Context-aware incremental learning-based method for personalized human activity recognition

Pekka Siirtola[1] · Juha Röning[1]

Abstract

This study introduces an ensemble-based personalized human activity recognition method relying on incremental learning, which is a method for continuous learning, that can not only learn from streaming data but also adapt to different contexts and changes in context. This adaptation is based on a novel weighting approach which gives bigger weight to those base models of the ensemble which are the most suitable to the current context. In this article, contexts are different body positions for inertial sensors. The experiments are performed in two scenarios: (S1) adapting model to a known context, and (S2) adapting model to a previously unknown context. In both scenarios, the models had to also adapt to the data of previously unknown person, as the initial user-independent dataset did not include any data from the studied user. In the experiments, the proposed ensemble-based approach is compared to non-weighted personalization method relying on ensemble-based classifier and to static user-independent model. Both ensemble models are experimented using three different base classifiers (linear discriminant analysis, quadratic discriminant analysis, and classification and regression tree). The results show that the proposed ensemble method performs much better than non-weighted ensemble model for personalization in both scenarios no matter which base classifier is used. Moreover, the proposed method outperforms user-independent models. In scenario 1, the error rate of balanced accuracy using user-independent model was 13.3%, using non-weighted personalization method 13.8%, and using the proposed method 6.4%. The difference is even bigger in scenario 2, where the error rate using user-independent model is 36.6%, using non-weighted personalization method 36.9%, and using the proposed method 14.1%. In addition, F1 scores also show that the proposed method performs much better in both scenarios that the rival methods. Moreover, as a side result, it was noted that the presented method can also be used to recognize body position of the sensor.

Keywords Personalization · Context-awareness · Human activity recognition · Incremental learning · Adaptive models

1 Problem statement and related work

Human activity recognition studies what a person wearing a sensor is currently doing. This information can then be used to create an automatic activity diary, to adapt user-interface based on the performed activity, or to create other activity-aware applications, etc. (Incel et al. 2013). People are different, and therefore, an inertial sensor-based activity detection model providing accurate results for one person, does not necessarily work accurately with somebody else's data as people are unique, for instance, in terms of physical characteristics, health state or gender (Albert et al. 2012). In addition, person's physical characteristics can also change over time, and thus the recognition models should be able to adapt to these changes. Therefore, when human activities based on inertial sensor data are recognized, the recognition should be based on adaptive personal or personalized models supporting continuous learning instead of static universal models. Moreover, often data from smartphone sensors are used in the human activity recognition studies and it is known that people have their own styles to use phones. Due to this, they carry their phones on different ways and places, such as pocket or backpack. It is shown that a human activity recognition model trained using data from one body position does not work well when it is tested using data from another position (Widhalm et al. 2019). Therefore, activity recognition models should not only adapt to user's personal moving style but also to context, and therefore, be context-aware and

✉ Pekka Siirtola
  pekka.siirtola@oulu.fi

[1] Biomimetics and Intelligent Systems Group, University of Oulu, P.O. Box 4500, 90014 Oulu, Finland

 Springer

not be depended on certain body position. This would allow more personal, accurate and user-friendly services.

In most of the inertial sensor-based human activity recognition studies, the recognition is based on user-independent models (Bulling et al. 2014). However, there are also studies relying on personal or personalized recognition models. Personal models rely purely on training data obtained from one user. Therefore, they are more accurate than user-independent models (Weiss and Lockhart 2012; Munguia 2008), but on the other hand, they require a separate data gathering session from each user making them non-usable out-of-the-box (Bulling et al. 2014). Moreover, personalized models are a hybrid of user-independent and personal models. They are mostly based on user-independent data but also use a small amount of personal data to adapt the model to better describe the needs of a specific user (Garcia-Ceja and Brena 2015). In fact, it has been shown in several studies that personalization improves the recognition rates, and several methods have been proposed for personalization. For instance, in Fallahzadeh and Ghasemzadeh (2017) the personalization was relying on transfer learning algorithm and unsupervised learning. The study showed the power of personalization: according to the study, the recognition rates of personalized models are much better than the recognition rates obtained using models that are not personalized. Moreover, in Akbari and Jafari (2020) a personalization method relying on deep learning and retraining was proposed. In this case, supervised personalization improved the recognition accuracy of a new user as much as 25%. Another approach for model personalization was presented in Ferrari et al. (2020) where personalization was based on physical characteristics of the persons and characteristics of the signals produced by these persons while performing activities. This was done by weighting the segments of the training data with the similarity between the subjects in the training and the subject of the test. In the study, this personalization improved the accuracy on average 11%. These studies show the power of personalization but the problem of these is that they cannot adapt to new environments, and they do not support continuous learning.

