IMSP: an Intelligent Mobile Sensing Platform for Automatized Abnormal Behavior Detection

Kobiljon Toshnazarov
Computer Science and Engineering
Inha University
Incheon, South Korea
kobiljon@nsl.inha.ac.kr

Youngtae Noh
Computer Science and Engineering
Inha University
Incheon, South Korea
ytnoh@nsl.inha.ac.kr

***Abstract*—In recent years, mobile sensing data collection has become crucial in Human-Computer Interaction research. When conducting mobile sensing studies using smartphones and wearable devices, researchers mostly address issues related to Data Quality (i.e., missing data, low accuracy, etc.). However, they do not address a problem of deviations from behavioral contexts, that can affect research findings and conclusions if not monitored efficiently in a timely manner. So far, no study has addressed this issue on an individual level. Therefore, this paper proposes Intelligent Mobile Sensing Platform (IMSP), a mobile sensing data collection platform that can efficiently detect abnormal participant behavior from mobile sensors utilizing Density-Based Local Outlier Factor and informs a researcher about it in real-time. Our study shows a promising result with our preliminary study conducted in a strictly controlled laboratory environment.**

***Index Terms*—Human-Computer Interaction, Behavior, Mobile, Streaming Sensor Data, Anomaly Detection**

I. Introduction

Collecting high-quality data in data collection campaigns can play a major role in making a confident research outcome. Thus, researchers try their best to keep the Data Quality [1] as high as possible during data collection campaigns. However, except Data Quality, there is one more crucial factor that can affect research outcomes, which is a collection of misleading data generated by abnormal behavior of participants. Currently, researchers that conducted mobile experimental studies (i.e., [2]–[4]) and currently existing mobile sensing data collection platforms (i.e., PurpleRobot [5], AWARE [6], mCerebrum [7], etc.) mostly limited themselves with monitoring data completeness (a dimension of Data Quality [1]). Therefore, they do not detect and properly handle deviations from individual behavioral contexts (abnormal behaviors). For example, let us imagine a GPS data collection study about a student's school commuting. Participant's data can be influenced by factors that are external to a study, and the issue can go unnoticed by a researcher. For instance, the participant could go for a trip during the school commuting study. This may lead to a collection of data about the GPS data that must not be included in the study as it doesn't fit the regular school life. This kind of a problem exhibits properties of an individual's behavior, rather than Data Quality, and we refer to it as *unexpected deviations from behavioral contexts*, and current works don't address this issue. Therefore, this paper proposes IMSP – a mobile sensing data collection platform that utilizes the concept of Density-Based Local Outlier Factor (LOF) [8] for detecting those deviations.

LOF is a score that represents how likely a certain data point to be an outlier (abnormal data) relative to a certain dataset. In other words, it presents how different a particular sensor data sample is from a participant's overall data. LOF of a point is a local density around the point compared to the local densities of its closest neighbors. If the density of a point is much smaller than the densities of its neighbors, it means the point is far from dense areas (clusters), hence, it can be considered as an outlier. Our platform fully utilizes this observation in detection of outliers in the sensor data that can represent a person's behavior.

Our platform tackles the issue that none of the existing platforms has addressed before. Thus, this work's main contribution is that our platform automatically detects participants' abnormal behaviors and alerts a researcher in a timely manner. In other words, IMSP makes it possible for researchers to effortlessly notice collection of misleading sensor data in real time, without any overhead of manual data processing. We also conducted a strictly controlled preliminary study, where we collected heart rate of a subject with a sedentary lifestyle, triggered deviations from behavioral contexts, and detecting the deviations using IMSP. Our preliminary study proved the effectiveness of IMSP in practice.

