Revisiting "Recognizing Human Activities User-Independently on Smartphones Based on Accelerometer Data" – What Has Happened Since 2012?

Pekka Siirtola*, and Juha Röning

Biomimetics and Intelligent Systems Group, University of Oulu (Finland)

Abstract

Our article "Recognizing human activities user-independently on smartphones based on accelerometer data" was published in the International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI) in 2012. In 2018, it was selected as the most outstanding article published in the 10 years of IJIMAI life. To celebrate the 10th anniversary of IJIMAI, in this article we will introduce what has happened in the field of human activity recognition and wearable sensor-based recognition since 2012, and especially, this article concentrates on introducing our work since 2012.

Keywords

I. Introduction

HUMAN activity recognition is a field of science where classification methods are applied to inertial sensor data to recognize human activities. Some early preliminary activity recognition studies had already been done in the 1990's (such as [1, 2]) but [3] can be considered as the first proper inertial sensor -based activity recognition article. It concentrated on recognizing daily activities using accelerometers, and how clothes could be made aware of context.

Our human activity recognition article "Recognizing human activities user-independently on smartphones based on accelerometer data" [4] was published in IJIMAI in 2012. For the study, a daily activity data set were collected from eight healthy subjects. The trousers' front pocket was fixed as the phone placement, but the subject was allowed to determine whether the phone was placed in the left or right pocket. The participants performed five different activities: walking, running, cycling, driving a car, and idling, that is, sitting/standing. The total amount of the data collected was about four hours. These activities were selected for the study because normal everyday life consists mainly of these five activities.

Study used a window length of 7.5 seconds, and altogether 42 features were extracted from windows. These included for instance standard deviation, mean, minimum, maximum, five different percentiles (10, 25, 50, 75, and 90), and a sum and square sum of observations above/below certain percentile (5, 10, 25, 75, 90, and 95). The classification was obtained using a two stage procedure. In the first classification stage, a model was trained to decide if the studied subject was active (walking, running or cycling) or inactive (driving a car or idling). In the second stage, the exact activity label was obtained, and therefore, one model was trained to classify an active activity as walking, running or cycling, and the other to classify an inactive activity as idling or driving.

* Corresponding author.

E-mail address: pekka.siirtola@oulu.fi

The models were trained offline using the collected daily activity data set. In addition, these models were implemented to smartphones (Symbian^3- and Android-phones) and also used in online tests. To compare different classifiers, the classification was performed using two different classification methods, kNN (k nearest neighbours) and QDA (quadratic discriminant analysis). The most descriptive features for each model were selected using a sequential forward selection (SFS). QDA classifiers for offline and online recognition were trained using the whole training data set, similar to kNN classifier for the offline recognition. However, because of the limited computational power of the smartphone, the activity recognition on the device using kNN was performed using only a limited number of randomly chosen instances from training data.

The offline recognition results show that the method enables accurate results. Each activity is recognized with high accuracy. The average classification accuracy using QDA was 95.4% and using kNN, 94.5%.

For the online experiments, the application for real-time classification was tested by seven persons carrying Nokia N8 smartphone in their trousers' front pocket. In addition, classification on Android device was tested by five subjects, again, carrying the phone on their trousers' front pocket. Based on the experiences gathered using Nokia phones, only QDA classifier was implemented to Android-phone. The online recognition rates were almost identical to offline results. On a device running Symbian^3 operating system the average classification accuracy using QDA was 96.2% and using kNN, 94.1%, and on Android phone the recognition accuracy using QDA was 94.5%.

The article showed that user-independent activity recognition works reliably and operating system independently. In fact, it was one of the first articles showing that human activities can be recognized reliably in real-time in real-life conditions using smartphone hardware and smartphone sensors. The article got immediately a positive reception from the research community, and so far, according to Google Scholar, it has been cited 146 times making it the most cited article ever published in IJIMAI (numbers checked 14.11.2018). In

addition, it was selected as the most outstanding article published in the 10 years of IJIMAI life. In addition, it was a key part of Dr. P. Siirtola's Doctoral thesis "Recognizing human activities based on wearable inertial measurements: methods and applications" [5] which was published in 2014. This article introduces how the field of human activity recognition and recognition based on wearable sensor data has changed, and especially, how the research of the authors of [1] has progressed since 2012.

