ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Person Identification and Discovery With Wrist Worn Accelerometer Data

Ryan McConville, Raúl Santos-Rodríguez, Niall Twomey

University of Bristol, UK

Abstract. Internet of Things (IoT) devices with embedded accelerometers continue to grow in popularity. These are often attached to individuals, whether they are a mobile phone in a pocket or a smartwatch on a wrist, and are constantly capturing data of a personal nature. In this work we propose a method for person identification using accelerometer data via supervised machine learning techniques. Further, we introduce the first unsupervised method for *discovering* individuals using the same accelerometer. We report the performance both in terms of classification and clustering using a publicly available dataset covering a large number of activities of daily living. While this has numerous benefits in tasks such as activity recognition and biometrics, this work also motivates the debate and discussion around privacy concerns of the analysis of accelerometer data.

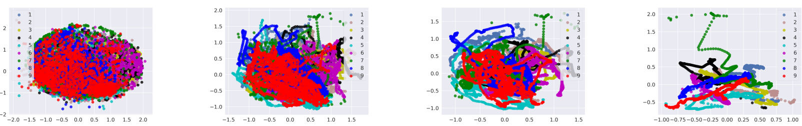
1 Introduction

As the proliferation of Internet of Things (IoT) devices continues, larger amounts of data are generated and stored on a daily basis. Increasingly, this happens inside homes with data of a personal nature. Wearable devices which collect accelerometer data raise a specific type of often neglected privacy question. Can an on-body accelerometer, such as those typically found in wrist worn devices, e.g. smartwatches, gather data that can uniquely identify the individual that it was generated from?

The motivation behind this work originates from the H2020 EurValve project, which includes an 'Smart Home in a Box' strand utilizing a wrist-worn accelerometer [1, 2]. A key concern with such systems being remotely deployed and used by participants themselves is to verify whether it is the actual participant is wearing the device, and not someone else. The potential to influence clinical decision making using this data means it is of the utmost importance that the correct participant is wearing the wearable throughout the study.

Accelerometers provide a rich amount of information; they measure acceleration and deceleration in three dimensions; the vertical x, the anteroposterior y and the mediolateral z. These devices are popular and in widespread use due to their very low cost yet broad applicability to tasks in different domains, for e.g. step counting in health and fitness. However, while other aspects of security and privacy in IoT have been well studied [3], the use of accelerometer data has attracted significantly less attention. Although there has been some recent attempts of using supervised machine learning methods to identify individuals [4], it appears that accelerometer data is still often overlooked as a personally identifiable data source. We extend existing work by proposing both fully supervised and unsupervised methods to respectively identify and discover individuals.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.



(a) 2 seconds (b) 30 seconds (c) 60 seconds (d) 720 seconds

Figure 1: The progression of signal separation as the window size increases.

We are the first to show, to the best of our knowledge, that it is possible to perform unsupervised person discovery, accurately discovering clusters corresponding to individuals.

2 Person Identification and Discovery

It is our hypothesis that each person has unique signature which can be extracted from their acceleration measured with a wrist worn device. In order to investigate this in a fair manner, we require a dataset with the following properties. (a) It should consist of people each carrying out the same activities in a controlled environment (i.e. a script). This prevents detecting specific behaviour patterns rather than a personal signature. (b) It should consist of the same script being performed again on a different occasion by each participant. The SPHERE challenge dataset [5] is a recent publicly available dataset meeting this criteria. Since the data is fully labelled, we will naturally first investigate supervised classification methods. Following this we will explore unsupervised methods to extend the framework from identification to discovery.

2.1 Dataset

First, we provide an overview of the SPHERE challenge dataset[1] which we used in order to validate our hypothesis. The data was collected from 10 people on two different occasions in a house fitted with numerous sensors. There were 8 males and 2 females, with 8 between the ages of 18 to 29 and 2 within the ages of 30 to 39. Each participant was wearing a wrist-worn accelerometer and was asked to complete a series of scripted activities, taking around 25 to 30 minutes in total. The activities included in the data covered ambulation activities (e.g. walking), posture activities (e.g. standing) and transitional activities (e.g. sit to stand). This script was carried out twice in full by each participant on different days. Due to a data recording issue with one participant in one instance, we exclude them from the experiment and are left with 18 different sets of scripted activities, each of roughly 25 to 30 minutes in length, from 9 different people.

2.1.1 Rotation correction

As described in [6], several realistic anomalies occur in the dataset, including data missingness and here we will describe a method of correcting misorientated

[1]We used a currently in preparation to be made public version of the dataset which provides information on who carried out each script. The current public version does not include this information.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

accelerometers to ensure that they conform to a standard reference. The technique we introduce is an important preprocessing step for person identification and activity recognition on the dataset at hand, but is also relevant in other settings, e.g. mobile phones in pockets.