In this study, incremental learning is used to personalize recognition models. This term refers to recognition methods that can learn from streaming data and adapt to new and changing environments by updating the model (Losing et al. 2018; Gepperth and Hammer 2016). Therefore, they support continuous life-long learning. Indeed, as the models are updated, they do not need to be retrained during the personalization process which has been the case in some other personalization methods, such as Cvetković et al. (2015) and Akbari and Jafari (2020). Due to this, there is no need to store the whole training dataset, which makes incremental learning-based applications suitable for devices with limited memory. In fact, incremental learning has already been used to detect human activities based on inertial sensor data (Wang et al. 2012; Mo et al. 2016; Ntalampiras and Roveri 2016; Abdallah et al. 2015). These studies show that the performance of activity recognition models can be improved by updating the models using incremental learning. However, in these articles the aim was not to personalize recognition models. Instead, the experiments performed in these studies concentrate more on showing that adapting models to evolving data streams increases the model accuracy, and that the accuracy of user-independent recognition model can be improved when more study subjects are added to the training data.

Incremental learning was used to personalize human activity recognition models in Yu et al. (2016), where classification was based on logistic regression models. In the study, user-independent model was trained and it was then updated and personalized based on personal data. In the study, the sensor had only one possible body position, and the labels for personal data were known, meaning that model update was based on supervised learning. This means that if the aim is to update model several times, a lot of personal data is needed to label making the approach laborious for the user. A different approach was introduced in Siirtola et al. (2018). In the study, the body position of the sensor was again fixed but the personalization process was fully autonomous, and therefore, model update was based on unsupervised learning where model update was based on personal data and predicted labels. While this kind of autonomous approach is easy for the user and enables continuous learning, it can easily learn wrong things and start to suffer from concept drift. To solve this problem, Mannini and Intille (2019) and Siirtola and Röning (2019) introduced methods where the process of updating and personalizing incremental learning-based human activity recognition models rely partly or totally on observations labelled by the user. In both studies, user did not have to label all the instances, instead, user-labelled only unreliable observations. In Mannini and Intille (2019), only these user labeled, previously unreliable observations, were used to update models while Siirtola and Röning (2019) used a combination of user-labelled and predicted labels in the updating process. Moreover, in both studies it was noted that updating models with personal data does improve the model accuracy, and in fact, according to Siirtola and Röning (2019), user labelled instances have a huge positive effect to the recognition rates. In addition, in Siirtola and Röning (2019) it was noted that updating models with user-independent data does not have the same effect. However, in these studies it was assumed that the sensor position is fixed to some body position. Therefore, they work only in one context. Unfortunately, this is not realistic assumption when sensor data from smartphone sensors is studied as people can carry their phones in several places, for instance at trouser's pocket, backpack, and handbag.

There are a lot of studies where context-aware models that can recognize the body position of the sensor are studied (for instance Fujinami et al. 2012; Sztyler and Stuckenschmidt 2016). Moreover, it has been showed that human activity recognition models can be trained as body position-independent which means that models work accurately regardless on which body position the sensor is located (such as Shi et al. 2017; Figueira et al. 2016; Almaslukh et al. 2018). However, the approaches presented in these studies do not support adaptation, personalization or learning from streaming data. On the other hand, there is at least one study where it is shown that also body position-aware models also benefit from personalization (Sato and Fujinami 2017). However, the method presented in the paper can only be used to detect the position of the sensor, and not the performed activity. Therefore, if the aim is to build context-aware human activity recognition model that can adapt to user's personal style of performing daily activities, other methods needs to be used.

This article introduces an incremental learning-based human activity recognition model that can adapt to user's personal movement, and changes in movement. What make the introduced method novel is that unlike other online learning-based personalizing methods (see for instance Siirtola et al. 2018; Mannini and Intille 2019; Siirtola and Röning 2019), it works in several contexts, even in previously unknown contexts. This means that it can recover from the drastic concept drift caused by changes in contexts, and also learn these new contexts. The context-awareness is obtained in this article based on novel base model weighting approach which gives weights to the base models based on their performance in the latest user-labeled data, and therefore, it gives biggest weights to those base models which perform the best in the current context. Contexts that are studied in this article are different body positions for inertial sensors. This means that in previous online learning-based personalization methods, the body position of the sensor was fixed, while the method proposed in this article is not similarly dependent on the body position of the sensor as it works in several body positions.

The article is organized as follows: Sect. 2 introduces the used dataset and Sect. 3 introduces and explains the methods used in context-aware models personalization. Experimental setup and results are in Sect. 4, discussion in Sect. 5, and finally, the conclusions are in Sect. 6.

2 Experimental dataset

Experiment of this study are based on a publicly open data set presented in Shoaib et al. (2014). This dataset contains 3D accelerometer, 3D magnetometer and 3D gyroscope data from seven physical activities (walking, sitting, standing, jogging, biking, walking up and walking down). The dataset contains equal amount of data from each activity. The data were collected using Samsung Galaxy SII and a sampling rate of 50 Hz, and from five body positions at the same time (belt, wrist, arm, trouser's left pocket, and trouser's right pocket). In this study, accelerometer and gyroscope data is used from four body positions: belt, wrist, arm, and trouser's right pocket. Therefore, in this study, these are considered as the possible contexts of the sensor. The data from trouser's left pocket was left out from the study as it is highly similar to data from trouser's right pocket. The data is collected from ten study subjects (male ages between 25 and 30). However, one of the study subjects had carried a smartphone in different orientation than others making the data totally different to other subjects' data. Thus, only nine persons data were used in the experiments.

