Anomaly Detection Techniques in Mobile App Usage Data among Older Adults

Athanasios I. Kyritsis, Michel Deriaz, and Dimitri Konstantas

Information Science Institute, GSEM/CUI

University of Geneva

Geneva, Switzerland

Email: $\{athanasios.kyritsis, michel.deriaz, dimitri.konstantas\}$@unige.ch

Abstract—We are living in an era of demographic ageing, and new technologies that support independent living are constantly being created. In this context, more and more mobile applications are developed for this target group. In this paper, we are presenting a multidimensional application that targets older adults. We are monitoring the usage of all different aspects of the app, the amount of daily activity in the form of daily steps and the resting time throughout the day from a connected bracelet the user is wearing. Data amounting to 402 user-days of 6 different users are collected. A set of different datasets are manufactured, and various anomaly detection techniques are employed to identify the abnormalities in the datasets. The results demonstrate that clustering can be of use to detect anomalies in the older adults' patterns that could be the trigger of appropriate actions, like informing family members or professional caregivers.

Keywords—Abnormality detection, ambient assisted living, mobile applications, senior citizens.

I. INTRODUCTION

After the massive adoption of the Internet, new forms of technology have been developed that have an impact on almost every aspect of daily life, including health. The term eHealth was created to refer to technologies and applications in the service of health and wellbeing [1]. With the latest advances in mobile communications and with the widespread use of smartphones and connected devices, there has been a high number of health-related mobile applications. These applications can focus on specific medical conditions [2], enable doctors to provide their professional services at a distance [3] and target to promote behaviour change for health improvements and disease management [4].

We are living in the ubiquitous computing era where connected devices form the internet of things and produce data faster than we can logically process. The plethora of sensors that every modern smartphone includes, along with the advances in telecommunication technologies, enable the creation of context-aware applications [5] that power the internet of things. Tracking various aspects of wellbeing with mobile applications has become a habit. Numerous consumer electronic devices can continuously monitor users and can assist in healthcare services [6].

Outlier detection, also known as anomaly detection is a broad domain that has many applications in different fields. Such applications include abnormal behaviour detection in network traffic [7], fraud detection in credit card usage [8], video surveillance systems [9], etc. According to the requirements of each problem, different frameworks [10] and approaches for anomaly detection have been proposed [11].

This study was conducted in the frame of the European Ambient Assisted Living (AAL) project named EDLAH2 (Enhanced Daily Living and Health 2). The goal of the project is to make the usage of smart technology easy and to promote wellbeing and health among older adults. This is achieved with the development of a tablet application targeting this age group. Fig. 1 presents the homescreen of the app. The functionality of the app includes an easy to manage photo library with photos sent by family members, integrated video/audio communication, a web browser, calendar functionality and some tablet games. Additionally, the platform includes a connected wearable device that will monitor health parameters, such as the number of steps and the amount of resting time. The tablet application is reporting app usage and health data to a web server, where family members and professional carers that have permission can view statistics and information. The goal of this paper is to explore ways of detecting abnormal behaviour among older people. To do so, we apply various anomaly detection techniques to data about the usage of the homescreen tablet app and the recorded health data.

The rest of the paper is organised as follows. In Section II we discuss related works on human abnormal behaviour detection and in Section III we present various anomaly detection techniques. In Section IV we discuss our case study along with the evaluation of the abnormal behaviour detection techniques that were employed. Finally, we conclude our work with Section V.

II. RELATED WORK

There have been many studies on detecting abnormal behaviour in humans. Body-worn sensors can be used for activity recognition in order to build a model of normal activities [12]. Then, activities that largely deviate can be characterised as abnormal. Using this approach however in uncontrolled environments is extremely difficult, if not impossible, because

Fig. 1. Screenshot of the homescreen of the EDLAH2 app.

Fig. 2. Example of a box plot.

of the infinite number of activities that should be included in the training dataset and be labelled as normal [13].

Detecting anomalies has also been a topic of interest in computer vision [14]. In relevant studies, human behaviours, motion patterns and activities are modelled from video footage. Statistical-based methods are used to characterise behaviours, even in crowded scenes [15]. Video-based approaches can be employed to extract useful information in surveillance and public areas. However, due to privacy concerns, and due to the fixed locations of the cameras, these techniques can not be used for applications where constant personalised monitoring is required.