We begin by computing the empirical distribution of each accelerometer sequence using histograms with K non overlapping bins of width $0.5g$ over the interval of $-4g$ to $4g$. A distance measure between probability distributions is required, and we have selected the Brier score [7] since it is a symmetric measure (unlike e.g. the KL divergence). One sequence is selected (arbitrarily) as a *reference*, and a set of prespecified candidate 3D rotations are applied to the remaining sequences. Finally, for each sequence the rotation that achieves minimal distance to the reference distribution is selected, and subsequently used in our analysis.

2.2 Methods

A natural way to approach this problem is to treat it as a multiclass classification problem where each person represents a different class. However, this relies on prior knowledge of which participant was wearing the wrist worn accelerometer which may not always be accessible. Thus we will also investigate an unsupervised method where we will cluster all of the accelerometer data with the objective that each discovered cluster will correspond to a single person.

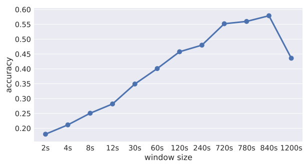


Figure 2: LR performance over various window sizes.

2.2.1 *Supervised*

First, for our supervised approach we train on each of the participants' first scripted instance and test on the second instance. While this is more challenging, it is also fair as it prevents issues such as learning to recognise the script that the participant is carrying out rather than the person themselves.

We extract the following simple features over a one second window; the mean, min, max, variance and standard deviation of each axis. However, calculating the features over just one second of activity may not contain enough of an identifiable signal. Thus we further compute the mean of each of these features over longer rolling time windows ranging from 2 seconds up to windows covering significant portions of the sequence. The importance of this can be seen in Fig. 1 where we can see the effect of increasing the window size (with PCA applied) for each of the 9 participants. With a 2 second window it is difficult to see any differentiating structure in the data, however as the window size increases to 30, 60 and then 720 seconds each individuals' signal becomes much clearer. We subsequently use a large window size, 720 seconds, to evaluate a number of classification algorithms for the task of person identification.

For each frame of each test set we use each model to predict which person (class label) the given frame belongs to. Due to the varied time it takes to carry out each script, the dataset is not equally balanced per script (person), although the random classifier performance on the dataset of 9 people is still close to 11%.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

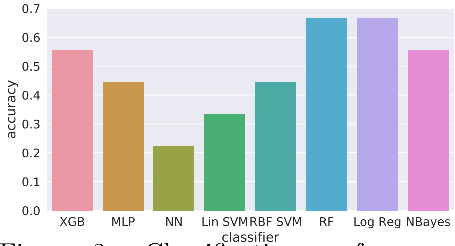


Figure 3: Classification performance over each script.

All models perform comparably, with the worst results around 2.5x better than random and the best over 5x better than random at 58% accuracy. The best model was a multinominal Logistic Regression with cross-validated parameters closely followed by a Random Forest. We also evaluated each model in a different way; as we knew that each set of accelerometer readings came from a single person, but not who that person was, we took the majority class predicted over all of the sequence for a given person. The results can be seen in Fig. 3. Most models performance improved in this scenario, with again the Logistic Regression and Random Forest among the best performers with 68% and 67% accuracy respectively. Recall the visualization discussed in Fig. 1 showing the window size to be very influential. In order to validate this, we plot the performance of the Logistic Regression model trained and tested on window sizes ranging from just 2 seconds up to 1200 seconds in Fig. 2. It is clear that up until 840 seconds classification accuracy improves consistently.

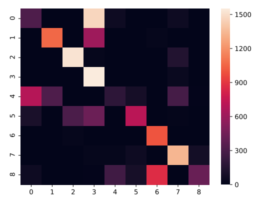


Figure 4: Confusion matrix from the LR classifier.

We further investigate the performance of the Logistic Regression (720s) model by analyzing the confusion matrix and classification report. From the confusion matrix in Fig. 4 it is clear that many people were correctly classified in the test set, with occasional confusion with at most one or two other people. We can see that person 4 is almost never predicted; investigation shows a recall of 0 and a precision of just 0.01. The rest of the people are detected much more accurately, with F-Scores ranging from 0.19 up to 0.93, with an average of 0.55.

2.2.2 *Unsupervised*

A more challenging scenario is to consider the problem as completely unsupervised; we have a set of accelerometer readings and we wish to discover which readings belong to which person.

As before we calculate the features over a rolling window of 720 seconds. Due to the structured nature of the data apparent in Fig. 1 we first attempt to use a density based clustering algorithm. Using DBSCAN [8] with the features standardized, removing the mean and scaled to unit variance, 65 clusters were found, with high scores by most metrics and far from the 9 expected clusters.