For feature extraction, the data were divided to time windows of 4.2 s (210 samples) and a 1.4 s slide (70 samples) was used. Features were then extracted from there windows. The features that were extracted are features that are commonly used in human activity recognition studies. These include statistical features such as standard deviation, minimum, maximum, median, and other percentiles (10, 25, 75, and 90). In addition to these, the extracted features include the sum of values above or below percentile (10, 25, 75, and 90), square sum of values above or below percentile (10, 25, 75, and 90), and number of crossings above or below percentile (10, 25, 75, and 90). Frequency domain features were extracted as well, for instance sums of small sequences of Fourier-transformed signals. All of the features were extracted from raw accelerometer and gyroscope signals (x, y, and z signals). Moreover, magnitude accelerometer and gyroscope signals was calculated ($\sqrt{x^2 + y^2 + z^2}$) and signals where two out of three accelerometer and gyroscope signals were combined, and same features were extracted also from these. Altogether, 202 features were extracted from the signals and eventually the data consisted of 8980 windows of data from each body position.

3 Methods for context-aware model personalization

Figure 1 shows a use case for model personalization, and the idea of personalization. Initially, when person uses application for the first time, the model for human activity recognition is user-independent. When the person starts to use the application, the activity recognition models are personalized based on streaming data. Moreover, when the recognition result obtained using recognition models is found unreliable, the application asks the correct activity label from the user. Based on this streaming data and labels obtained via user or the recognition model,
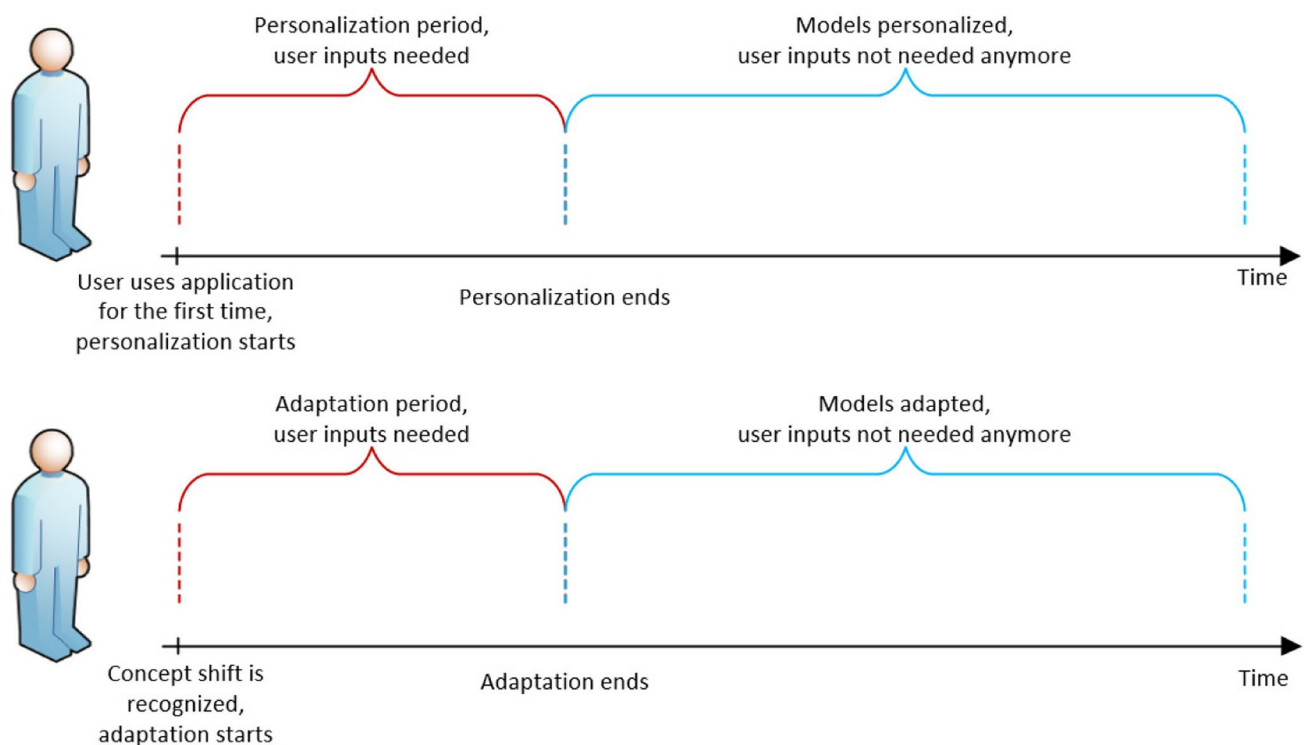
Fig. 1 Use case for personalized human activity recognition models relying on user inputs

the models can be updated and personalized. After the personalization period is over, the application can run on it's own and it provides accurate personalized recognition results to the user and user does not need to provide labels anymore.

The real-life situations and environments are often unpredictable, and the data used to train the personalized ensemble model cannot contain all the real-world situations. These can cause concept drift that slowly affects to the measured signals and recognition rates, or they can cause drastic concept shift that rapidly affects to the measured signals and have a sudden negative effect to the recognition rates (Roggen et al. 2013). When concept shift is recognized, for instance based on low posterior values of the predicted activity labels, the personalization and adaptation process needs to be performed again, and therefore, user needs to label some of the instances again to obtain models that are reliable when applied to sensor data measured from changed environment. Again, streaming data and labels obtained via user or the recognition model can be used to update and adapt models to to make the model accurate again.

