Improving wearable activity recognition via fusion of multiple equally-sized data subwindows

Oresti Banos[1][0000-0001-5434-4253], Juan-Manuel Galvez[1][0000-0003-2139-0565], Miguel Damas[1][0000-0003-2599-8076], Alberto Guillen[1][0000-0001-9918-3238], Luis-Javier Herrera[1][0000-0003-3220-9389], Hector Pomares[1][0000-0002-8528-828X], Ignacio Rojas[1][0000-0002-1021-0496], and Claudia Villalonga[2][0000-0003-4253-7909]

[1] Research Center for Information and Communication Technologies, University of Granada, Spain
{oresti,jmgg,mdamas,aguillen,jherrera,hector,irojas}@ugr.es
[2] School of Engineering and Technology, Universidad Internacional de La Rioja, Spain
claudia.villalonga@unir.net

Abstract. The automatic recognition of physical activities typically involves various signal processing and machine learning steps used to transform raw sensor data into activity labels. One crucial step has to do with the segmentation or windowing of the sensor data stream, as it has clear implications on the eventual accuracy level of the activity recogniser. While prior studies have proposed specific window sizes to generally achieve good recognition results, in this work we explore the potential of fusing multiple equally-sized subwindows to improve such recognition capabilities. We tested our approach for eight different subwindow sizes on a widely-used activity recognition dataset. The results show that the recognition performance can be increased up to 15% when using the fusion of equally-sized subwindows compared to using a classical single window.

Keywords: Activity recognition · Segmentation · Data window · Data fusion · Wearable sensors.

1 Introduction

After some decades of research, the automatic recognition of human physical activity continues to attract an enormous interest from the scientific community. One major reason for such enduring interest is that activity recognition systems are used in a variety of application domains, ranging from industry [22, 17], healthcare [11], transportation [12], gaming [13] or housekeeping tasks [16], to name a few. Beyond research prototypes, we can already find activity recognition-based systems in a handful of products, mostly oriented to wellness applications such as fitness trackers. However, various challenges remain for activity recognition systems to penetrate in more critical domains where a high level of accuracy and performance are required.

In the particular case of wearable activity recognition, i.e. the recognition of activities based on the automatic analysis of user's data collected through wearable sensors, some challenges have to do with the placement [19], orientation [20], robustness [4], battery [21] or number [8] of the sensors. Other important challenges relate to the design of the different phases of the so-called recognition chain, as to maximise the chances of recognising the activity carried out by the user. One of such phases is the segmentation, which refers to the partitioning of the continuous data stream produced by the sensors into discrete data windows. Each data window is sequentially processed to identify the activity performed by the user during that period of time. In this direction, we showed in a previous work [2] the great impact of the segmentation phase on the accuracy of the recognition models. Amongst other findings, this work proved the existing relation among activity categories and involved body parts with the window size utilized during the segmentation process. Driven by the goal of increasing the recognition performance, we also proposed in another work a multiwindow fusion mechanism that exploits the classification capabilities of several recognisers working on different window sizes [5]. This approach shows to generally improve the recognition accuracy at the expense of running as many recognition systems as window sizes are considered.

In this paper, we explore however the potential of fusing multiple subwindows but of a similar size. We hypothesize that while using a single window, especially of a relatively long duration, the system may not fully capture the dynamics of the performed activity, thus potentially leading to an erroneous recognition. To that end, we propose a methodology for splitting and characterising each data window with a number of subwindows (Section 2). Next, we evaluate the proposed approach on a well-known activity recognition dataset (Section 3 and discuss the results for each design choice (Section 4). Final conclusions are summarized at the end of the paper (Section 5).

2 Methods

The most common approach for recognising activities from sensor data involves a sequence of steps, a.k.a. activity recognition chain or pipeline, normally consisting of sensor data acquisition, segmentation, feature extraction and classification. Depending on the quality of the collected data, some level of preprocessing might be required (e.g., band-pass filtering to remove bias and high frequency noise). Such preprocessing is however most often available on the device as a hardware component; therefore, we assume here that the sensor data stream is ready for segmentation. Likewise, feature selection or dimensionality reduction techniques are sometimes used to determine the optimal set of features to be extracted from each data window. While the feature selection could be also considered as another step of the pipeline, we rely on the system designer's choice of features, and as such, this step is not explicitly addressed in the proposed method.

Let us consider a classical activity recognition system which segments the sensor data stream into fixed windows of P seconds (i.e., recognition period).