An approach to building a personalised model able to identify abnormal instances would be to track users indoors. In a similar study [16], motion and door sensors were used to track the activity of an older adult indoors. Binary dissimilarity measures, such as the classical hamming distance and the fuzzy hamming distance, are then used to measure the degree of resemblance between activity patterns. When designing a smart home environment [17], such sensors along with electricity power usage and bed/sofa pressure sensors can be installed to monitor the day-to-day activities of the inhabitants.

A similar problem to inferring behaviour through the usage patterns of a mobile application is the one of app prediction on smartphones [18]. Homescreen applications can monitor various spatiotemporal contexts [19], including the time of day, the location of the user and the previously opened app. The app can then build a personalised prediction model that will exploit the relationships between those attributes and the next application that will be executed by the user. By being able to predict smartphone app usage, one can identify usage patterns and even identify anomalies when significant deviations from the expected patterns are noticed.

In this study, we aim at detecting anomalies in the data that were recorded from the developed tablet application. Those data include app usage logs and activity data from the connected bracelet. Before proceeding with extracting useful features, we examine some approaches that have been developed for anomaly detection.
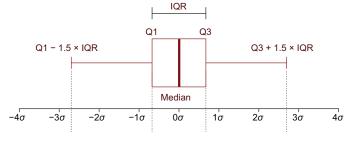
III. ANOMALY DETECTION TECHNIQUES

It is generally assumed that the anomalies that need to be detected are scarce in the given dataset. The approach for detecting anomalies can be either supervised or unsupervised, depending on whether the training set is labelled or not. Approaches can be further categorised into discriminative and generative, parametric and nonparametric, and into univariate and multivariate [11].

A. Dataset Characteristics

1) Univariate Techniques: Univariable methods are the ones that examine one variable a time. A way to identify abnormal observations is to use a variance or standard deviation-based measure. Abnormal cases would be the ones that are several standard deviations away from the mean. This measure is called z-score. However, this approach is problematic because the anomalies will influence the mean and the standard deviation in the first place, so it is less likely that they will be later identified as anomalies.

More common techniques use quartiles or percentile-based measures. In those approaches, the distance between each observation from the rest of them is based on the Interquartile Range (IQR) or the middle 50% of the scores. Groups of numerical data are typically depicted with box plots, with an example of such plot presented in Fig. 2. The box of the plot represents the area between the first (lower) and the third (upper) quartiles. The bar in the box represents the median, and the whiskers (fences) typically represent a distance of 1.5 IQRs away from the lower and the upper quartile. IQR is equal to the distance between the upper and lower quartiles. The observations outside the area defined by the whiskers are typically characterised as outliers.

2) Multivariate Techniques: In bivariate approaches, a pair of parameters is examined at a time. Such methods include distance measures, where the distance of an observation is calculated from the centre. A more comprehensible approach is the bivariate normal distribution with various confidence regions that are depicted with ellipses over a scatter plot. Observations that fall out of a selected confidence ellipse are the anomalous ones. Kernel density estimates are topographical maps that follow the density of the data and can have irregular shapes.

Clustering is usually used for unsupervised learning problems. There are many available clustering algorithms, and the

best one depends on the application and the characteristics of the available data [20]. A popular algorithm is the k-means clustering one. This algorithm partitions the available dataset into a predefined number of clusters by attempting to minimise intra-partition distances. Some other methods do not need to specify the number of clusters in advance. The mean shift and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are two such methods. The mean shift algorithm assumes that there is a probability density function and tries to place the centroids of the clusters at the maxima of that function. DBSCAN is an algorithm that assumes clusters in dense regions of data.

B. Model Characteristics

1) Discriminative Models: A way to approach an outlier detection problem is first to build a model based on a given dataset and then to compute a score for each new observation. The given dataset contains both normal and anomalous observations. The discriminative techniques consist of a similarity function that measures the similarity between two observations. Using this function, the observations are clustered so that within a single cluster the similarity is maximised, while at the same time the similarity between different clusters is minimised. An observation's anomaly score is defined as the distance of the observation from the centroid of its closest cluster. The parameters of discriminative techniques roughly consist of the definition of the similarity function and the clustering method.

2) Generative Models: In unsupervised generative techniques, a model is trained using a dataset that is known to be clean of anomalous observations. Then for every new observation, the probability of generation of such observation from the trained model is calculated. An observation is categorised as anomalous in case this probability is low, or normal if the probability is high.