In section present methods to obtain a personalized context-aware human activity recognition models that works like the method presented in Fig. 1.

3.1 Learn++

In this article, the personalization and learning from streaming data is based on incremental learning, and more specifically on Learn++ algorithm (Polikar et al. 2001). Learn++ is an ensemble method and the idea of this method is not to handle all the incoming streaming data measurements separately, instead, it gathers several measurements and processes them as chunks. For each chunk, one or several new weak base models are trained. The ensemble is updated by adding these to a set of previously trained weak base models through weighted majority voting as an ensemble model (Losing et al. 2018). Once the new base models are trained and added to the ensemble, the data chunk is not needed anymore. Therefore, old training data does not need to be stored to memory. Moreover, when more streaming data are obtained, a new chunk of data can be gathered, more weak base models can be trained based on it and added to the ensemble.

Learn++ is one of the many algorithms that can be used for incremental learning. It was decided to be used in this study as in Losing et al. (2018), it is shown that Learn++ is one of the most accurate algorithm for incremental learning, but still it is less complex than many other methods. Moreover, it has been shown that ensemble methods perform really

well when applied to human activity recognition problems (Subasi et al. 2018, 2019). Therefore, Learn++-based classification models are ideal to be implemented into devices with low memory and calculation capacity such as wearable sensors. Yet another big boon of Learn++ is that it can use any classifier as the base classifier. Therefore, for this study, it was natural to choose base classifiers that are commonly used in the previous human activity recognition studies. The selected base classifiers are (1) linear discriminant analysis (LDA), (2) quadratic discriminant analysis (QDA), and (3) classification and regression tree (CART). These all three suitable for devices with low memory and calculation capacity, as they are really simple and light classifiers.

3.2 Personalizing human activity recognition models and the importance of human inputs

A method to personalize human activity recognition model based on incremental learning was first introduced in Siirtola et al. (2018), and an improved version the model, based on

human AI collaboration was presented in Siirtola and Röning (2019), see Fig. 2.

The idea of the method is to first train user-independent activity recognition models and add these to the Learn++-based ensemble model (Phase 1). This is initially used in the recognition process when user installs the application and starts to use it. However, when the user starts to use the application, personal streaming data is obtained. This can be used to personalize the recognition model.

In the Phase 2, streaming data is gathered and label for it is predicted using the ensemble model. The posterior value of the prediction is has an important role. If the posterior is above some pre-defined threshold th, the predicted label and the data related to it, can be used to update the ensemble model. On the other hand, if the posterior is below the threshold th, the prediction is considered as uncertain and user needs to label it before it is used in the model updating process. Moreover, as the studied human activities are of long-term (Siirtola et al. 2011), it valid to assume that also the data windows right before and after the user labeled
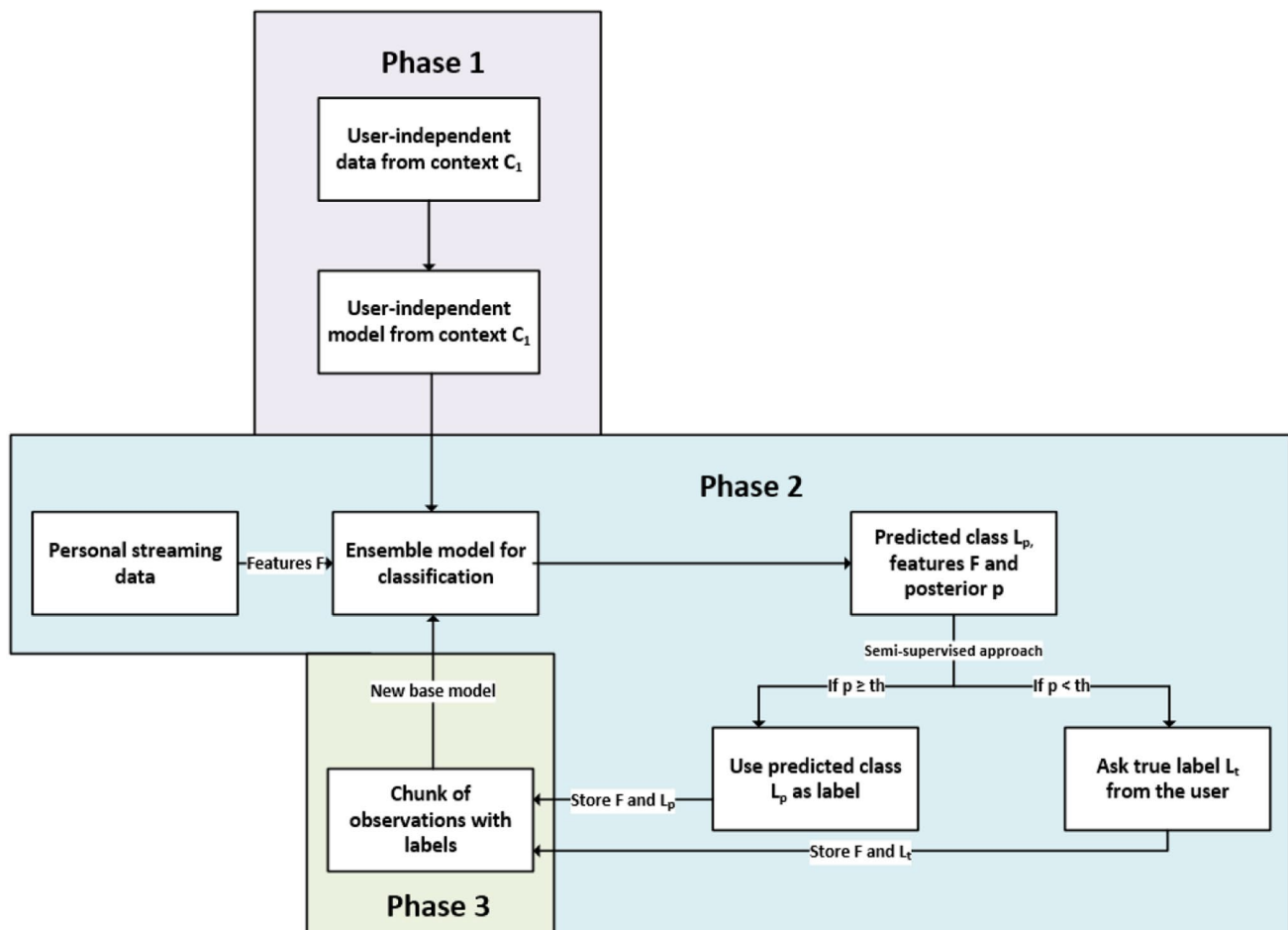