II. Background

The need for a platform that can detect the deviations from behavioral contexts is valid. One of the examples can be the case study [2] about personality traits. The goal of the study is to find out if personalities of people from different countries (cultural backgrounds) can be generalized. In a more technical language, the goal is to see how well a machine learning model can perform if it is trained on sensing data that represent personality from people that live in several different countries, and tested on people that live in a country that was not in the training data. The study collects data across 5 countries during 3 weeks, which covers all 4 seasons overall. It is also a common knowledge that seasonal effects have various effects on behavior [9] (i.e., changing circadian rhythm, mental well-being, etc.). In other words, seasonal effects can influence our behavior, and behavior is linked with our personality as *personality traits* are *"stable" behavioral characteristics*. The study considers the issue on a sample level, and not on a timely manner. However, in order to confidently state whether personality traits are generalizable across cultural contexts or not, these kinds of studies need to efficiently detect each participants' deviations from their individual behavioral contexts in a timely manner, to find out if any external factor like seasonal effect have affected the data being collected. If observed on time, these external factors trigger abnormal behavioral data relative to subject's normal behavior. A technique that can be exploited in this case is Local Outlier Factor (LOF) [8]. LOF can detect local outliers in behavioral sensor data (generated by abnormal behavior) on an individual level, and help researchers identify misleading sensor data to produce more accurate, and confident research outcomes.

A. Density-Based Local Outlier Factor

Density-Based LOF is well-known to be effective for detecting outliers in a skewed dataset, which presents various distributions of its data clusters. The score shows how high the local density around the specified point is compared to its neighbors' local densities, meaning how far away is the data point from a dense area (data cluster). If the LOF score of a point is close to 1, then the point is an inlier (as points in a cluster share almost the same local densities), while a high LOF score ($LOF \gg 1$) means the specified data point is an outlier.

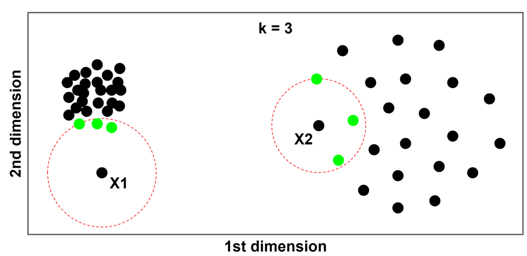

Fig. 1. LOF calculation based on local densities

LOF's calculation process is straightforward. First, we need to make a few definitions:

- $d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$, which is the Euclidean distance between points p and q, and n is the dimension of the points.
- $N_k(p)$ is a set of closest k neighbor points relative to the p point
- $d_k(p)$ is a distance between the point p and its k^{th} nearest neighbor.
- $reach\_dist_k(p,q) = max\{d(p,q), d_k(q)\}$, which is a reachability distance of q from point p, where at least k points must be within the reachability distance.
- $lrd_k(p) = (\frac{1}{k}\sum_{q \in N_k(p)} reach\_dist_k(p,q))^{-1}$, where lrd stands for "local reachability density" that defines how dense the local neighborhood of the point p is.

Finally, we can calculate the LOF of a point p as:

$$LOF_k(p) = \frac{1}{k}\sum_{q \in N_k(p)} \frac{lrd_k(q)}{lrd_k(p)}$$

It is clear from the Fig. 1 that point X_1 is an outlier due to its low density compared to its closest neighbors' densities ($LOF_k(X_1) \gg 1$), and X_2 is an inlier due to having approximately the same density as its neighbors ($LOF_k(X_2) \approx 1$). In the same fashion, our platform performs LOF calculations on behavioral sensor data and alerts a researcher when the LOF score is abnormally high. The further section explains how IMSP works in details.