3) Parametric and Nonparametric Models: A learning model that can summarise data using a predefined number of parameters is called a parametric model. These models are used in occasions where there is prior knowledge of the problem. This technique may simplify learning but may limit the learning capabilities.

On the contrary, nonparametric techniques do not make strong assumptions about the underlying distribution or the form of the mapping function. These methods seek to best fit the training data, while being able to generalise to unseen data.

IV. CASE STUDY

A. System Overview

A tablet that had the EDLAH2 application installed was given to 6 older adults above 65 years old. The users were also asked to wear an accompanying connected bracelet, the Xiaomi Mi Band 2. All of the participants were living in their own house. The data that have been collected amount to 402 user-days of tablet usage and activity tracking.

TABLE I
DATA MONITORED BY THE EDLAH2 APPLICATION

Category	Content
Browser	Interaction with the web browser
Games	Interaction with a game
Launch_activity	Execution of a feature of the app
Photos	Interaction with the photos feature of the app
Resting_time	Time windows of resting time
Steps	Hourly number of steps

The tablet application was periodically sending the app usage information along with data collected from the bracelet to a web server. The collected data are summarised in Table I. Data about the usage of the different features of the tablet app were sent to the server immediately on every interaction of the user with the tablet. Data from the bracelet, i.e. the number of steps and the amount of resting time, were grouped and sent whenever the bracelet synced with the tablet, and a Wi-Fi connection was available.

B. Feature Extraction

The R programming environment was used in order to filter out incomplete days and to process the available data. The data were aggregated on a per day manner. A number of different ways to build the features that will be later used for clustering were explored. Table II summarises the characteristics of the 4 datasets that were built and tested. Dataset1 includes features about steps, resting time and the independent usage of every aspect of the app (10 features), amounting in total to 12 features to be used for clustering. For the missing values, imputation was employed using the k nearest neighbours methods with $k = 3$. The missing values solely belonged to the app usage classes. The resting time feature accounted for the number of minutes the user was estimated to rest during the day.

Dataset2 had the same number of features as dataset1, but instead of using data imputation, all missing values were set to 0. Dataset3 was formed by keeping only 2 features of dataset2, the step count and the resting time. Finally, for dataset4 the step count and the resting time features of dataset2 were also used, but instead of independent features for every aspect of the app usage, all those features were combined into a single aggregated app usage feature.

C. Univariate Outliers

Initially, we have proceeded with detecting outliers from independent variables. Fig. 3 presents the boxplots for each variable of dataset4. An outlier, in this case, is defined as an observation that is located outside the whiskers of the boxplot.

At this stage, we have identified that some days had a value of 0 for the steps feature or for the resting time one. This can be attributed to several possibilities. It might have been that the user was not wearing the bracelet or that the battery was depleted or that the algorithm measuring the steps and the resting time malfunctioned. However, it might have been a problem that the user had, a case that an alarm should trigger from the platform.

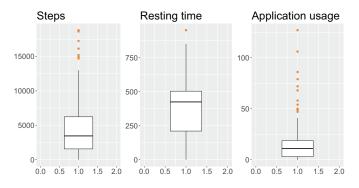| Dataset No	No of features	Missing values imputation	Missing values = 0	Steps & Resting time	Usage per feature	Aggregated usage
Dataset1	12	✓		✓	✓	
Dataset2	12		✓	✓	✓	
Dataset3	2		✓	✓		
Dataset4	3		✓	✓		✓

Fig. 3. Boxplots for each feature of dataset4.

D. Bivariate Outliers

Our next approach is to evaluate abnormality in our data in pairs. We have selected the bivariate normal distribution confidence ellipse method. Fig. 4 graphically presents the results from this technique from the features of dataset4 for a selected confidence level of 90%. The observations that are not encircled by the ellipses in the scatterplots are considered to be anomalous. This approach gives a first impression of how the variables relate to each other and enables the identification of patterns that exist across all variables.

E. Multivariate Outliers

1) Clustering: We have used different clustering techniques in order to detect anomalies. Theoretically, if daily routine data are clustered together, the odd data will be far from the normal cluster. Also, observations that are near their cluster centroid may be considered as normal, while the rest observations that are located far from their cluster centroid as abnormal. The clustering techniques that we have used are the mean shift, DBSCAN and k-means clustering with $k = 2, 3, 4$.

