Model Update in Wearable Sensors Based Human Activity Recognition

Heli Koskimäki and Pekka Siirtola
Biomimetics and Intelligent Systems Group, University of Oulu
PO BOX 4500, 90014 University of Oulu, Finland
Email: heli.koskimaki@ee.oulu.fi, pekka.siirtola@ee.oulu.fi

***Abstract*—In this article wearable sensors based human activity recognition is approached with a case where personal data collected has a high inner activity variety. With this kind of approach, the model adaptivity as well as update becomes more important issues for the activity recognition models. In authors' previous article it was shown that with this kind of data the personal models do not always outperform the user-independent models and a self-organising maps distance based approach was introduced to be used in personal and UI model fusion. In this work, the idea is developed further and it is shown that the same procedure can and should be used as predictive model as well as in data labeling also when updating the models in real-life. It is shown that compared to results of updating personal model our approach gives from 0.6 to 8.2 percentage units improvement in overall accuracy. In practice, just updating personal model with its given somewhat mislabeled data can in the worst case even drop the recognition accuracy by several percentage units which is an unwanted effect in real world applications.**

I. Introduction

The wearable sensor market is currently one of the most rapidly growing area in consumer electronics. The global market for wearables is estimated to reach $34 billion by 2020 [1] and to almost $70 billion by 2025 [2]. In research perspective, this has enabled that mobile sensors based recognition (activities, gestures, symptoms, diagnosis) to become one of the fastest developing areas of machine learning. The remarkable progress in the actual sensor development including improved memory and battery properties has making possible to measure human physiology 24/7, and more importantly with such accurate readings that has previously been possible only in laboratory settings. Moreover, the most popular area concentrating on wearable sensors based activity recognition has already been taken outside the research institutes and the results have been implemented in several commercial wrist-worn products like Polar Loop [3], Garmin Vivoactive [4] and Fitbit Blaze [5] and as mobile phone apps (Android [5], iOS [6]).

From data mining perspective, the overall activity recognition process includes a data set collected from the activities wanted to be recognized, preprocessing, segmentation, feature extraction, and classification [7]. Activity recognition is used in recognizing, for example, daily activities [8], [9], in sport sector [10], [11], [12] and in monitoring of assembly tasks [13], [14]. Quite recent trend in the research area is adaptive activity recognition where mostly the adaptiveness of the models is studied from personalization point of view. By now, the user-independent (UI) solutions have been preferred in actual applications while the main barrier for personal models is that they need personal data to be used to train the recognition models. Within end-products this is not reasonable while the user cannot be deployed to label their own data. As a solution for this approaches labeling the personal data automatically has been introduced for example in [15], [16]. In both, the solution is based on using human independent models given classification results for labeling data which then can be used to train personalized models. In addition to those, there are also studies deploying special meta-classifier approach that decides between human-independent and personal model given estimates and select the more appropriate [17], [18].

Although, it is common that personal recognition models are more accurate than user-independent [19] there actually are cases where this is not true. For example, in [20] it was shown by the authors that in the worst case the classification rate with personal model can be below a random guess. As a solution, the article presented, a novel self-organising maps (SOM) distance based approach for personal and user-independent model fusion. In the article it was shown that when there are plenty of variation within the personal data the classifier probability based solution like used in [17], [18] is not suitable to be used as a fusion criteria while it tends to emphasize the personal solution too much. It was also argued, in the article [20] that the self-organising maps approach gives a straightforward solution to a question of when to retrain the personal models. In this article, the truthfulness of the argument as well as the most efficient way to carry on the actual update are studied.

This article is organized as follows: Section II introduces the methodological aspects of the previous article [20] including the data and methods related to the activity recognition process and to the model fusion. In Section III the update procedure is described. The results and discussion of the findings are covered out in Sections IV and V while the whole study is concluded in Section VI.

II. Data set and Methodology

A. Activity recognition process

As mentioned earlier the overall activity recognition process includes a data set collected from the activities wanted to be recognized, preprocessing, segmentation, feature extraction, and classification, all introduced in this section from this study perspective.