Fig. 2 A method to personalize human activity recognition models using incremental learning based on human AI collaboration (Siirtola and Röning 2019). The model works only in one context

observation have the same label as the observation labeled by the user. Due to this, it is assumed that user labeled observation and two data windows right before and after belong to the same class. Therefore, one user input gives information to five data windows, and thus, increases the number of the correct labels used in the model updating process.

Finally, in the Phase 3, new personal models can be trained based on data set containing both user labeled and predicted labels. Therefore, the new base model is trained and added to the ensemble model. In fact, this phase can be performed several times for one data set, so more than one base model can be trained based on the data set and added to the ensemble. As a results of Phase 3, the ensemble model is personalized and it is not user-independent anymore. Therefore, when more streaming data is obtained, the activity recognition can be done using personal ensemble model, and the outcomes of this model can be used to further personalize the model.

It is shown in Siirtola and Röning (2019) that this method is really powerful as it out-performs the performance of the user-independent recognition model in the every tested scenario, and with every tested dataset. However, it is designed for applications where the sensor position is fixed, therefore, the methods presented in Siirtola and Röning (2019) can be used for instance to create smartwatch application, as the body position of the watch is always the same, but they are not the best option when creating smartphone application, as phone can be positioned in more than one body position.

Moreover, the correct labels for human activities cannot be measured, they can only be observed directly. Therefore, the correct labels for the training data can only be obtained from the user, and in order to not disturb user too often, mostly the labels of training data are predictions instead of true labels. Thus, when combining the results of the base models, in Siirtola and Röning (2019) equal weight was given to base models. Other option would have been to give them weights based on the training data that can contain false labels and is not fully reliable. Equal weights reduce the possibility learn wrong things due to false labels, and therefore, the possibility to start suffering from the concept drift. However, it also means that models cannot quickly recover from concept shift, as the new base models, trained based on the data after concept shift have the same weight as the base models trained based on data obtained before the concept shift. Therefore, this method cannot be used in situations where context can change.

3.3 Context-aware personalization method

This article introduces a novel context-aware personalization method for human activity recognition and it is presented Fig. 3. This methods aims to fix the weaknesses of the method presented in Fig. 2. Firstly, the proposed method is context-aware, and therefore, the body position of the sensor is not fixed. Secondly, it can better react to the changes and new situations by recovering more rapidly from concept shift than the previous version of the personalized human activity recognition model.

The novel approach to build context-aware recognition models is divided into three phases, see Fig. 3. In the Phase 1, the recognition of streaming data is based on one or multiple user-independent models, for each context own user-independent base models are trained and added to the ensemble. Initially, each user-independent base model has equal weight.

When personal streaming data is obtained (Phase 2), similar to the approach presented in the previous subsection, the data is divided into windows and features $\mathbf{F}$ are extracted from them. The user-independent ensemble is used to predict class label for these windows. Again, if the posterior value of the predicted label L_p is above the pre-defined threshold th, the predicted label is considered as the correct label of the window. These observations having reliable predicted label are collected as a own separate chunk $\bigcup L_p$. On the other hand, if the posterior is below th, the prediction is considered as unreliable and user is asked to label the observation. Moreover, the data windows right before and after the user labeled window are considered to have the same class label. These user-labeled observations are collected as a own separate chunk S_t. After obtaining several observation, some having predicted labels and some user-defined labels, in Phase 3, new personal base models are trained based on combined data chunk $\bigcup L_p \cup S_t$. When a new base model is trained, it is added to the ensemble model. Several new base models can be trained from $\bigcup L_p \cup S_t$, and all of these are added to the ensemble.