2) Evaluation: How well a particular unsupervised learning method performs depends on why unsupervised learning is used in the first place. In our case, we are using clustering methods expecting that the majority of the data should be considered as normal behaviour and will form a cluster with the observations relatively near the centroid of the cluster. At the same time, the observations of abnormal behaviour will be excluded from the normal cluster and may form clusters on their own.

We took the step of manually labelling the dataset by using empirical intuitions that would characterise an observation and thus a day as abnormal. In order to characterise an observation as normal, and therefore safe not to trigger any alarms, we have set thresholds for the steps and the resting time features.

Dataset	Mean shift	DBSCAN	2-means	3-means	4-means
1	47.8% (18)	- (1)	31.7%	29.4%	65.5%
2	36.7% (21)	25.8% (4)	28.3%	32.7%	33.8%
3	56.7% (3)	37.5% (2)	97.5%	56.7%	44.3%
4	41.7% (5)	43.3% (2)	89.2%	54.4%	52.5%

We have assumed that during any normal day the user should have walked more than 100 steps and rested between 4 and a half and 12 hours. We expect that our method should be able to detect otherwise so that a carer is informed via an alarm. We did not apply any thresholds to the app usage features. Obviously, days that are categorised as abnormal by this empirical approach may be a result of erroneous samples or days that the user was not wearing the bracelet. However, these might also be occasions that signal an alarming situation. We, therefore, characterised abnormal patterns as those that we assume to be of interest and should attract the attention of the carer.

To evaluate the clustering, we have assumed in all cases that the biggest cluster is the cluster containing the observations of normal behaviour, while the rest contain abnormal behaviours. In the mean shift and the DBSCAN methods, the number of clusters is automatically inferred, while in the k-means clustering the number of clusters is preselected. Standardisation was used on all datasets before clustering. For each method, three values were calculated to evaluate the performance of the abnormal behaviour detection system, the sensitivity (also called the true positive rate), the specificity and the Positive Predictive Value (PPV).

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (1)$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \quad (2)$$

$$PPV = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3)$$

Tables III, IV and V include the sensitivity, the specificity and the PPV of all clustering algorithms for all datasets. The number of clusters that were inferred from the mean shift and the DBSCAN methods is displayed in the corresponding parentheses in the tables.

We notice that the sensitivity of observing an abnormal day varies across clustering methods and datasets and peaks for
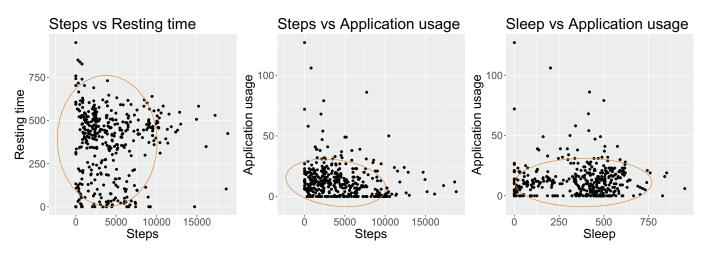
Fig. 4. Scatter plots with 90% confidence ellipses for each pair of features of dataset4.

TABLE IV
SPECIFICITY OF THE CLUSTERING ALGORITHMS

Dataset	Mean shift	DBSCAN	2-means	3-means	4-means
1	69.9% (18)	67.9% (1)	67.8%	67.4%	93.4%
2	68.6% (21)	66.8% (4)	66.67%	68.4%	69.9%
3	69.9% (3)	68% (2)	95.4%	92.5%	89.7%
4	68.2% (5)	70.1% (2)	95.2%	91.3%	91.9%

TABLE V
POSITIVE PREDICTIVE VALUE OF THE CLUSTERING ALGORITHMS

Dataset	Mean shift	DBSCAN	2-means	3-means	4-means
1	17.1% (18)	0% (1)	14.7%	15.5%	88.4%
2	14% (21)	12.4% (4)	21.7%	49.6%	56.6%
3	13.2% (3)	2.3% (2)	89.9%	88.4%	88.4%
4	3.9% (5)	22.5% (2)	89.9%	86.8%	88.4%

the dataset3 and the dataset4 using the 2-means clustering method. The fact that this method performs equally well for both datasets indicates that the single app usage feature had little contribution to the clustering. How 2-means clustering worked for dataset4 of our case is presented in Fig. 5 for each pair of the available features.