For this study purposes 3D acceleration signals of four different gym activities from 21 persons were collected using a writs-worn inertial measurement unit. From these test subjects, 8 went to gym weekly, 4 occasionally and 9 never or rarely. Nevertheless, the exercises were selected so that they were easy to perform even for the first timers and every one could train the movements before the data collection. The actual exercises were dumbbell alternate bicep curl, side lateral raise with dumbbells, overhead triceps extension with dumbbell and sit-ups, collected as three distinct sets of ten repetitions; 1) with light weights, 2) with medium weights and 3) with heavy weights. This approach was used to ensure the difference within activities of the same persons. Nevertheless, this is not an artificial distinction while the same person's gym program can include warm up with light weights, long series with medium weights (muscle stamina) and short series with heavy weights (muscular strength).

The signals collected at a frequency of 100 Hz were divided into segments using the sliding window method, where window length of three seconds with a slide of 0.75 seconds between two sequential windows was used. For every of the windows, altogether 79 features were calculated including standard deviation, mean, percentiles and values above certain percentiles summed or square summed from each individual channels as well as from the magnitude acceleration signal ($=\sqrt{x^2+y^2+z^2}$). In addition correlation values between individual channels were used.

Based on the previous study, only the linear discriminant analysis (LDA) was used as a classifier, although, the model fusion method introduced is classifier independent. The LDA models the class-conditional densities parametrically as multivariate normals [21] by using linear decision boundaries. The methods is fast to train, easy to implement and the memory requirements are small thus making it applicable within wearable devices.

B. Self-organizing maps distance based model fusion

The self-organizing maps (SOM) distance based model fusion includes five steps.

1) Train human independent and personal models with a method of your choice
2) Use human independent data to train SOM
3) Use Euclidean distance to select the closest model vector m_i of the trained SOM for each data vectors of the personal data
4) Save the indexes i of the SOM neurons selected in previous step
5) For every new observation:
    a) Use Euclidean distance to select the closest model vector m_i of the trained SOM for the data vector
    b) **IF** the index of the closest matching unit of the new observation belongs to the saved indexes **THEN** use personal model, **ELSE** use person independent model

The self-organizing map itself is an unsupervised neural network method that presents the statistical dependencies of high-dimensional data typically in a two-dimensional space. This is done by keeping the topologic and metric relations of the two-dimensional space as close as possible to the relations of the initial high-dimensional space.

A SOM is usually formed of neurons on a regular low-dimensional grid with a hexagonal lattice. The neurons are model vectors $m_i = [m_{i1}, m_{i2}, \ldots, m_{in}]$, where n is the dimension of the input space. Training is done by choosing a data sample x and finding the closest model vector m_c (the best-matching unit). When the best-matching unit is found, it and its topologically closest neighbors are updated with the equation

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)], \quad (1)$$

where t is time, $\alpha(t)$ is the learning rate factor (a decreasing function of time) and $h_{ci}(t)$ is the neighborhood kernel centered on the winner unit c. Training continues by choosing a new data sample and iterating the updating equation [22].

In this article, the lattice of SOM was selected to be 20x20, meaning altogether 400 nodes.

III. Update procedure

In previous article [20] it was stated that the self-organizing maps distance based model fusion gives a straightforward solution to decide when to retrain the personal models, namely, that the model should be retrained when new observations are achieved that do not belong to the neurons selected based on the personal data used already in model training. Nevertheless, how this update effects to the actual classification rates and how the personal data should be labeled to achieve the best possible solution was not studied.

Thus, to study the effect of this update procedure 13 different solutions were compared; first three of them are static models for comparison purposes, models 4-7 cover updating personal model approaches, and models 8-13 cover the updating fusion model approaches. Our approach is presented as model 13. In this study, the update is done in every case as a new model training with data achieved by that time point.

1) New data predicted using static personal model
2) New data predicted using static user-independent model
3) New data predicted using fusion of static personal and static user-independent models
4) New data predicted using updating personal model; new training data labeled with exact labels (reference data)
5) New data predicted using updating personal model; new training data labeled with UI model given labels
6) New data predicted using updating personal model; new training data labeled with static personal model given labels
7) New data predicted using updating personal model; new training data labeled with updating personal model given labels

Number	Prediction model			Labeling approach				
	Static personal	UI	Updating personal	Training data	Reference	UI model	Static personal	Updating personal
1	x			x				
2		x		x				
3	x	x		x				
4			x	x	x			
5			x	x		x		
6			x	x			x	
7			x	x				x
8	x		x	x	x			
9	x		x	x		x		
10	x		x	x			x	
11	x		x	x				x
12	x		x	x		x	x	
13	x		x	x		x		x

TABLE I: Prediction procedure with 13 different solutions; with cyan background is the prediction model information and with blue background the labeling approach.