II. Ready-To-Use Activity Recognition

A sequel to [4] was presented in "Ready-to-Use Activity Recognition for Smartphones" [6]. It extended the work by introducing body-position independent human activity recognition method. This means that while in [4] the phone position was fixed as trousers' pocket, in [6] there was more options as the phone position. In fact, a data set containing data from five body positions (trousers' front pocket, jacket's pocket, at backpack, at brachium, and at the ear) were gathered for this study. The participants performed five different activities: walking, running, cycling, driving a car, and sitting/standing. However, there is no data from each activity from each body position. For instance, subjects were not allowed to cycle while holding a phone at the ear because of safety issues. Moreover, data were collected when a phone was laying on the table. Therefore, six activities were recognized. The total amount of the data collected was about fifteen hours.

A 3D accelerometer is a sensor consisting of three accelerometers that are approximately perpendicular to each other. However, as a side results, it was noted in the article that in reality, in each phone these sensors are aligned a bit differently due to manufacturing differences. Because of this difference, accelerometer values from different phones are at a different level, and this leads to differences in measured values and, eventually, misclassifications. On the other hand, it was noted that despite the differences in sensor values, the shapes of the signals are approximately the same but absolute values differ by some constant. This difference is normally fixed by using automated or user-driven calibration. However, because the difference in the signal level is the only major difference caused by different calibration, in our study, differences were eliminated simply by subtracting the mean of the window's values from each value of the window. This way, calibration differences can be eliminated already in the feature extraction phase and automated or user-driven calibration was not needed.

For the model training, 19 features were extracted from magnitude signal and from the signals combining two out of three acceleration channels after calibration differences were eliminated, so together 76 features were extracted and models for offline and online experiments were trained using these. QDA was decided to be used as a classifier, and in order to achieve the highest possible recognition rates, the most descriptive features for each model were selected using SFS method. Moreover, to obtain reliable user-independent results, the training was performed using the leave-one-out method, so that each person's data in turn was used for testing and the rest of the data were employed for model training.

Offline experiments showed that using the presented method, it was possible to recognize activities with a high accuracy user-independently. In addition, all the activities were detected almost perfectly from each body-position. Moreover, hardware variations between phones was not an issue: training data was collected using five different smartphones and activities from each phone were detected reliably.

Again, online experiments were performed using phones running two different operating systems, Symbian^3 and Android. Online recognition accuracies were also high, however, not as high as offline recognition results. The difference between recognition rates of online and offline scenarios was most likely caused by real-life situations from

which there was no training data. Moreover, it was not only noted that recognition can be done body position –independently without caring hardware differences, it was also noted that mobile application was light. In fact, Symbian^3 -version of the application used only around 15% of CPU capacity of the Nokia N8, and thus, the application did not eat too much battery.

III. Towards Personalized Activity Recognition Models

The main weakness of articles [4] and [6] was that they were based on user-independent models. When dealing with data collected from humans, the challenge is that people are different: they are unique for instance in terms of physical characteristics, health state or gender. All of these affect to the inertial data that are measured. In fact, it is shown that user-independent models do not work accurately for instance if they are trained with healthy study subjects and tested with subjects who have difficulties to move [7]. One more challenge is real-life, real-time conditions. It has been shown that when models that work in laboratory conditions are used in real-life conditions outside the laboratory, the results can be far from excellent [8]. In such cases, the recognition model is not general enough, and therefore, it cannot react to the changing and unseen conditions. It is especially relevant to overcome problems arising from real world conditions when the aim is to build a model that is used outside laboratory. In fact, non-stationary environments are considered one of the modern machine learning's greatest challenges [9]. Therefore, to be able to use models in real life problems, model used in the recognition process should be non-stationary instead of stationary. Moreover, because of problems related to different types of people, the focus of research should be on personal and personalized prediction models instead of user-independent models. However, the challenge of personal and personalized models is that they require personal training data. This normally would require an extensive, separate data collection session for each user.

A. From User-Independent to Personal Human Activity Recognition Models Using Smartphone Sensors

One method to obtain personalized recognition models without user-interruption was presented in [10]. The study presented a method for smartphones to obtain light weight personalized human activity recognition models unobtrusively by using the sensors of a smartphone. The proposed method consisted of four phases:

1. In the first phase, sensor fusion-based recognition model is trained and used to recognize activities from the streaming data. To maximize the recognition rate of this model, it is trained using a large number of features and these features can be based on more than one type of sensors of a smartphone (for instance accelerometers, GPS values, gyroscopes, and magnetometers).

2. When classifying streaming data using sensor fusion-based user-independent model, it can be assumed that recognition process is reliable leading to reliable classification results. Therefore, by combining these recognition results, and using them as labels, and the data related to them, a personal training data set can be gathered without user-interruption.