The next step is unique to the proposed method, and was not included to the previous version of the personalization method making the proposed method novel: a unique weight is defined to each base model. In fact, to make the recognition process context-aware, new weight vector $\mathbf{w}$ is defined containing new weight w_i for each base model m_i, old and new, personal and user-independent, of the ensemble based on their performance on S_t which contains all the user labeled observations. As all of the data of S_t are user labelled, the labels of this dataset can be considered as reliable. Therefore, by using this set of data to define weights for the base models, it is possible to obtain a realistic, non-bias, estimation of the performance of each base model, and therefore, the risk to start suffering from the concept drift because of relying on mislabelled data is low. As a results of Phase 3, the ensemble model is now personalized and context-aware. It can give bigger weight to base models that are performing the best is the studied context. Therefore, when more streaming data is obtained, the activity recognition can be done using personal ensemble model, and the
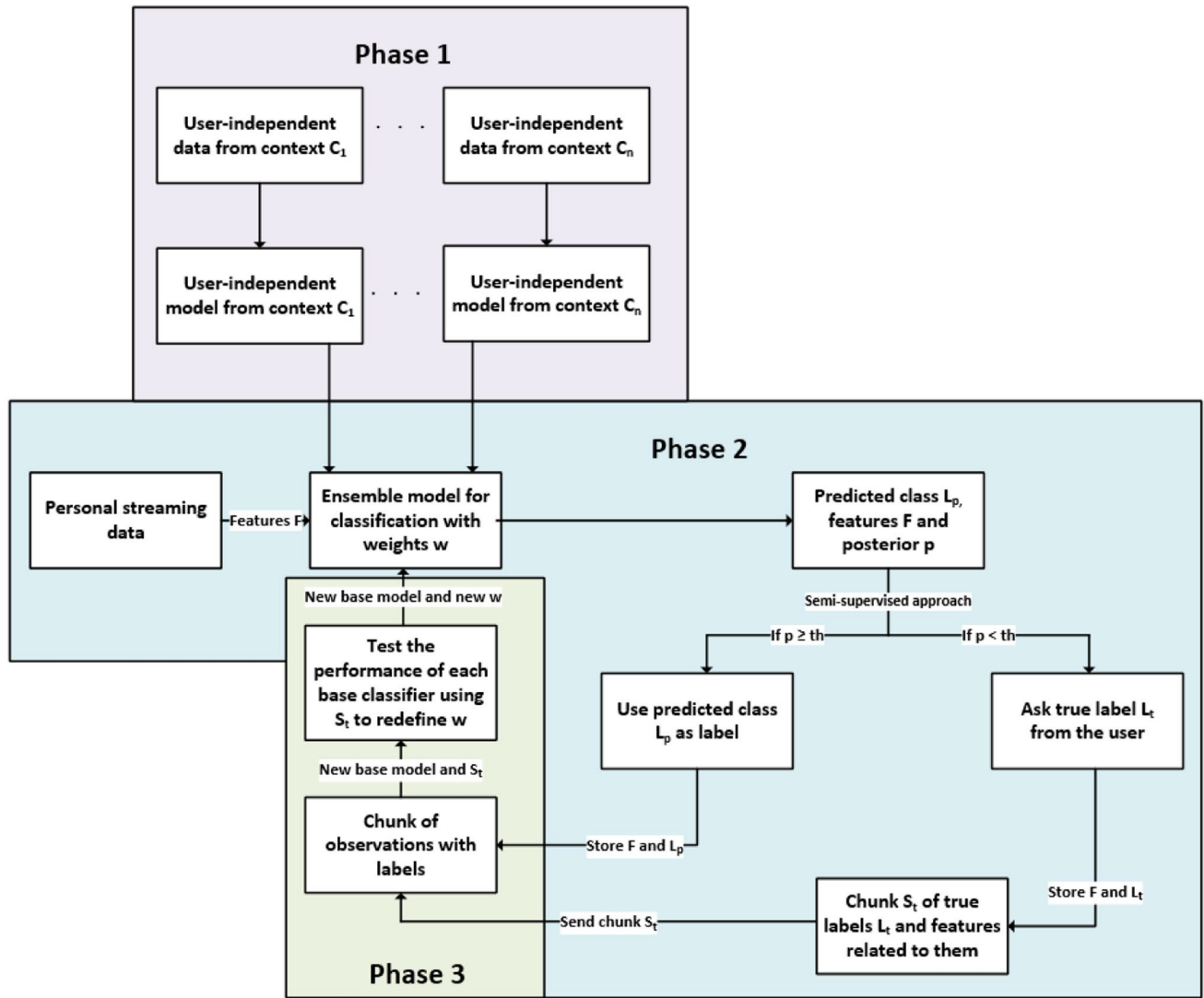
Fig. 3 A novel context-aware incremental learning-based method for human activity recognition gives weights for the base models based on their performance on the latest user-labeled data

outcomes of this model can be used to further personalize the model. Thus, the proposed method supports life-long learning. Moreover, weights for the base models are redefined after each updating round, and they are not depended on the history data. In fact, by weighting the models only based on their performance on the latest user-defined labels, and data related to them, enables fast adaptation to changes in context and faster adaptation to user's personal style of performing activities.