8) New data predicted using fusion of updating personal model and static UI model; new training data labeled with exact labels (reference data)
9) New data predicted using fusion of updating personal model and static UI model; new training data labeled with static UI model given labels
10) New data predicted using fusion of updating personal model and static UI model; new training data labeled with static personal model given labels
11) New data predicted using fusion of updating personal model and static UI model; new training data labeled with updating personal model given labels
12) New data predicted using fusion of updating personal model and static UI model; new training data labeled with fusion of static models given labels
13) New data predicted using fusion of updating personal model and static UI model; new training data labeled with fusion of static UI model and updating personal model given labels.

This same information is also gathered as Table I. While the model fusion approach is not basic procedure with adaptive or semi-supervised activity recognition the update versions most commonly used are 1, 2 and 4. Nevertheless, in this article the different possibilities were intentionally used to help the community to see the effect of these slight modifications.

IV. Results

While the data set consisted 3D acceleration signals of four different gym activities collected as three distinct sets of ten repetitions; with light weights, with medium weights and with heavy weights three different versions of model update procedure were conducted:

1) using light weight data in training, and medium and heavy weight data in testing
2) using medium weight data in training, and light and heavy weight data in testing
3) using heavy weight data in training, and light and medium weight data in testing.

For every person at a time, all the 13 different solutions were compared and the averages are shown as Tables II - IV.

The first notable result in the tables is that the solutions 4 and 8 outperforms the other approaches in every case. This is an expected result while the model update is done in those cases based on the exact reference labels, meaning that in every update step the actual activity is known for whole training data. In reality, this however means, that every person would need to collect and label their own data with changing weights and also due technique improvements and so on. This is a burdensome approach and, in practice, making the whole activity recognition process pointless.

When changing the focus on Table II where the light weight data is used in model training and the medium and heavy weight data as testing a case where the novel data is quite dissimilar with the existing data is seen. By concentrating on personal models 5-7 in Table II it can be seen that the solutions relying on personal model given labels (models 6 and 7) are, regardless of the updating, even worse than the static UI model (model 2) given accuracies. In practice, the update procedure improves the personal model only when UI model given labels are used (model 5). The similar phenomena stays true with model fusion approach, for example, models 10 and 11 trusting personal labels give the most inaccurate readings while when the model fusion given labels are utilized in the update procedure the accuracy of 93.7 % is achieved. This is the closest to the readings of the reference cases and 3.8 and 2.7 percentage units higher when compared to static personal model and updating personal model, respectively.

A little bit different example can be seen as Table III. In that case the medium weight data is used as training data, meaning that in this case it is expected that the unseen data is the most similar with the existing data. With this data set the most accurate reading (after the reference data label cases) are achieved using the static UI model for data labeling (models 5 and 9) although the model fusion given labels with model fusion based prediction (model 13) is only 0.4 percentage units smaller. Nevertheless, there is only 2.6 percentage unit difference between the most inaccurate (static personal model, 1) and the most accurate model (model fusion approach, UI model + updating personal model using UI model given labels, model 9) meaning that the results did not scatter widely.