3. When personal data from each of the recognized activities is available, a new user-dependent recognition model can be trained. In order to make this personal recognition model light, only a small number of features extracted from a one sensor are used to train the model.

4. Streaming data can then be classified using a computationally efficient, single sensor-based user-dependent model.

The data used in the study was the same as the one used in [6], but in this case both accelerometer and magnetometer data was used.

Therefore, the user-independent model used in phase 1 was trained using features extracted from both accelerometer and magnetometer signals, and personal user-dependent model trained in phase 3 was trained using accelerometer reading only, making it more energy efficient than user-independent model.

Experiments were done with LDA (linear discriminant analysis), and QDA classifiers, and the experiments showed that the presented method improved classification accuracy when compared to traditional user-independent model. In fact, the recognition accuracy improved in nine tested cases out of ten, on average the improvement varied from 3 to 4 %-units.

B. Personalizing Human Activity Recognition Models Using Incremental Learning

Using the approach presented in [10], recognition accuracy can be improved but the problem of the approach is that personalization is based on model re-training. Therefore, in order to build personal recognition models, all the streaming data needs to be kept stored. This is problematic as it requires a lot memory, and model re-training requires a lot calculation capacity.

Incremental learning refers to recognition methods that can learn from online information and adapt to new environments. The advantage of incremental learning is that this adaptation can be done without model re-training and user-interruption. Instead, the idea is that models can be updated, instead of re-training, automatically based on streaming data [11]. Therefore, in the case of human activity recognition, incremental learning can not only be used to adapt to new environments, but also to data of new unseen person to build personalized recognition models.

Our method to personalize human activity recognition models using incremental learning was presented in [12]. In fact, our study does not only show that personalization based on incremental learning improves the recognition rates compared to results of user-independent model, it also compares three base classifiers: LDA, QDA and CART (classification and regression tree).

The experiments were made using publicly open data set [13]. This data contains data from seven physical activities (walking, sitting, standing, jogging, biking, walking upstairs and downstairs). Data set contains measurements from 10 study subjects. However, only nine persons data were used in the experiments, as apparently one of the study subjects had placed sensor in different orientation than others making the data totally different to other subjects' data. In fact, because of this problem, and other problems that can be found from publicly open datasets [14], in 2018 we introduced OpenHAR [15] which is a MATLAB toolbox to provide easy access to 10 publicly open human activity datasets.

In [12], incremental learning was based on Learn++ algorithm [16]. Learn++ is an ensemble method where the idea is to process incoming streaming data not as single values but instead as chunks. For each chunk, a new group of weak base models are trained and combined to a group of previously trained base models through weighted majority voting as ensemble model [17]. The following idea was used in our study to personalize models: in the first place, user-independent models were trained and added to ensemble. When streaming data from a new unseen person were obtained, ensemble model was used to label this new data. This data and predicted labels can then be used to train personal recognition models which can be added to ensemble. This means that instead of re-training the whole recognition model, personalized models can be obtained by updating the existing model by adding new base models to ensemble. Therefore, once the new base models were added to ensemble, the data used to train these base models were no longer needed, and it can be erased from the memory. This makes the approach very efficient computationally. Moreover, every time a new base model was added to ensemble, the model

becomes more personal, and simultaneously this continuous learning also enables a solution to adapt to new environments and unseen situations. However, problem with the data chunks used to personalize the recognition process was that they were small. Thus, they did not contain much variation leading easily to over-fitted models. To avoid over-fitting, noise injection method we presented in [18] was applied to training data sets to increase the size of training data and increase its variation.

According to the experiments shown in [12], in most cases personalization reduces the average error rate: when new base models were added to Learn++, error rates decreased. In fact, the improvement was significant: QDA and CART improved results in 7 cases out of 9 and LDA with all study subjects. With CART the average error rate dropped from 18.0% to 15.7% (13.1% improvement), with LDA from 14.1% to 9.5% (33.1% improvement), and with QDA from 11.1% to 9.1% (17.9% improvement). Comparison was made to user-independent model. Therefore, while the average error rate using QDA was the smallest, the biggest benefit from personalization can be achieved when LDA was used as a base classifier.