Weight w_i for base model m_i is calculated using equation

$$\left(log\frac{1}{lim}\right)^2, \tag{1}$$

where $lim = D/(1 - D + \epsilon)$, and D is the sum of the weights of the incorrectly classified instances and $\epsilon > 0$. In this study, equal weight $1/N$, where N is the total number of instances, is given to each observation. In the original Learn++ article, the weights for the base models are given in the same way but without raising the log-value to the power of 2. However, it this article, it was decided to raise the log-value to the power of 2 to underline the importance incorrectly classified instances, and this way to avoid concept drift caused by inaccurate base models. Moreover, in the original Learn++ article (Polikar et al. 2001) weights for the models of the ensemble were defined based on their performance in the training data when correct labels for all the instances were available. However, as already mentioned, similar approach cannot be used in this article, as the the phenomenon studied

in this article can only be observed indirectly. Therefore, the labels for the training data can only be obtained from the user, and in order to not disturb user too often, mostly the labels of training data are predictions instead of true labels. Moreover, the weighting method presented in Polikar et al. (2001) was dependent on the history data, unlike the method presented in this article. Therefore, it cannot handle changing situations as fast as the presented approach.

Therefore, the key element to obtain these context-awareness is weighting which is based on user labeled observations. Thus, user labeled instances have two roles in the method presented in Fig. 3. In Siirtola and Röning (2019) it was shown that already a small number of user-labeled

instances has a huge positive impact to the recognition rates. Therefore, to improve our previous study, this study uses the same chunk of user-labeled observations not only to improve the quality of the labels, but also to define unique weights to the base models, and this way adds context-awareness and an ability to fast react to the changing situations to the system.

4 Experimental setup and results

Experimental setup used in this article is presented in Fig. 4. For the experiments, user-independent models are trained using inertial data from three contexts: arm, wrist and waist
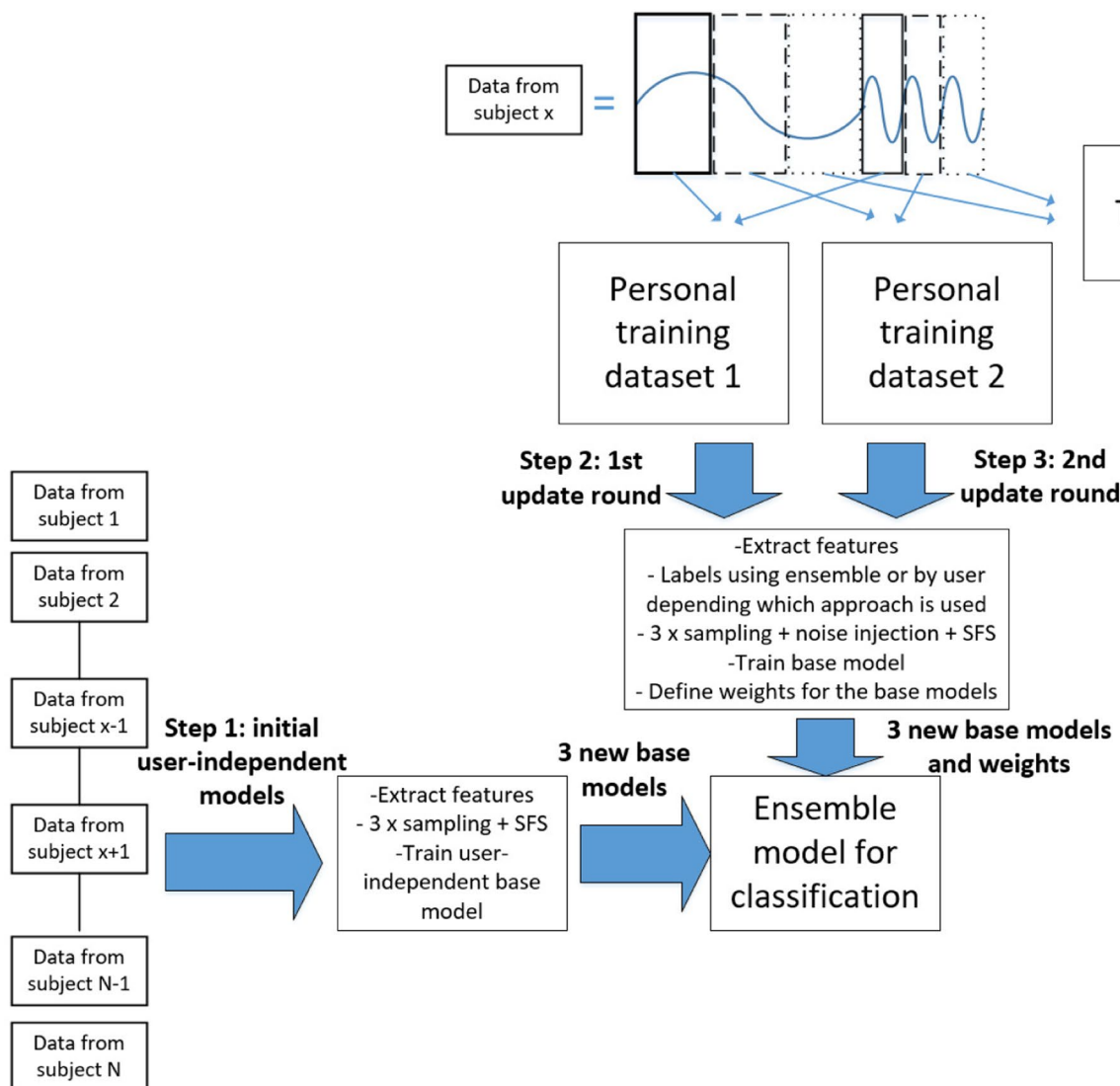