III. Intelligent Mobile Sensing Platform

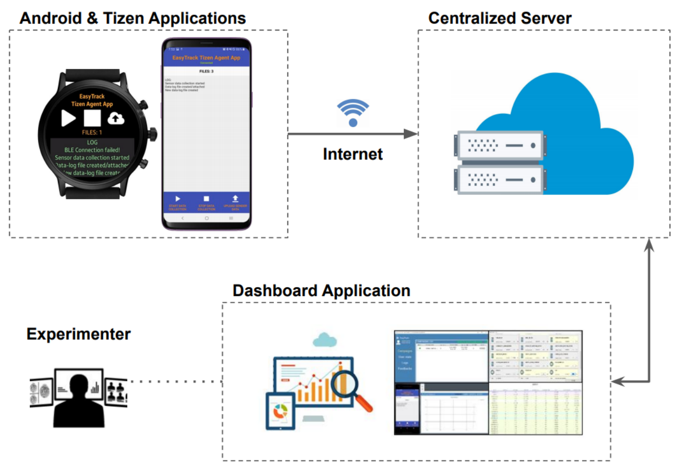

Fig. 2. Abnormal Behavior Detection platform architecture

The IMSP consists of thee parts as shown in the Fig. 2, which are: *Mobile & Wearable Applications*[1,2] for participants to collect and provide sensor data to a Centralized Server. *Centralized Server*[3] is the core part of the platform that gathers sensor data from participants and stores in a database, after which it performs operations needed for abnormal behavior detection; and *Dashboard Application*[4] is where the researcher, who conducts a mobile sensing study, is notified when detected outlying behavioral sensor data from behavioral contexts.

A. Mobile & Wearable Applications

The Mobile & Wearable Applications work as background services that collect sensor data (i.e., accelerometer, heart rate, light, etc.) into locally stored files and provide them to the Central Server as soon as the devices are connected to the Internet using WiFi. There are two basic steps for a participant to start sensor data collecting, which are: logging in, and starting the data collection with a button click. The rest (running sensor data collection, uploading sensor data, etc.) is automatically performed by the applications in background

[1]https://github.com/Qobiljon/EasyTrack_AndroidAgent
[2]https://github.com/Qobiljon/EasyTrack_TizenAgent
[3]https://github.com/Qobiljon/EasyTrack_Server
[4]https://github.com/Qobiljon/EasyTrack_Dashboard

services. The applications are for Android OS and Tizen Operating systems respectively, developed in Java and C# programming languages, and no external libraries were used for the data collection applications. The job of the Mobile & Wearable Applications finishes here, and the further work is done by the Centralized Server.

B. Centralized server

Centralized Server is the core part of IMSP, which gathers participant data, processes it, and calculates LOF scores for data points. Participant's sensor data goes through the following four procedures in the Centralized Server.

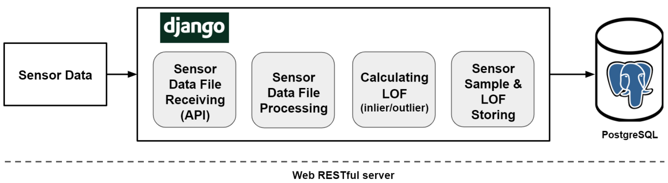

Fig. 3. Data processing flow in Centralized Server

The 1st step, which we refer to as the *Sensor Data File Receiving* step is: receiving and storing a sensor data file using an API (as in Fig. 3) when a participant P submits one. So as not to make a client wait for the server to process the whole sensor data after it is received, server immediately responds to a client with a success message after storing the file locally, which releases the client's HTTP request. After that we move to the 2nd step called the *Sensor Data File Processing* step. It is an always-running background service that continuously waits for new files that arrive from clients. Each row of the sensor data files represents a single data point/sample. This step reads each data point p, passing it to the next step for further processing. We call the 3rd step as *Calculating LOF* step. In this step we run LOF algorithm [8] for the target point p, where the dataset provided to the LOF algorithm is a set of P's previously submitted data points d_P where d is set of all data points in the PostgreSQL database, and d_P is acquired by filtering P's data points from d. After running the algorithm, the final *Sensor Sample & LOF Storing* step stores the resulting LOF value with the target point p in the PosgreSQL database. The Centralized Server is a RESTful Web service developed using a Django [10] Web framework, that uses PostgreSQL [11] as a database management system.

After the 4th step we have the LOF values stored in the database, and we already know that a high LOF ($LOF \gg 1$) means the data point is an outlier. The rest of the work is handled by our Dashboard Application.

C. Dashboard Application

Dashboard Application serves as a GUI application for visualizing data collection status, detecting abnormal behavior using LOF values loaded from the Centralized Server, and alerting a researcher when required.