Thus, we evaluate an approach where we can specify the number of clusters in advance. We use the agglomerative hierarchical clustering Ward [9] method and specify the number of clusters as 9. We further specify connectivity constraints built using a K-Nearest Neighbour (kNN) graph where k is 50. By all metrics we achieve good performance in discovering clusters associated with each person

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

	All	**Trans.**	**Posture**	**Ambul.**	**None**
Homogeneity	0.63	0.48	0.48	0.54	0.50
Completeness	0.66	0.50	0.51	0.56	0.56
Adj. Mut. Info	0.63	0.48	0.48	0.54	0.49
Silhouette Coef.	0.13	0.21	0.19	0.28	0.34
Purity	0.68	0.57	0.53	0.57	0.65
Num. Points	32395	7855	23816	5082	1345

Table 1: Cluster quality metrics for 9 clusters.

across the two different sets (Table 1).

We note that the same issue occurred with the clustering method as did with the classification method; the person 4 was not reliably clustered into their own cluster. Instead they were often misclustered. We also found that the activity that the participants were carrying out had no significant influence, with all performing similarly (Table 1).

As compared to the accuracy in classification methods, from around 50% with a 2 minute window to 58% with a 14 minute window, when it comes to unsupervised clustering methods we can achieve high quality clusters corresponding to participants with all quality metrics tried (when having access to the number of individuals expected). Interestingly, the cluster purity, which is most similar to a classification accuracy score, is around 68%, which is around a 10% improvement over the classification accuracy. This can be explained due to the fact that the unsupervised method had more data to learn from since only 50% of the data for each person was utilized in classification for training.

3 Related Work

While some existing supervised person identification work exists [10], perhaps the most relevant work is that of Hernandez et al. [4] where a method for both posture and person identification using a wrist worn smartwatch (among others) is proposed. It extracts the ballistocardiography (BCG) signal from the participant using the wearable, then classifies the wearer as they carry out three different stationary activities; specifically sitting, lying and standing over two different one minute periods. Using a Galaxy Gear smartwatch sampling at 100Hz they achieved an accuracy of 42.93% on their dataset, when the random accuracy would be 8.3%. Our work differs, and is an improvement in a number of ways. First, we sample at 20Hz and use much less complex statistical features calculated on the accelerometer data. Their method relies on the participant standing still for 10 seconds (while ours does not) in order to extract the BCG signature. Our dataset is also significantly larger, with around 60 total minutes of data for each of our 9 participants. Further, our data is split into two different periods, where the same script consisting of 20 types of activities is repeated by each participant twice. By training on one period, and testing on the other, we ensure a fair evaluation, for e.g. that we avoid learning to recognise the specific occasion rather than the person.

The task of identification based on time series is also common in related fields such as speech processing (speaker recognition or verification [11, 12]).

4 Conclusion

In this work we proposed a strategy for person identification and discovery using a wrist-worn accelerometer on a large dataset consisting of many activities of daily living using supervised and unsupervised methods. We were able to achieve classification scores over 5x better than the random baseline. Further we are the first to propose a method of unsupervised clustering of accelerometer data for person discovery achieving a performance across many clustering evaluation metrics, including 68% cluster purity. While this has many useful beneficial applications in areas such as health, it also illustrates the argument for further debate and work on privacy of pervasive accelerometer data.

Acknowledgement

This work was supported in part by EurValve (Personalised Decision Support for Heart Valve Disease), Project Number: H2020 PHC-30-2015, 689617 and the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

References

[1] James Pope et al. Sphere in a box: Practical and scalable eurvalve activity monitoring smart home kit. In *IEEE Workshop on Networks of Sensors, Wearable, and Medical Devices*, 2017.

[2] Ryan McConville et al. Understanding the quality of calibrations for indoor localisation. In *IEEE 4th World Forum on Internet of Things*, 2018.

[3] S. Sicari et al. Security, privacy and trust in internet of things: The road ahead. *Computer Networks*, 76(Supplement C):146 – 164, 2015.

[4] Javier Hernandez et al. Bioinsights: extracting personal data from still wearable motion sensors. In *Wearable and Implantable Body Sensor Networks*, pages 1–6, 2015.

[5] Niall Twomey, Tom Diethe, et al. The sphere challenge: Activity recognition with multimodal sensor data. *arXiv preprint arXiv:1603.00797*, 2016.

[6] Tom Diethe, Niall Twomey, Meelis Kull, Peter Flach, and Ian Craddock. Probabilistic sensor fusion for ambient assisted living. *arXiv preprint arXiv:1702.01209*, 2017.

[7] Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, 78(1):1–3, 1950.

[8] Martin Ester et al. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, 1996.

[9] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

[10] Gary M. Weiss and Jeffrey W. Lockhart. Identifying user traits by mining smart phone accelerometer data. In *SensorKDD*, pages 61–69, 2011.

[11] Tomi Kinnunen and Haizhou Li. An overview of text-independent speaker recognition: From features to supervectors. *Speech Commun.*, 52(1):12–40, January 2010.

[12] Dario García-García and Raúl Santos-Rodríguez. Sphere packing for clustering sets of vectors in feature space. In *IEEE ICASSP*, pages 2092–2095, 2011.