ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-1015

Using multiple contexts to distinguish standing from sitting with a single accelerometer

Hristijan Gjoreski[1] and **Simon Kozina**[1] and **Mitja Luštrek**[1] and **Matjaž Gams**[1]

Abstract. Activity recognition with a single accelerometer placed on the torso is fairly common task, but distinguishing standing from sitting in this way is very difficult because the torso is oriented the same way during both activities, and the transition between the two is very hard to classify into going down or up. We propose a novel approach based on the Multiple Contexts Ensemble (MCE) algorithm which classifies the activity with an ensemble of classifiers, each of which considers the problem in the context of a single feature. The improvement stems from using multiple viewpoints, based on accelerometer data only, designed specifically to distinguish standing from sitting. This approach improves the accuracy on the two activities by 24 percentage points compared to regular machine learning.

1 INTRODUCTION

Accelerometers are becoming increasingly common because of their lowering cost, weight and power consumption. This drives the development of an ever wider range of applications relying on wearable accelerometers, many of which involve the recognition of the user's activities. For the sake of the user's convenience, these applications are often limited to a single accelerometer. The best accelerometer placement for activity recognition (AR) [1] – and even more so for other applications built on top of AR, such as fall detection [1] and human energy expenditure estimation [2] – is on the torso. However, two very common activities – sitting and standing – do not differ in the torso orientation, so distinguishing them with a single 3D accelerometer is very challenging. The transition, i.e. going up or down, might provide an insight into the right activity, but measurements show that humans in real-life perform transitions in so many ways with so many additional movements, that render correct classification unreliable.

The research on AR is fairly extensive, but the issue addressed in this paper is largely sidestepped. Some researchers simply do not include both standing and sitting among the activities to recognize [3]. However, since both activities are fairly common, we believe this is not justified for everyday-life monitoring applications. Others solve the problem by merging these two classes [4]. And finally, some do report it [5], although this is not common. We also observed this problem in our previous work [6]. Experiments indicated that the recognition accuracy of these two activities is significantly lower compared to the accuracy of the other activities, and that usually these two activities are mutually misclassified. We additionally tested our algorithms in a living lab at the EvAAL AR competition. Even though our system was the most accurate, the recognition accuracy of these activities was poor. Therefore, in this paper we propose a novel approach to distinguishing standing from sitting that uses information from accelerometers beyond what is customarily used for AR and machine learning (ML).

2 LEARNING WITH CONTEXTS

In recent years, learning and reasoning using context proved to be effective in the ambient intelligence and ubiquitous-computing domain. For instance, our Multiple Contexts Ensemble (MCE) approach outperformed the existing approaches in the human energy expenditure estimation task [7]. In that study the idea was to use the data from multiple sensors as contexts, then to learn regression models for each context value and finally combine the output of the regression equations in order to estimate the energy expenditure of the user. Context-based approaches also proved successful in the fall detection field. Li et al. [8] and Gjoreski et al. [1] combined the user's activity and the context information extracted from environmental or wearable sensors, in order to detect a fall situation.

To distinguish standing from sitting, one first needs to separate these two activities from others. We used the AR procedure described in [6] for this task. It uses the raw chest accelerometer data, preprocesses it, extracts numerous features and applies a Random Forest (RF) classification model trained to recognize activities in real-time on one-second intervals. The RF proved most successful compared to other machine learning (ML) methods in our previous tests. Figure 1 shows how the MCE is included in the AR. MCE is used only when the RF model classifies the current activity as standing or sitting.

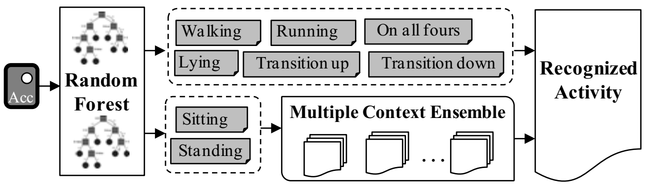

Figure 1. Overview of the AR process. RF combined with MCE for the sitting and the standing activity.

The MCE scheme for AR is presented in Figure 2. It consists of eleven contexts with appropriate context values and for each context value a classification model is learned. Once a data sample is classified by the appropriate models, the final decision about the recognized activity is performed using majority voting.

The contexts were carefully chosen based on years-long AR experiences: (1) current activity as classified by the RF, (2) previous activity (not transition) as classified by the RF, (3) last

[1] Jozef Stefan Institute, Ljubljana, Slovenia, email: {hristijan.gjoreski, simon.kozina, mitja.lustrek, matjaz.gams}@ijs.si

transition (LT) as classified by the RF, (4) LT as classified by RF trained to distinguish transitions, (5) LT as classified by a KNN using DTW metrics, (6) Angle of sensor vertical inclination (AVI) during a sit or stand activity, (7) Standard deviation (STD) of the AVI during a sit or stand activity, (8) STD of the x-axis acceleration during a sit or stand activity, (9) same as (8), but for y-axis, (10) same as (8), but for z-axis, (11) standing or sitting as classified by a KNN model with DTW during sit or stand activity.

For each context value a different model is learned. Therefore, for a given testing data sample a custom ensemble is assembled from the models that correspond to the context values of that sample. Figure 2 shows an example of a data sample containing "sitting" as a current-activity, "walking" as a previous-activity, and "transition down" as a last-transition. Thus, those three models are included in the ensemble. The learning dataset for each of the models consists of all the data samples that have that particular context value. MCE therefore constructs different subsets of the original dataset for each particular context value.