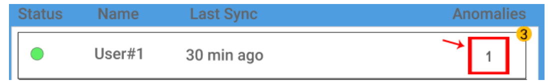

Fig. 4. Presenting a detected abnormal behavior on our Dashboard Application

A researcher needs to perform 3 basic steps to be able to detect participants' abnormal behaviors using the IMSP's Dashboard Application. The three steps are: (1) authentication (logging into the platform); (2) campaign creation (picking sensors, picking participants, specifying study duration, specifying LOF threshold, etc.); and (3) monitoring dashboard's LOF visualization. The Dashboard Application automatically loads and tracks LOF scores calculated on the Centralized Server, after which it alerts the researcher when a specified LOF threshold is exceeded by any user. As an example, Fig. 4 shows that a participant "User#1" has once exceeded a given LOF threshold by behaving abnormally, and a researcher is capable of monitoring it on using the Dashboard Application.

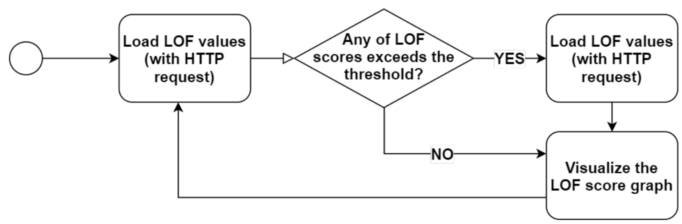

Fig. 5. Logic behind the Dashboard Application

If we look at the dashboard from a technical perspective as in Fig. 5., there is an always running background service, that keeps requesting calculated LOF values from Centralized Server, and whenever any LOF value exceeds a specified threshold, application alerts an experimenter with a simple alert window, and presents a graph of overall LOF values for the particular participant. Since a researcher has a control over the threshold, they can set how sensitive the dashboard should be to fluctuations of participants' behaviors ($threshold \approxeq 1$ means very sensitive to changes, and $threshold \gg 1$ means non-sensitive, as a very high LOF value can be only achieved if the target data point is significantly far away from dense areas).

It is the first work in the Human-Computer Interaction research to detect unexpected deviations from behavioral contexts in mobile sensing studies. Thus, IMSP creates an opportunity for researchers to detect misleading data, which was not considered by current works. Finally, we prove the effectiveness of the platform by discussing about our preliminary study in the next section.

IV. Preliminary Evaluation Study

In order to test the platform on detection of deviations from behavioral contexts, we conducted a preliminary study in a strictly controlled laboratory environment. Our goal was to observe LOF fluctuations in cases of abnormal behavior and validate that IMSP can detect abnormal behavior. Further, we

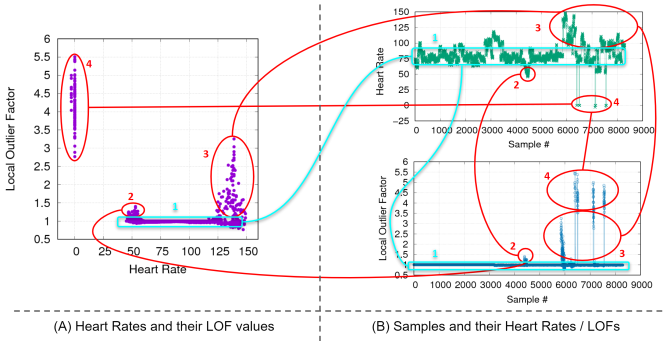

Fig. 6. LOF fluctuations based on heart rate values of the subject

explain our preliminary study's setup and finally evaluate its outcomes. For this study, we only considered a single subject (participant).

A. Setup

To achieve a valid research outcome, we had to cherrypick a sensor that could represent a subject's behavior, and that has values that are easy to understand. So, we ended up with Heart Rate Monitor as a heart rate is easy to understand at a glance, and it is well-utilized in Positive Computing research for studying behavior [12]. We considered a participant with a sedentary lifestyle and collected heart rate values approximately for 12 hours (10 AM–10 PM). The frequency of sampling was 0.2 Hz (sampling once every 5 seconds), collecting 8296 heart rate samples by the end of the day (dataset is available[5]). We considered 4 different cases, labeling the cases with numbers from 1 to 4, where the heart rate data represented:

- 1 - subject behaving normally / collection of subject's behavioral context (during the first 5 hours of the study)
- 2 - smartwatch being temporarily worn by a different person (within the the 6^{th} hour of the study)
- 3 - subject deviating from their normal behavior by performing physical activity, for example: running, jumping, doing squats in order to cause abnormal heart rates (between 8^{th} and 10^{th} hours of the study)
- 4 - subject wearing the smartwatch device loose (between 9^{th} and 11^{th} hours of the study)

We needed to first collect initial data for a participant for IMSP to know the subject's behavioral context. Then we needed to test if IMSP could detect abnormal heart rates coming from a different subject that would temporarily wear the smartwatch. After which, we needed to test if abnormal behavior (extra physical activities) would be detected. Additionally, we planned to test if IMSP would detect possible zero values generated by wearing the smartwatch loose.

B. Evaluation

The Fig. 6(A) shows the LOF scores of subjects heart rate values, and the Fig. 6(B) shows the sampled heart rate / LOF values in the sampling order.

[5]https://github.com/Qobiljon/ABD_Observation_HR_Dataset

We wanted the platform to be highly sensitive to abnormal behavioral data in our preliminary study, and it fulfilled our expectations. As we can see from the Fig. 6, IMSP was precise on detection of the deviations from behavioral contexts whenever a subject behaved abnormally during the preliminary study and provided on-time alerts to a researcher. Therefore, IMSP is highly effective in practice and can be used by researchers to collect clean and useful data that are not affected by external factor rather than the study's goal.

V. Conclusion

Detection of deviations from behavioral contexts leads to collection of clean, and useful sensor data. We proposed IMSP, a platform that detects those deviations using LOF, and alerts a researcher when neecessary. Results from our preliminary evaluation study proved IMSP to be effective in practice. Therefore, IMSP can serve as a foundation for making valid, and confident research outcomes in Mobile Sensing studies by detecting deviations from behavioral contexts and alerting a researcher in real-time.

References

[1] Wikipedia contributors, "Data quality — Wikipedia, the free encyclopedia," 2019, [Online; accessed 2-January-2020]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Data_quality&oldid=930823029

[2] M. Khwaja, S. Vaid, S. Zannone, G. Harari, A. Faisal, and A. Matic, "Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–24, 09 2019.

[3] N. Lathia, G. Sandstrom, C. Mascolo, and P. Rentfrow, "Happier people live more active lives: Using smartphones to link happiness and physical activity," *PLOS ONE*, vol. 12, p. e0160589, 01 2017.

[4] G. M. Harari, S. D. Gosling, R. Wang, F. Chen, Z. Chen, and A. T. Campbell, "Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods," *Computers in Human Behavior*, vol. 67, pp. 129 – 138, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563216307282

[5] "Purple Robot," https://tech.cbits.northwestern.edu/purple-robot/, [Online; accessed 29-December-2019].

[6] D. Ferreira, V. Kostakos, and A. Dey, "Aware: Mobile context instrumentation framework," *Frontiers in ICT*, vol. 2, 05 2015.

[7] S. M. Hossain, T. Hnat, N. Saleheen, N. J. Nasrin, J. Noor, B.-J. Ho, T. Condie, M. Srivastava, and S. Kumar, "Mcerebrum: A mobile sensing software platform for development and validation of digital biomarkers and interventions," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3131672.3131694

[8] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '00. New York, NY, USA: Association for Computing Machinery, 2000, p. 93–104. [Online]. Available: https://doi.org/10.1145/342009.335388

[9] R. Foster and T. Roenneberg, "Human responses to the geophysical daily, annual and lunar cycles," *Current biology : CB*, vol. 18, pp. R784–R794, 10 2008.

[10] "The web framework for perfectionists with deadlines." https://www.djangoproject.com/, [Online; accessed 29-December-2019].

[11] "PostgreSQL: The World's Most Advanced Open Source Relational Database," https://www.postgresql.org/, [Online; accessed 29-December-2019].

[12] A. Waytz, "Psychology beyond the Brain," https://www.scientificamerican.com/article/the-neuroscience-of-heart/, 2010, [Online; accessed 29-December-2019].