IV. Using Human Activity Recognition Methods In Health Applications

Back in 2012, most of the activity recognition studies were based on inertial sensor data, as devices did not normally include any other sensors. However, nowadays the situation is different. For instance, wrist-worn Empatica E4 device includes not only inertial sensor, but also thermometer, electrodermal activity sensor, which is used to measure galvanic skin response, and photoplethysmography sensor, which can measure blood volume pulse, heart rate, and heart rate variability. This has opened new research possibilities, and nowadays, similar methods that were used to train human activity recognition models can be used to train models for health and medical application.

In [19] we used data collected using Empatica E4 to early detection of migraine attacks. Data was collected from seven volunteer study subjects. Five of them were women and two were men, and the age of the study subjects varied from 30 to 60 years. They had different types of migraines, for instance five of them had aura symptoms, while two did not have. Moreover, most of them did not use preventive medication. However, all of them used medication during the migraine attacks. All of the study subjects had migraine attacks quite often. In fact, this was a criterion for joining the study as it helped the data gathering process. Frequent attacks enabled a shorter data collection period, which ensured that data consisted of several migraine attacks for every study subject. Data gathering session was long; study subjects wore Empatica E4 on their non-dominant hand approximately 27 days. Altogether, data set included 200 days of data.

When data was pre-processed, it was noted that the quality of signals was not good during the daytime due to physical activity which caused disturbances to signals. Due to this problem, only sleep time data was used in the study. Moreover, sleep time data were divided into two classes: (1) nights before a day without a migraine attack and (2) nights before a day with a migraine attack, and therefore, class (2) contains information and measured values about the pre-ictal stage of a migraine attack. The idea behind this approach is to inform the user after he/she wakes up in the morning if he/she will have a migraine attack that day, and therefore, user can take predictive medication if needed.

Similar to human activity recognition studies, features were extracted from sleep time data by considering one night as one window. Altogether, 110 features were extracted from sleep data, these included for instance standard deviation, mean, max, min, different percentiles from each signal, as well as correlations between different signals. However, the problem was that though the data set was extensive, 27 nights of data per

study subject, considering one night as one window compressed the data set so much that reliable models cannot be built based on it. Moreover, the data set was imbalanced, set includes only a few nights of data from class (2), and most of the data were from class (1).

In order to increase the number of observations, it was decided to not base the recognition on the features extracted from each night. Instead, we used the differences between nights as features. Differences were calculated so that (1) nights before a migraine attack were compared to nights before a day without a migraine, and (2) nights before a day without a migraine were compared with each other. This approach increased the number of data significantly: before this procedure, data set had 200 samples (data from 200 nights), while after applying this approach, the number of observations was 2265. In addition, to avoid over-fitting the number of observations were increased using our noise injection presented in [18].

Experiments were again made using LDA and QDA classifiers. The results showed that migraine attacks cannot be detected beforehand using user-independent model. However, the results using personal models were encouraging: balanced accuracy for detecting attacks one night prior was 70% using LDA, and as high as 84% using QDA. While the average detection rate using QDA was high, the results also show that balance accuracy varies greatly between study subjects (from 60% to 95%), which shows how complicated the problem actually is. In fact, the future work is to determine reasons for this variation. For instance, it is possible that some migraine types are more difficult to predict than others are, or it is possible that our data set was not comprehensive enough to build reliable models for different study subjects.

V. Conclusion

A lot has happened in the field of human activity recognition using wearable sensors since 2012 when our article "Recognizing human activities user-independently on smartphones based on accelerometer data" was published in IJIMAI. Our article was one of the first articles where human activity recognition was done in real-time on smartphone using smartphone's own sensors. Therefore, it had its own small role in shaping the field as it is now.

While the field has developed, also our research has progressed and diversified: we have moved from stationary recognition models to models that enable continuous life-long learning, from user-dependent models to personalized models, and from movement recognition to more comprehensively measuring humans, which enables understanding what is happening inside human body.

The future of human activity recognition, and especially, recognition based on data from wearable sensors, looks interesting. Wearable devices and their market develop rapidly, and new sensors are introduced to devices which enables new types of applications. Moreover, market development enables bigger user tests for researchers, and more business opportunities for application developers.

References

[1] J. Farringdon, A. Moore, N. Tilbury, J. Church & P. Biemond: Wearable sensor badge and sensor jacket for context awareness. Proc. Wearable Computers, 1999. Digest of Papers. The Third International Symposium on, 107–113.

[2] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz & M. Depairon (1999) Physical activity monitoring based on accelerometry: validation and comparison with video observation. Medical & Biological Engineering & Computing 37(3): 304–30.