Person	Model Number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	82.8	92.9	88.3	96.2	92.7	83.2	92.1	94	92.6	82.5	91.8	91.2	93.2
2	85.8	92.2	88.9	94.8	92.4	88.4	93.1	94.7	93.7	90.2	94.2	93.0	93.5
3	93.7	92.0	92.5	95.4	92.5	93.2	92.7	93.5	91.7	92.3	92.4	92.0	91.6
4	84.9	62.1	78.5	88.8	71.0	84.0	80.0	85.7	69.5	82.0	79.1	79.9	80.1
5	83.2	90.7	86.0	94.7	92.8	82.1	84.4	94.2	92.7	83.8	85.9	86.5	96.0
6	96.1	90.2	93.0	95.5	92.2	95.2	95.7	93.4	91.9	93.1	93.2	92.2	91.3
7	65.9	76.7	76.8	88.0	80.4	70.4	56.0	87.4	79.7	74.5	65.4	80.1	79.7
8	94.4	97.0	95.9	97.9	95.5	96.4	96.8	97.8	95.9	96.4	96.3	95.5	95.5
9	88.7	93.0	91.9	94.8	91.3	90.1	91.0	94.0	92.3	89.9	91.2	92.0	92.1
10	89.5	96.6	90.3	98.1	99.6	90.0	90.0	97.3	98.3	89.2	89.2	95.5	98.3
11	82.9	95.3	93.2	98.5	97.6	88.4	82.6	97.5	97.5	95.5	85.2	97.5	97.5
12	87.5	98.5	94.1	96.8	96.8	86.8	96.5	98.3	98.3	90.3	97.8	96.2	97.4
13	98.3	97.7	97.7	99.6	98.6	98.2	98.2	98.6	97.7	98.6	98.6	97.7	97.7
14	98.3	97.5	97.5	98.4	98.0	98.4	97.6	97.5	97.5	97.5	97.5	97.5	97.5
15	97.4	97.7	97.7	98.4	98.0	97.4	98.2	97.9	98.3	97.3	97.3	98.3	98.3
16	89.9	93.7	94.0	98.5	94.0	90.0	93.0	95.1	94.4	91.7	94.0	94.4	94.4
17	92.0	84.0	90.2	96.3	93.0	91.9	90.7	92.7	90.9	91.6	91.2	90.9	91.1
18	90.6	93.0	92.8	95.8	93.6	91.7	92.8	95.6	94.7	93.7	93.7	94.7	94.2
19	96.7	97.5	97.9	96.8	96.3	96.4	96.4	97.5	96.7	97.9	97.9	97.5	97.5
20	96.1	92.3	94.9	98.0	95.3	96.0	97.7	96.8	94.8	95.7	96.4	97.0	96.6
21	92.7	93.9	92.8	95.8	95.1	94.4	94.7	95.1	94.5	93.7	93.6	94.1	93.7
Average	89.9	91.6	91.7	**96.1**	93.2	90.6	91.0	**95.0**	93.0	91.3	91.5	93.0	**93.7**

TABLE II: Person-wise and average recognition rates when using 13 different prediction procedures; light weight data used for training, medium and heavy weight data used as testing. The best average results are bolded, while the accuracies under 80 % are highlighted with gray cells.

Person	Model Number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	89.2	87.9	89.4	98.1	92.3	88.6	96.4	92.9	90.4	85.7	91.8	90.7	90.7
2	93.3	92.9	92.1	94.0	92.6	91.9	92.6	94.1	93.2	93.3	93.7	92.6	94.3
3	90.7	91.9	92.0	92.7	94.8	91.9	92.1	95.3	95.4	94.1	94.3	95.8	95.4
4	78.4	68.3	73.7	86.6	79.6	76.3	79.9	84.6	78.5	77.2	80.1	79.2	80.5
5	84.1	94.2	86.7	88.7	92.3	82.1	88.5	88.2	92.0	83.5	89.9	89.2	82.8
6	95.4	94.3	95.9	98.0	94.3	94.9	95.4	96.6	94.8	94.9	95.1	95.2	95.6
7	80.5	90.1	88.3	91.5	88.1	83.1	87.3	89.2	88.0	84.7	86.4	88.0	86.7
8	94.3	98.6	95.5	98.7	98.8	95.8	98.8	98.7	98.7	97.0	98.7	97.1	98.3
9	91.6	88.1	89.3	94.5	90.0	92.5	90.6	92.9	89.0	91.5	90.2	90.1	90.5
10	98.3	95.8	97.3	98.8	94.2	97.3	89.7	97.6	95.0	96.8	89.4	95.9	95.0
11	98.0	99.0	99.5	99.0	99.0	98.5	99.0	99.5	99.5	99.0	99.5	99.5	99.5
12	89.8	96.3	92.8	97.0	96.9	89.9	89.6	96.2	96.9	89.6	89.9	92.6	96.5
13	96.0	97.4	97.4	97.9	97.0	95.0	94.1	97.9	97.5	96.3	96.4	97.5	97.5
14	85.4	93.3	91.5	96.8	92.0	86.9	90.5	95.1	93.3	91.9	92.7	92.5	92.5
15	96.9	97.8	98.4	98.2	97.5	97.0	97.5	98.6	98.2	97.3	97.8	98.0	98.2
16	92.8	95.8	95.5	96.8	94.7	93.9	95.8	95.8	95.4	94.7	95.0	95.4	95.4
17	92.0	89.3	94.4	95.9	93.6	90.0	91.9	94.4	94.4	93.5	94.0	94.4	94.0
18	88.9	92.4	93.3	97.2	96.6	91.9	97.0	96.5	96.6	93.0	95.7	95.9	96.6
19	96.3	94.1	95.5	97.2	95.7	95.5	96.0	96.1	95.7	95.7	95.7	95.7	95.7
20	92.6	93.3	93.7	96.8	94.2	93.2	93.2	95.7	94.0	93.2	93.2	94.4	93.0
21	89.5	90.2	90.2	93.4	90.9	89.2	90.2	93.3	90.4	89.4	90.4	90.4	90.0
Average	91.1	92.4	92.5	**95.6**	93.6	91.2	92.7	**94.7**	**93.7**	92.0	92.9	93.3	93.3