Such system will provide, after processing the considered data window, a label identifying the activity performed by the person during those P seconds. The approach proposed here revolves around the idea of splitting such window into smaller windows, hereafter subwindows, processing each subwindow separately, and then fusing the classification decisions or labels yielded from each of these subwindows, delivering eventually a unique recognised activity. For practical reasons, the subwindow sizes should be divisors of the system's recognition period. The complete structure of the proposed model is shown in Figure 1, while its mathematical foundation is described in the following.

Generally, let us consider a problem with K classes or activities, $k = 1, ..., K$. Also, let us consider the system's data collection is composed by M sensors, each one producing a separate data stream u_m, with $m = 1, ..., M$. Each sensor data stream u_m is segmented by using N different window sizes, W_n with $W_{n-1} < W_n$ and W_n divisor of W_N for all $n = 1, ..., N$, and W_N formally representing the system's recognition period. This leads to the creation of MxN segmentation pipelines, in which every data window of size W_N, i.e., s^{W_N}, is split into W_N/W_n subwindows of size W_n, i.e., $\{s_{m1}^{W_n}, ..., s_{mi}^{W_n}, ..., s_{mW_N/W_n}^{W_n}\}$, for all $i = 1, ..., W_N/W_n$. All subwindows $s_{mi}^{W_n}$ are transformed into features, $f(s_{mi}^{W_n})$, which are then aggregated across sensors into feature vectors $\{f(s_{1i}^{W_n}), ..., f(s_{mi}^{W_n}, ..., f(s_{Mi}^{W_n}))\}$. These feature vectors are input to each corresponding classifier, yielding a recognised activity or label per subwindow, $c_i^{W_n}$.

As a result of the previous process, we obtain a vector with W_N/W_n labels, $\{c_1^{W_n}, ..., c_i^{W_n}, ..., c_{W_N/W_n}^{W_n}\}$. These labels are then used to determine the most probable activity for each data window (W_N). Different approaches can be used, however, for the sake of simplicity, we propose to use a majority voting rule:

$$D_{ik}\left(c_i^{W_n}\right) = \begin{cases} 1, \; c_i^{W_n} = k \\ 0, \; c_i^{W_n} \neq k \end{cases} \tag{1}$$

The eventual recognised class is defined as the one obtaining the highest score:

$$c^{W_n} = \underset{k}{\operatorname{argmax}} \left(\sum_{i=1}^{W_N/W_n} D_{ik}\left(c_i^{W_n}\right) \right) \tag{2}$$

3 Evaluation

3.1 Experimental setup

For the evaluation of the proposed approach we use one of the most popular wearable activity recognition datasets publicly available [1, 3]. This dataset comprises motion data, namely, acceleration, rate of turn and magnetic field orientation, recorded for 17 volunteers while carrying out 33 physical activities ($K = 33$). A set of nine inertial sensors ($M = 9$) attached to different parts of their bodies was

