation formula that produces changes on timestamps to mirror the human perception of your time. Moreover, its approach is light-weight enough that it is often run on little devices, like good watches, and so reduces the network and privacy value of causing knowledge to the cloud. Moreover, changing raw timestamps to temporal granularities increase the accuracy of the FBP identification that is influenced by completely different values of temporal graininess, the section of the day and also the detector kind. These findings assist the system in characteristic the acceptable runtime and detector impact of the behavioral pattern identification.

REFERENCES

[1] Reza Rawassizadeh, Elaheh Momeni, Chelsea Dobbins, Joobin Gharibshah, and Michael Pazzani, “Scalable Daily Human Behavioral Pattern Mining from Multivariate Temporal Data,” IEEE Trans. on Knowledge and Data Engg, vol. 28, no. 11, Nov. 2016

[2] R. Dobbins and R. Rawassizadeh, “Clustering of physical activities for quantified self and mHealth applications,” in Proc. IEEE Int. Conf. Ubiquitous Comput. Commun., 2015, pp. 1423–1428.

[3] A. Campbell and T. Choudhury, “From smart to cognitive phones,” IEEE Pervasive Comput., vol. 11, no. 3, pp. 7–11, Jul. 2012.

[4] D. Ferreira, A. K. Dey, and V. Kostakos, “Understanding human smart phone concerns: A study of battery life,” in Proc. 9th Int. Conf. Pervasive Comput. 2011, pp. 19–33.

[5] A. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos, “Getting closer: An empirical investigation of the proximity of user to their smart phones,” in Proc. 13th Int. Conf. Ubiquitous Comput., 2011, pp. 163–172.

[6] N. Eagle and A. Pentland, “Reality mining: Sensing complex social systems,” Pers. Ubiquitous Comput., vol. 10, no. 4, pp. 255–268, 2006.

[7] N. Kiukkonen, J. Blom, O. Dousse, G.-P. Daniel, and J. Laurila, “Towards rich mobile phone Datasets: Lausanne data collection campaign,” Proc. ACM Int. Conf. Pervasive Services, 2010.