TABLE III: Person-wise and average recognition rates when using 13 different prediction procedures; medium weight data used for training, light and heavy weight data used as testing. The best average results are bolded, while the accuracies under 80 % are highlighted with gray cells.

The biggest differences between updating personal models and updating fusion models are seen in Table IV. In this case, the static models are trained with heavy weight data, and the testing and update procedure is done with light weight data and medium weight data, and more specifically, in that order. While the results of Table II were accumulated so that the model was updated with medium weight data before it was used to predict the heavy weight events, in this case the light weight events were predicted based on the heavy weight based model (very dissimilar), and the medium weight events were predicted based on heavy weight and light weight data. Because of this order even 8.2 percentage units higher accuracies are achieved with our proposed approach (model 13) compared to updating personal model (model 7). Moreover, the updating personal model using its own labels (model 7) performs on average over 4 percentage units poorer than the static personal model (model 1). Nevertheless, with our approach the 93.9 % overall accuracy is achieved regardless of the order the new

Person	Model Number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	80.2	89.0	87.7	96.6	93.0	82.4	85.8	92.2	91.8	82.8	84.4	90.3	92.2
2	88.9	89.9	90.3	92.5	89.9	89.6	89.9	91.8	91.1	90.6	90.6	91.9	92.4
3	87.2	92.6	89.6	93.2	92.0	88.4	92.3	92.4	92.8	88.5	92.3	93.7	92.7
4	75.4	82.1	78.5	92.5	84.5	75.1	72.6	92.6	85.3	77.7	76.7	83.1	83.2
5	79.6	94.0	92.4	94.3	92.4	80.0	56.5	95.4	94.7	84.8	81.7	94.1	95.2
6	96.3	90.2	94.1	96.5	92.4	94.5	93.2	93.7	92.1	92.9	92.0	93.3	93.2
7	77.8	89.0	85.4	90.3	85.0	77.3	55.0	90.7	88.3	79.4	64.3	86.1	85.5
8	92.6	91.7	92.9	96.5	90.9	93.0	93.4	96.4	92.7	93.3	93.0	92.6	94.8
9	97.0	93.4	95.0	98.2	95.1	97.2	98.2	96.1	95.4	96.0	96.1	95.4	95.4
10	92.2	93.0	91.1	99.1	96.1	89.5	73.3	96.9	96.1	87.1	76.8	94.3	96.9
11	97.8	99.5	99.0	98.9	99.5	99.4	99.4	99.0	99.0	99.5	99.5	99.5	99.0
12	97.6	97.7	97.6	98.2	98.5	97.5	96.5	98.3	98.3	98.3	96.7	98.3	98.3
13	95.3	91.1	94.2	97.4	93.8	96.1	95.4	95.8	93.5	95.3	95.3	94.5	94.5
14	93.2	94.6	95.7	97.6	94.8	92.6	90.9	96.6	95.9	94.6	94.5	95.4	95.4
15	94.7	96.8	95.9	98.1	95.2	94.1	91.9	97.1	96.6	95.7	93.5	96.3	96.3
16	97.1	98.0	98.6	96.3	96.3	96.7	95.8	97.5	97.5	98.0	97.5	98.0	98.0
17	87.2	93.5	87.8	96.5	93.5	79.7	93.0	94.4	93.0	81.8	93.0	88.9	93.0
18	79.2	92.1	92.3	93.8	88.4	77.3	73.3	93.5	88.3	82.6	80.4	88.8	92.3
19	91.1	94.9	93.7	98.0	94.8	90.9	68.9	98.3	96.5	92.8	74.7	96.2	96.9
20	95.4	90.0	96.2	96.0	95.1	94.7	93.7	96.4	96.5	95.9	95.4	95.7	95.1
21	93.1	91.3	90.9	95.3	91.9	92.0	91.6	92.8	91.2	91.6	90.0	91.2	91.2
Average	89.9	92.6	92.3	**96.0**	93.0	89.4	85.7	**95.1**	93.7	90.4	88.5	93.2	**93.9**