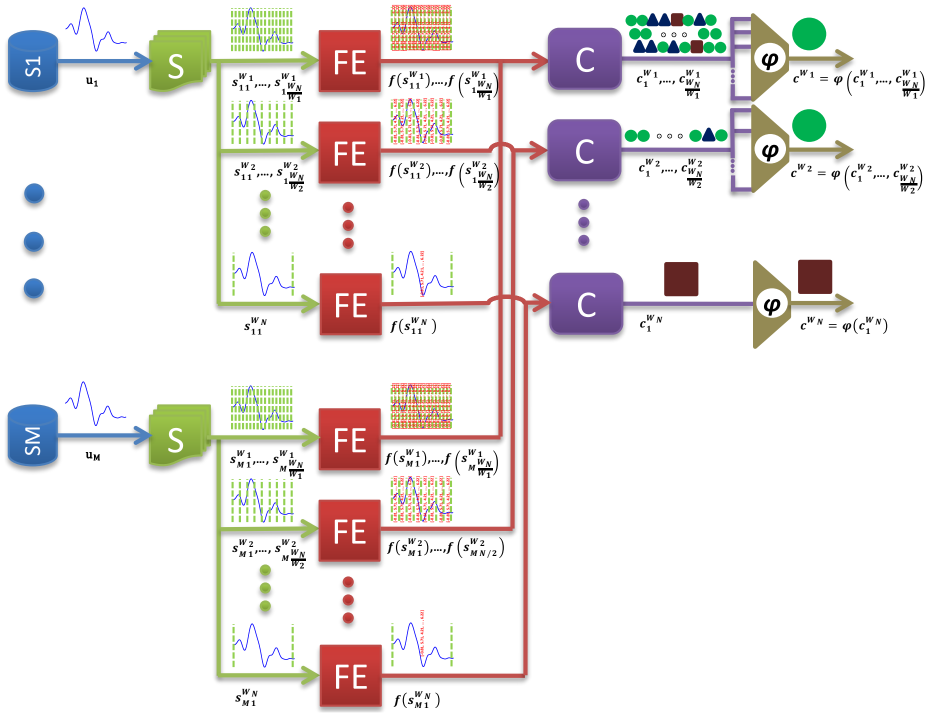

Fig. 1. Fusion of multiple equally-sized data subwindows. Given M raw sensor data streams, u_m with $m \in [1, M]$, each one is segmented into n data windows of size W_n, $s_{mi}^{W_n}$, with $n \in [1, N]$, $i \in [1, k]$ $\forall$ $k|N$. A set of features are extracted from each data subwindow, $f(s_{mi}^{W_n})$, which are then aggregated into a vector combining the computed features for the given subwindow for all sensors $\{f(s_{1i}^{W_n}), f(s_{2i}^{W_n}), ..., f(s_{Mi}^{W_n})\}$. Each feature vector is input to the n-*th* classifier (one per subwindow size), yielding a recognised activity label, $c_i^{W_n}$. The recognised labels are fused to determine, for each of the considered subwindow sizes, the eventually recognised activity c^{W_n}.

used for the motion recording. From all measured magnitudes, only the acceleration data is considered here since this turns to be the most prevalent sensor modality in previous activity recognition works. The potential of this dataset stems from the number of considered activities, diversity of body parts involved, as well as the variety in intensity and dynamicity of the actions. Moreover, all the recordings were collected in the wild, with no constraints whatsoever on the way the activities must be executed.

The activity recognition models devised for evaluation are described next. As it was stated in Section 2, no preprocessing of the data is applied to avoid the removal of any relevant information. This is normal practice when the activities are of a diverse nature. A window size of 6 seconds is considered as the reference value for the recognition system, as this figure has been used in a number of previous works. Thus, in order to meet the criterion of the subwindow sizes being divisors of the window size, the following sizes are considered for this study, $W_1 = 0.25$, $W_2 = 0.5$, $W_3 = 0.75$, $W_4 = 1$, $W_5 = 1.5$, $W_6 = 2$, $W_7 = 3$ and $W_8 = 6$ (i.e., $N = 8$), all in seconds. Three feature sets (FS) are respectively used for evaluation: FS1 = 'mean', FS2 = 'mean and standard deviation' and FS3 = 'mean, standard deviation, maximum, minimum and mean crossing rate'. These are features frequently used in activity recognition [9, 14] for their discrimination potential and ease of interpretation in the acceleration domain. Four well-known machine learning techniques widely utilized in previous activity recognition problems are also considered for classification, namely, C4.5 decision trees (DT, [7]), k-nearest neighbors (KNN, [6]), naive Bayes (NB, [18]) and nearest centroid classifier (NCC, [15]). The k-value for the KNN model is particularly set to three as it has been shown to provide good results in related works. The evaluation of the multiwindow fusion models is performed through a ten-fold random-partitioning cross validation process applied across all subjects and activities. The process is repeated 100 times for each method to ensure statistical robustness.

3.2 Results

The results obtained for the fusion of multiple equally-sized data subwindows for all possible combinations of segmentation sizes, feature sets and classification models are summarised in Table 1. No fusion is explicitly performed for the recognition models using a subwindow size of 6 seconds, as this implies using just one window. Thus, the results presented for this case coincide with the performance obtained at the classification level, i.e., before applying majority voting.

Generally speaking, the segmentation into subwindows has a great effect on the improvement of the recognition capabilities of the system. This result is observed across all feature sets and classification paradigms. Thus, for example, we can find a growth between 10% and 15% with respect to the baseline (i.e., 6s) when using subwindows of 0.25s and DT. For the same subwindow size, the