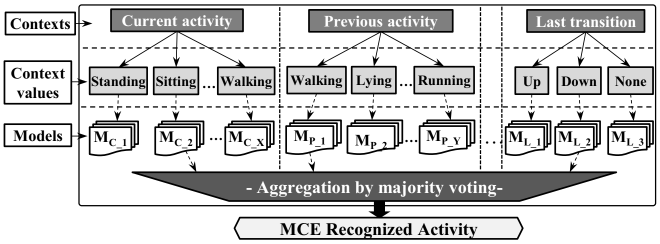

Figure 2. MCE scheme for AR.

3 EXPERIMENTAL RESULTS

To evaluate the proposed approach, we used a 90-minute activity scenario, recorded by 10 people, assembled for the Chiron project. The scenario was designed to realistically capture the real-life conditions of a person's behavior and everyday activities, although it was recorded in a laboratory. It included 8 elementary activities: walking (W), standing (Stand), sitting (Sit), lying (L), running (R), on all fours (A4), transition down (TD), and transition up (TU). The method evaluation was performed with the leave-one-person-out cross-validation technique; that is, models were trained on the data of nine people and tested on the remaining person. This procedure was repeated ten times.

Table 1 shows the confusion matrix, recall and precision values for the RF model and the combination of the RF with the MCE. The improvements of the MCE for the standing and the sitting activities are significant, i.e., 16 percentage points (p.p.) and 31 p.p. for the recall, and 17 p.p. and 22 p.p. for the precision, respectively. The further analysis of only these two activities (2x2 sub-confusion matrix marked in Table 1) shows that the accuracy is improved by 24 p.p., i.e. from 62 to 86%.

4 DISCUSSION AND CONCLUSION

Distinguishing standing from sitting in real-life circumstances has so far been too demanding to perform accurately with one accelerometer attached to the human body. But several "impossible tasks" get resolved in time when sophisticated methods are introduced. Our optimism was based on the overall fast improvement of AI and ML algorithms in particular in real-life circumstances in recent years. The trick in our case was to represent the available information from multiple viewpoints and to intelligently integrate them. Indeed, the approach proposed in this paper made a large difference of over 24 p.p. to the activity-recognition accuracy on these two activities, from 62% to 86%. The process consisted of two steps. First, we introduced additional features designed specifically to distinguish standing from sitting that could be used as contexts. Second, we classified the activities with a context-based ensemble of classifiers.

We plan to apply the MCE scheme to data from accelerometers placed on other parts of the body, e.g., the thigh. This will require some modifications: for instance, initial results show that for the thigh placement, the problematic activities are sitting and lying, so they are the ones to which the MCE should be applied. It should also be possible to apply the MCE to all the activities.

The problem discussed in this paper may appear narrow, focusing only on the recognition of the standing and sitting activities. However, since AR with accelerometers is fairly mature, it seems time to tackle the remaining weaknesses such as this one. Also, since these two activities are significantly different from the health perspective, distinguishing them is important for health-promotion applications whose popularity is on the rise.

The MCE approach is fairly general and can be applied to many problems where the available information can be presented from multiple viewpoints. Its limitation, however, is that considerable human effort is needed to present the information appropriately.

Table 1. The confusion matrix of the RF and the enhanced RF with MCE.

	W	Stand RF | RF+MCE	Sit RF | RF+MCE	TD	TU	A4	R	L	Recall RF | RF+MCE
W	**8995**	262	4	49	104	9	43	1	95.0%
Stand	93	**4265 | 5086**	1086 | 321	8	27	3	0	0	**76.4% | 92.3%**
Sit	47	2687 | 1074	**1793 | 3432**	18	19	0	0	0	**43.6% | 74.9%**
TD	155	184	19	**212**	77	50	0	36	28.9%
TU	43	127	16	31	**234**	75	1	58	40.0%
A4	1	512	0	15	10	**3286**	0	72	84.3%
R	75	5	2	3	1	0	**3523**	0	97.6%
L	2	315	0	47	59	242	10	**12786**	95.0%
Precision RF | RF+MCE	95.6%	**50.4% | 67.6%**	**61.1% | 83.4%**	55.4%	44.1%	89.7%	98.5%	98.7%	

REFERENCES

[1] H. Gjoreski, M. Gams and M. Luštrek, Context-Based Fall Detection and Activity Recognition Using Inertial and Location Sensors. Journal of Ambient Intelligence and Smart Environments, in press.

[2] M. Luštrek, B. Cvetković and S. Kozina, Energy Expenditure Estimation with Wearable Accelerometers. *Proceedings of the IEEE ISCAS*, 5–8, 2012.

[3] O.D. Lara, A.J. Péreza, M.A. Labradora and J.D. Posada, Centinela: A Human Activity Recognition System Based on Acceleration and Vital Sign Data. *Pervasive and Mobile Comp*. 8(5), 717–729, 2012.

[4] D. Curone, G.M. Bertolotti, A. Cristiani, E.L Secco and G. Magenes, A Real-Time and Self-Calibrating Algorithm Based on Triaxial Accelerometer Signals for the Detection of Human Posture and Activity. *IEEE TITB*, 14(4), 1098–1105, 2010.

[5] S.K. Song, J. Jang and S.J Park, Dynamic Activity Classification Based on Automatic Adaptation of Postural Orientation. *Proceedings of the EMBC*, 6175–6178, 2009.

[6] S. Kozina, H. Gjoreski, M. Gams and M. Luštrek, Three-Layer Activity Recognition Combining Domain Knowledge and Meta-Classification. *JMBE*, 33(4), 406-414 (2013)

[7] H. Gjoreski, B. Kaluža, M. Gams, R. Milić and M. Luštrek, Ensembles of multiple sensors for human energy expenditure estimation. *Proceedings of the Ubicomp'13*, 359-362, 2013.

[8] Li Q., and Stankovic, J. Grammar-based, posture and context-cognitive detection for falls with different activity levels. In *Proc. Wireless Health*, (2011).