TABLE IV: Person-wise and average recognition rates when using 13 different prediction procedures; heavy weight data used for training, light and medium weight data used as testing. The best average results are bolded, while the accuracies under 80 % are highlighted with gray cells.

Training data \ Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1) light weight	59	73	57	106	58	55	66	65	76	53	58	45	37	45	88	80	45	60	75	54	46
2) medium weight	66	63	60	112	62	45	79	61	82	57	60	57	44	57	83	69	51	49	73	46	62
3) heavy weight	69	77	58	74	82	56	88	51	52	40	38	41	46	48	71	48	40	46	59	58	56

TABLE V: Person-wise amount of updates.

events arrive.

From the different models our proposed approach (model 13) and the fusion of updating personal model and static UI model by using UI model given labels (model 9) outperformed the rest. Nevertheless, with this data set the overall accuracy of the UI model was quite high (over 90%) thus effecting positively the accuracy of model 9. The model 13 is not too sensitive to the initial UI accuracy and thus more reliable in general.

On the other hand, when considering the accuracies achieved for individual persons in Tables II - IV, it can be seen that in all the cases the person 4 has had the poorest overall accuracy. This is not surprising, while the test subject is one of those who have exercised at gym never before or rarely. Due this novelty aspect the actual movement trajectories include plenty of variation showing also in acceleration signals. Nevertheless, it is surprising that the test subject 5 in Table IV who have only 56.5 % accuracy with model 7, is one of the subjects visiting gym weekly. With him/her the movement trajectories change drastically when changing weights although the actual exercises are well-known.

V. Discussion

In authors' previous article, it was argued that self-organising maps distance based model fusion, gives a straightforward solution to a questions of when to retrain the personal models. In this article it was clearly seen, that it indeed holds true and the approach is also most efficient, automatic way to label the training data and, moreover, to be used as predictive model. Naturally, if it can be assumed that the personal data is unchangeable or exact labels can be achieved easily, this approach is not necessary. Nevertheless, the performance of the most activities change due the time, or as simplest when walking with different shoes, thus even affecting on that the person independent models can be more accurate in long run. It was also shown, that if the adaptiveness of the model is achieved using somewhat mislabeled data in update procedure the accuracy of the personal models can drop significantly.

In this article, the models were retrained when new observations were achieved that did not belong to the closest matching units of the personal data used in model training. Although the lattice of used UI model based SOM consisted of 400 nodes in every case, the amount of nodes covered with personal initial training data or new nodes selected by the new observations differed quite noticeable. For example, in Table V the maximum amount of updates was for person 4 using medium weight data as initial training data and light weight and heavy weight data for testing altogether 112 updates. On the other hand, with person 13 and using light weight data for training only 37 updates were needed.

It has to be noted, that although only one classifier was used, due the comparison results in previous article, the approach is model independent. There is no necessity to use same classifier as personal and UI model but case-specific models can be utilized. Moreover, the approach is not activity recognition specific, but it can be used also in other application areas where human centered data is used.

VI. Conclusion

The main concentration in this article was to discover the human activity recognition model update in cases where the personal data collected has high inner variety. In most of the activity recognition studies, the activities are collected as one continuous session and thus the inner variety can be considered very low. In worst cases, the training and testing data sets for personal models are selected with random sampling causing possibility that due the overlapping windows some of the training and testing samples can have three quarters of the data the same with each other. Naturally, in these kind of approaches the personal models outperform the UI models drastically.

Nevertheless, it was shown in the article that by using the self-organising maps distance based approach for personal and UI model fusion as well as in data labeling when updating the models in real-life it was shown that our approach gives from 0.6 to 8.2 percentage units improvement in overall accuracy compared to results of updating personal model. Thus making the solution advisable approach with human centric data applications. On the other hand, it was also shown that by using mislabeled data in model training the recognition accuracies can even drop by several percentage units.