	DT			NB			NCC			KNN		
Window size (s)	FS1	FS2	FS3	FS1	FS2	FS3	FS1	FS2	FS3	FS1	FS2	FS3
0.25	**0.9995**	**0.9992**	**0.9994**	0.9193	0.9366	0.9675	0.7340	0.8343	0.8842	**1.0**	**1.0**	**1.0**
0.5	0.9900	0.9987	0.9987	0.9345	0.9727	0.9851	0.8306	0.9164	0.9566	**1.0**	**1.0**	**1.0**
0.75	0.9827	0.9972	0.9980	0.9450	0.9813	**0.9873**	0.8626	0.9379	0.9707	0.9995	0.9999	0.9999
1	0.9817	0.9961	0.9970	**0.9479**	**0.9824**	0.9863	0.8843	**0.9561**	**0.9765**	0.9994	0.9999	0.9999
1.5	0.9702	0.9890	0.9908	0.9423	0.9792	0.9821	**0.8906**	0.9552	0.9750	0.9970	0.9997	0.9998
2	0.9559	0.9795	0.9782	0.9307	0.9800	0.9830	0.8751	0.9504	0.9703	0.9952	0.9988	0.9993
3	0.9128	0.9410	0.9324	0.9208	0.9675	0.9661	0.8565	0.9377	0.9604	0.9793	0.9885	0.9924
6	0.8506	0.8961	0.8929	0.8781	0.9337	0.9249	0.7956	0.8977	0.9011	0.9363	0.9643	0.9476

Table 1. Multisubwindow fusion performance ($F - score$) for all subwindow sizes ($W_1 = 0.25$, $W_2 = 0.5$, $W_3 = 0.75$, $W_4 = 1$, $W_5 = 1.5$, $W_6 = 2$, $W_7 = 3$ and $W_8 = 6$), classification paradigms (DT, NB, NCC, KNN) and feature sets (FS1 = mean, FS2 = mean and standard deviation, FS3 = mean, standard deviation, maximum, minimum and mean crossing rate).

results vary from 3% to 6% for KNN and up to 4% for NB, with nevertheless a notable decrease in performance of up to 6% for NCC. In the case of NB we obtain up to 8% improvement with respect to baseline when using subwindows of 0.75s and 1s. For subwindows of 1 and 1.5 seconds we can reach up to 10% improvement with respect to baseline for NCC. The already good-enough results for baseline in KNN are improved around 4% to 6% for subwindows of 0.25 and 0.5 seconds. It should be also noted that a perfect recognition capabilities (F-score = 1) are reached with this model.

4 Discussion

Choosing a proper window size for the activity recognition model can often be the key to realising a system with good or poor recognition capabilities. While a simple solution may be to use those sizes that worked just fine for previous recognition systems, we could jeopardise by doing so the capabilities of our system. Hence, it is generally recommended to pay special attention to the segmentation phase and fine tune the window size value based on the considered sensors and targeted activity set.

In view of the results in this work, a valid alternative to selecting a single window size seems to be to use a combination of small windows, thus avoiding the need of engineering the optimal window size. In our earlier work [2] we demonstrated that very small windows (0.5s, 0.75s) normally lead to relatively poor recognition results, with some activity recognisers yielding F-scores below 0.5. However, in this paper we show that such small windows have a lot to offer when used in combination, outperforming even the performance reached with the optimal window sizes (1 to 2 seconds).

The above conclusion extends to any of the considered feature set and classification paradigms but for the NCC case. We do believe that such behaviour may have to do with the limitations of this classification model to capture the dynamics of the activities when considering such short windows, already reported in previous works when detecting short actions or gestures [10].

Despite the significant performance improvement, the proposed model is not without limitations. In this work we did not evaluate the complexity added from using a single window size to multiple subwindows. However, as it can be derived at a glance from Figure 1, the smaller the subwindow size is the more subwindows have to be processed (i.e., features extracted and classifications made). The ever growing computational resources available on both edge and cloud computing devices make this however less of an issue compared to past years. Either way, this should be taken into consideration specially when it comes to embedded computing systems where battery optimisation is of much relevance.

With the latest breakthroughs in machine learning, the engineering of activity recognition systems has somehow shifted to the background. While deep learning models show splendid results within a number of application domains, surpassing in many cases expert-driven feature extraction and probabilistic classification, some genuine aspects of the activity recognition problem, like the sensor data stream segmentation, are still ahead of these models. By means of this work, we take the opportunity to highlight the importance of valuing the engineering of all phases of the activity recognition chain, as well as the existing need for defining new segmentation strategies.

5 Conclusions

Using a proper window size has been shown in prior works to play a major role in the activity recognition system capabilities, which becomes quite often more important than the choice of features or machine learning models. In this work we propose a new segmentation technique that exploits the potential of equally-sized subwindows to capture the dynamics of different activities. Our results prove this approach to perform better than relying on a single window size, at the expense of increasing computational costs. Future work will explore the validation of our preliminary findings for other sensor modalities, e.g., angular velocity, and activity types, e.g., activities of the daily living.