Fig. 5. Scatter plots of 2-means clustering for each pair of features of dataset4.

V. CONCLUSION AND FUTURE WORK

In this paper, we have investigated the possibility of detecting abnormal behaviour of older adults through monitoring the use of a tablet application along with the activity and resting habits of the users as these were monitored with a connected bracelet. A case study from 6 different users was presented where 402 user-days were recorded. A set of datasets containing different features were built, and different univariate and multivariate techniques, including clustering algorithms, were evaluated. Although some cases have both high sensitivities and specificities, more experimental investigations are needed to explore more reliable ways of detecting abnormal behaviours in everyday life situations.

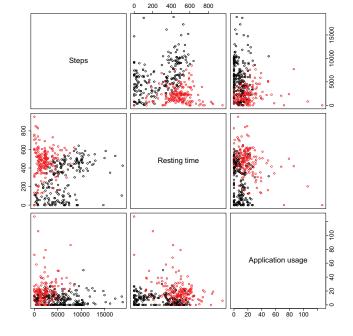Our future work aims at exploring more features that could be used for abnormal behaviour detection. Occupancy sensors in the rooms of the end users will enable us to further explore habits from the user's presence in various rooms of the house. Furthermore, the users will also be asked to fill questionnaires throughout the trials that will help us understand them better so that we better establish a ground truth for their daily behaviour.

REFERENCES

[1] H. Oh, C. Rizo, M. Enkin, and A. Jadad, "What is ehealth (3): a systematic review of published definitions," *Journal of medical Internet research*, vol. 7, no. 1, 2005.

[2] B. Martínez-Pérez, I. De La Torre-Díez, and M. López-Coronado, "Mobile health applications for the most prevalent conditions by the world health organization: review and analysis," *Journal of medical Internet research*, vol. 15, no. 6, 2013.

[3] R. S. Weinstein, A. M. Lopez, B. A. Joseph, K. A. Erps, M. Holcomb, G. P. Barker, and E. A. Krupinski, "Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers," *The American journal of medicine*, vol. 127, no. 3, pp. 183–187, 2014.

[4] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel, and A. Haines, "The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis," *PLoS medicine*, vol. 10, no. 1, p. e1001363, 2013.

[5] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, "Enabling real-time context-aware collaboration through 5g and mobile edge computing," in *Information Technology-New Generations (ITNG), 2015 12th International Conference on*. IEEE, 2015, pp. 601–605.

[6] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Computer networks*, vol. 54, no. 15, pp. 2688–2710, 2010.

[7] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[8] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using hidden markov model," *IEEE Transactions on dependable and secure computing*, vol. 5, no. 1, pp. 37–48, 2008.

[9] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Detecting contextual anomalies of crowd motion in surveillance video," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 1117–1120.

[10] S. Lee, J. Kim, S. Shin, P. Porras, and V. Yegneswaran, "Athena: A framework for scalable anomaly detection in software-defined networks," in *Dependable Systems and Networks (DSN), 2017 47th Annual IEEE/IFIP International Conference on*. IEEE, 2017, pp. 249–260.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[12] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[13] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[14] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognitiona review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865–878, 2012.

[15] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded scene analysis: A survey," *IEEE transactions on circuits and systems for video technology*, vol. 25, no. 3, pp. 367–386, 2015.

[16] S. M. Mahmoud, A. Lotfi, and C. Langensiepen, "Abnormal behaviours identification for an elder's life activities using dissimilarity measurements," in *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2011, p. 25.

[17] A. Lotfi, C. Langensiepen, S. M. Mahmoud, and M. J. Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of ambient intelligence and humanized computing*, vol. 3, no. 3, pp. 205–218, 2012.

[18] R. Baeza-Yates, D. Jiang, F. Silvestri, and B. Harrison, "Predicting the next app that you are going to use," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 285–294.

[19] Y. Xu, M. Lin, H. Lu, G. Cardone, N. Lane, Z. Chen, A. Campbell, and T. Choudhury, "Preference, context and communities: a multi-faceted approach to predicting smartphone app usage patterns," in *Proceedings of the 2013 International Symposium on Wearable Computers*. ACM, 2013, pp. 69–76.

[20] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on neural networks*, vol. 16, no. 3, pp. 645–678, 2005.