To achieve more generalizable results more versatile data sets would be preferred. The problem, however, is that the most of the open activity recognition data sets are collected as a one ensemble during a single data collection session causing very little variance in the data. Thus to test the approach, for example, to daily activity recognition either an artificial distinction has to made (e.g. in walking and running by using different shoes or different running base material) or naturally letting the time pass between these data collection periods. Nevertheless, to collect that kind of data will be left as a future work.

Acknowledgment

The authors would like to thank Academy of Finland for funding this work (Decision 257468, Postdoctoral Researcher project: Mobile Sensors and Behavior Recognition in Real-world).

References

[1] "Ccs insight," http://www.ccsinsight.com/press/company-news/2516-wearables-momentum-continues, accessed: 2016-06-10.

[2] "16 global retail tech trends at a glance," https://www.jda.com/-/media/jda/knowledge-center/white-papers/16-retail-trends-2016.ashx, accessed: 2016-06-10.

[3] "Polar loop," http://www.polarloop.com/, accessed: 2016-06-10.

[4] "Garmin vivoactive," https://buy.garmin.com/en-US/US/into-sports/health-fitness/vivoactive-/prod150767.html, accessed: 2016-06-10.

[5] "Fitbit blaze," https://www.fitbit.com/fi/blaze, accessed: 2016-06-10.

[6] "Motionactivity," https://developer.apple.com/library/ios/documentation/CoreMotion/Reference/CMMotionActivityManager_class/, accessed: 2016-06-10.

[7] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 33:1–33:33, Jan. 2014. [Online]. Available: http://doi.acm.org/10.1145/2499621

[8] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas, "Daily living activity recognition based on statistical feature quality group selection," *Expert Systems with Applications*, vol. 39, no. 9, pp. 8013–8021, 2012.

[9] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 553–560, May 2013.

[10] K. Chang, M. Chen, and J. Canny, "Tracking free-weight exercises," *UbiComp 2007: Ubiquitous Computing*, pp. 19–37, 2007.

[11] H. Koskimäki and P. Siirtola, "Recognizing gym exercises using acceleration data from wearable sensors," in *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*. IEEE, 2014, pp. 321–328.

[12] M. Muehlbauer, G. Bahle, and P. Lukowicz, "What can an arm holster worn smart phone do for activity recognition?" *15th Annual International Symposium on Wearable Computers (ISWC)*, pp. 79 – 82, 2011.

[13] T. Stiefmeier, D. Roggen, G. Tröster, G. Ogris, and P. Lukowicz, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42–50, 2008.

[14] H. Koskimäki, V. Huikari, P. Siirtola, P. Laurinen, and J. Röning, "Activity recognition using a wrist-worn inertial measurement unit: a case study for industrial assembly lines," *The 17th Mediterranean Conference on Control and Automation*, pp. 401–405, 2009.

[15] D. Roggen, K. Förster, A. Calatroni, and G. Tröster, "The adarc pattern analysis architecture for adaptive human activity recognition systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 2, pp. 169–186, 2011. [Online]. Available: http://dx.doi.org/10.1007/s12652-011-0064-0

[16] P. Siirtola, H. Koskimäki, and J. Röning, "From user-independent to personal human activity recognition models using smartphone sensors." ESANN 2016, April 2016.

[17] B. Cvetković, M. Luštrek, Boštjan, Luštrek, and M. Gams, "Semi-supervised learning for adaptation of human activity recognition classifier to the user," *Workshop on Space, Time and Ambient Intelligence*, 2011.

[18] B. Cvetković, B. Kaluža, M. Luštrek, and M. Gams, "Multi-classifier adaptive training: Specialising an activity recognition classifier using semi-supervised learning," pp. 193–207, 2012.

[19] G. Weiss and J. Lockhart, "The impact of personalization on smartphone-based activity recognition," in *AAAI Workshop on Activity Context Representation: Techniques and Languages*, 2012.

[20] H. Koskimäki and P. Siirtola, "Adaptive model fusion for wearable sensors based human activity recognition," in *International Conference on Information Fusion*, 2016, accepted.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[22] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, Berlin, Heidelberg, 2001.