Acknowledgments. This work was partially supported by the Spanish Ministry of Economy and Competitiveness (MINECO) Projects TIN2015-71873-R and TIN2015-67020-P together with the European Fund for Regional Development (FEDER). This work was also partially funded by the "User Behaviour Sensing, Modelling and Analysis" contract OTRI-UGR-4071.

References

1. Banos, O., Damas, M., Pomares, H., Rojas, I., Toth, M.A., Amft, O.: A benchmark dataset to evaluate sensor displacement in activity recognition. In: ACM International Conference on Ubiquitous Computing. pp. 1026–1035 (2012)
2. Banos, O., Galvez, J.M., Damas, M., Pomares, H., Rojas, I.: Window size impact in human activity recognition. Sensors **14**(4), 6474–6499 (2014)
3. Banos, O., Toth, M.A., Damas, M., Pomares, H., Rojas, I.: Dealing with the effects of sensor displacement in wearable activity recognition. Sensors **14**(6), 9995–10023 (2014)
4. Banos, O., Damas, M., Guillen, A., Herrera, L.J., Pomares, H., Rojas, I., Villalonga, C.: Multi-sensor fusion based on asymmetric decision weighting for robust activity recognition. Neural Processing Letters **42**(1), 5–26 (2015)
5. Banos, O., Galvez, J.M., Damas, M., Guillen, A., Herrera, L.J., Pomares, H., Rojas, I., Villalonga, C., Hong, C., Lee, S.: Multiwindow fusion for wearable activity recognition. In: International Work-Conference on Artificial Neural Networks. vol. 9095, pp. 290–297 (2015)
6. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory **13**(1), 21–27 (1967)
7. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Edition). Wiley-Interscience (2000)
8. Ertugrul, O.F., Kaya, Y.: Determining the optimal number of body-worn sensors for human activity recognition. Soft Computing **21**(17), 5053–5060 (2017)

9. Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.P.: Preprocessing techniques for context recognition from accelerometer data. Personal and Ubiquitous Computing **14**(7), 645–662 (2010)
10. Forster, K., Roggen, D., Troster, G.: Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain? In: International Symposium on Wearable Computers. pp. 77–84 (2009)
11. Guo, X., Liu, J., Chen, Y.: Fitcoach: Virtual fitness coach empowered by wearable mobile devices. In: IEEE Conference on Computer Communications. pp. 1–9. IEEE (2017)
12. Hur, T., Bang, J., Kim, D., Banos, O., Lee, S.: Smartphone location-independent physical activity recognition based on transportation natural vibration analysis. Sensors **17**(4), 1–21 (2017)
13. Jablonsky, N., McKenzie, S., Bangay, S., Wilkin, T.: Evaluating sensor placement and modality for activity recognition in active games. In: Australasian Computer Science Week Multiconference. p. 61 (2017)
14. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. In: International Conference on Knowledge Discovery and Data Mining. vol. 12, pp. 74–82 (2011)
15. Lam, W., Keung, C.K., Ling, C.X.: Learning good prototypes for classification using filtering and abstraction of instances. Pattern Recognition **35**(7), 1491 – 1506 (2002)
16. Liu, K.C., Yen, C.Y., Chang, L.H., Hsieh, C.Y., Chan, C.T.: Wearable sensor-based activity recognition for housekeeping task. In: International Conference on Wearable and Implantable Body Sensor Networks. pp. 67–70. IEEE (2017)
17. Malaisé, A., Maurice, P., Colas, F., Charpillet, F., Ivaldi, S.: Activity recognition with multiple wearable sensors for industrial applications. In: International Conference on Advances in Computer-Human Interactions (2018)
18. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press, 4th edn. (2008)
19. Villalonga, C., Pomares, H., Rojas, I., Banos, O.: MIMU-Wear: Ontology-based sensor selection for real-world wearable activity recognition. Neurocomputing **250**, 76 – 100 (2017)
20. Yurtman, A., Barshan, B.: Activity recognition invariant to sensor orientation with wearable motion sensors. Sensors **17**(8), 1–24 (2017)
21. Zappi, P., Roggen, D., Farella, E., Tröster, G., Benini, L.: Network-level power-performance trade-off in wearable activity recognition: A dynamic sensor selection approach. ACM Transactions on Embedded Computing Systems **11**(3), 68:1–68:30 (2012)
22. Zhao, W., Lun, R., Gordon, C., Fofana, A.B.M., Espy, D.D., Reinthal, M.A., Ekelman, B., Goodman, G.D., Niederriter, J.E., Luo, X.: A human-centered activity tracking system: Toward a healthier workplace. IEEE Transactions on Human-Machine Systems **47**(3), 343–355 (2017)