[3] K. Van Laerhoven & O. Cakmakci (2000) What shall we teach to our pants? Wearable Computers, 2000. The Fourth International Symposium on, 77–83.

[4] P. Siirtola & J. Röning (2012) Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data. International Journal of Interactive Multimedia and Artificial Intelligence 1(5): 38-45.

[5] P. Siirtola. Recognizing Human Activities Based on Wearable Inertial Measurements-Methods and Applications. Doctoral dissertation, Department of Computer Science and Engineering, University of Oulu, (Acta Univ Oul C 524), 2015.

[6] P. Siirtola & J. Röning (2013). Ready-to-use activity recognition for smartphones. In Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on (pp. 59-64). IEEE.

[7] M. Albert, S. Toledo, M. Shapiro, and K. Kording (2012) Using mobile phones for activity recognition in parkinsons patients, Frontiers in neurology, vol. 3, 2012.

[8] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola & I. Korhonen (2006). Activity classification using realistic data from wearable sensors. IEEE Transactions on information technology in biomedicine 10(1): 119-128.

[9] M. Sugiyama & M. Kawanabe (2012). Machine learning in non-stationary environments: Introduction to covariate shift adaptation. MIT Press.

[10] P. Siirtola, H. Koskimäki & J. Röning: From User-independent to Personal Human Activity Recognition Models Exploiting the Sensors of a Smartphone, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2016., Bruges, Belgium 27-29 April 2016, 471--476.

[11] A. Gepperth & B. Hammer (2016). Incremental learning algorithms and applications. In European Symposium on Artificial Neural Networks (ESANN).

[12] P. Siirtola, H. Koskimäki & J. Röning: Personalizing human activity recognition models using incremental learning. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2016., Bruges, Belgium 25-27 April 2018, 627--632.

[13] M. Shoaib, S. Bosch, O.D. Incel, H. Scholten. and P.J.M Havinga. (2014) Fusion of Smartphone Motion Sensors for Physical Activity Recognition. Sensors, 14, 10146-10176.

[14] P. Siirtola, H. Koskimäki & J. Röning: Experiences with Publicly Open Human Activity Data Sets - Studying the Generalizability of the Recognition Models, Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, 291--299, 2018.

[15] P. Siirtola, H. Koskimäki & J. Röning: (2018, October). OpenHAR: A Matlab Toolbox for Easy Access to Publicly Open Human Activity Data Sets. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (pp. 1396-1403). ACM.

[16] R. Polikar, L. Upda, S. S Upda, & V. Honavar. Learn++: An incremental learning algorithm for supervised neural networks. IEEE transactions on systems, man, and cybernetics, part C (applications and reviews), 31(4):497–508, 2001.

[17] V. Losing, B. Hammer, & H. Wersing. Choosing the best algorithm for an incremental on-line learning task. In European Symposium on Artificial Neural Networks (ESANN), April 27-29 2016.

[18] P. Siirtola, H. Koskimäki & J. Röning: (2016). Personal models for eHealth-improving user-dependent human activity recognition models using noise injection. In Computational Intelligence (SSCI), 2016 IEEE Symposium Series on (pp. 1-7). IEEE.

[19] P. Siirtola, H. Koskimäki, H. Mönttinen & J. Röning: (2018). Using Sleep Time Data from Wearable Sensors for Early Detection of Migraine Attacks. Sensors (Basel, Switzerland), 18(5).

Pekka Siirtola

Pekka Siirtola received his M.Sc. degree in Mathematics in February 2006 and Dr. Tech degree in April 2015, both from University of Oulu, Oulu, Finland. Currently, he is working as a post-doc researcher at the Biomimetics and Intelligent Systems Group at the University of Oulu. His research interests include data mining, and machine learning, and especially, how data from wearable sensors can be used in health and wellness applications.

Juha Röning

Juha Röning is Professor of Embedded System at the University of Oulu. He serves also as Visiting Professor of Tianjin University of Technology, P. R. China. He is principal investigator of the Biomimetics and Intelligent Systems Group (BISG). In 1985 he received Asla/Fullbright scholarship. From 1985 to 1986 he was a visiting research scientist in the Center for Robotic Research at the University of Cincinnati. From 1986 to 1989 he held a Young Researcher Position in the Finnish Academy. Prof. Röning was nominated as Fellow of SPIE in 2000. He has two patents and has published more than 300 papers in the areas of computer vision, robotics, intelligent signal analysis, and software security. He is currently serving as a Board of Director for euRobotics aisbl. He is also a steering board member of ARTMIS-IA.