Fig. 4 Experimental setup used in this study. Each person's data is divided into three three chunks and each chunk contains the same amount of data from each activity. The first two chunks are used for personalizing the model and the last chunk is used only for testing

the ensemble model accuracy. In the experiments of this study, 10% of the personal training data was randomly selected to be labeled by the user

positions. Three user-independent base models are trained from each context, and training data set to train these base models is randomly selected from the pool containing all the data from certain body position. The ensemble model containing these nine user-independent models base models is initially used in the recognition process. Moreover, each person's x data are divided into three chunks and each chunk contains the same amount of data from each activity. When a user-independent model is personalized for subject x, two chunks of the data are used for personalizing the model and the last chunk is used for testing. Therefore, the third chunk is only used to test the performance of recognition model, and it is not used in any point of training or updating the base models. The two personal training data sets are small, and therefore, to avoid over-fitting, the noise injection method (Siirtola et al. 2016) to increase the size and variation of the data set is applied to the data set before training personal base models. The observations used in the model training process are selected from the noise injected training data set the using random sampling (in this case, sampling with replacement), and sequential feature selection [SFS (Bishop 2006)] is applied to it to select the best features for the selected observations. This sampling, feature selection and training phase is performed three times for one data set. Therefore, eventually the ensemble contains six personal and nine user-independent base models.

Three types of models were compared: traditional static user-independent model trained using leave-one-subject-out method and data from three body positions (arm, waist and wrist), non-weighted personalized model that gives equal weights for each base model (Siirtola and Röning 2019), and the proposed method for personalized context-aware models. According to Fig. 3, it is decided based on the posterior values of the predicted labels which instances needs to be labeled by the user. However, the obtained posterior values are dependent on the used personalization method. Therefore, to avoid the bias caused by that and to better compare personalization methods and to see the effect of weighting, instead of selecting user-labeled instances based on posteriors, 10% of the personal training data was randomly selected to be labeled by the user. However, as the studied activities are of long-term, it was assumed that the instances right before and after the user-labeled window have the same label as the window labeled by the user. Due to this, the total ratio of user-labeled instances was around 40%.

The experiments were performed in two scenarios: (S1) adapting models to a known context, and (S2) adapting models to a previously unknown context. In both scenarios, the models had to also adapt to the data of previously unknown person, as the initial user-independent dataset did not include any data from the studied user. In S1, aim is to show that it is possible build personalized models that are context-aware: initial base models are trained using data from arm, wrist, and waist positions (three user-independent base models from each position), and personal data is also coming from one of these positions. In S2, the same initial base model set is used, but the aim is to adapt to a totally new context, which in this case is the pocket-position. Therefore, in this scenario, the initial base model set does not include any information from the source context, and therefore, the context and the body position of the test data is different to the context of the training data of initial base models.

The results from scenario 1 are shown in Figs. 5 and 7 shows the results from scenario 2. In these figures, y-axis shows the error rate of the model, using balanced accuracy as the performance metric, and x-axis is related to the number of base models, $x = 1$ shows the error rate using only user-independent base models, and starting from $x = 2$ it is shown how personalization effects to error rate when new personal base models are added to ensemble one by one. Eventually ($x = 7$), the ensemble contains six personal and nine user-independent base models. Similarly, the results from scenario 1 using F1 score as a performance metric are shown in Figs. 6 and 8 shows the same results from scenario 2. Moreover, Learn++ contains random elements, and therefore, is does not provide the same results on each run. Due to this, the classification was performed five times using non-weighted and the proposed method. The results did not have much variation between the runs, and the presented results are average error rates and F1 scores from these five runs. In addition, Tables 1 and 2 compares the error rates of different approaches, and how much the performance of the approaches vary between the study subjects.

5 Discussion

When the results from scenario 1 are studied, the first thing that can be noted from Figs. 5 and 6 is that in many cases the initial Learn++ model containing user-independent base models from three body positions ($x = 1$) performs badly compared to the static user-independent model. This is due to trying to classify instances using a group of models where most are not trained using data from the same body position from where the test data originates. The static user-independent model does not perform as badly, as it is trained using a combination of data from all three contexts.

While the performance of the initial ensemble model is not good, it can be noted from Fig. 5 that the error rate quickly decreases when personal base models are added to the ensemble, and eventually, in each context better results are obtained using the proposed context-aware personalization method than using user-independent model (Table 1). Especially the difference is big when arm and waist positions are studied. However, when wrist-position is studied, the difference between the proposed method and
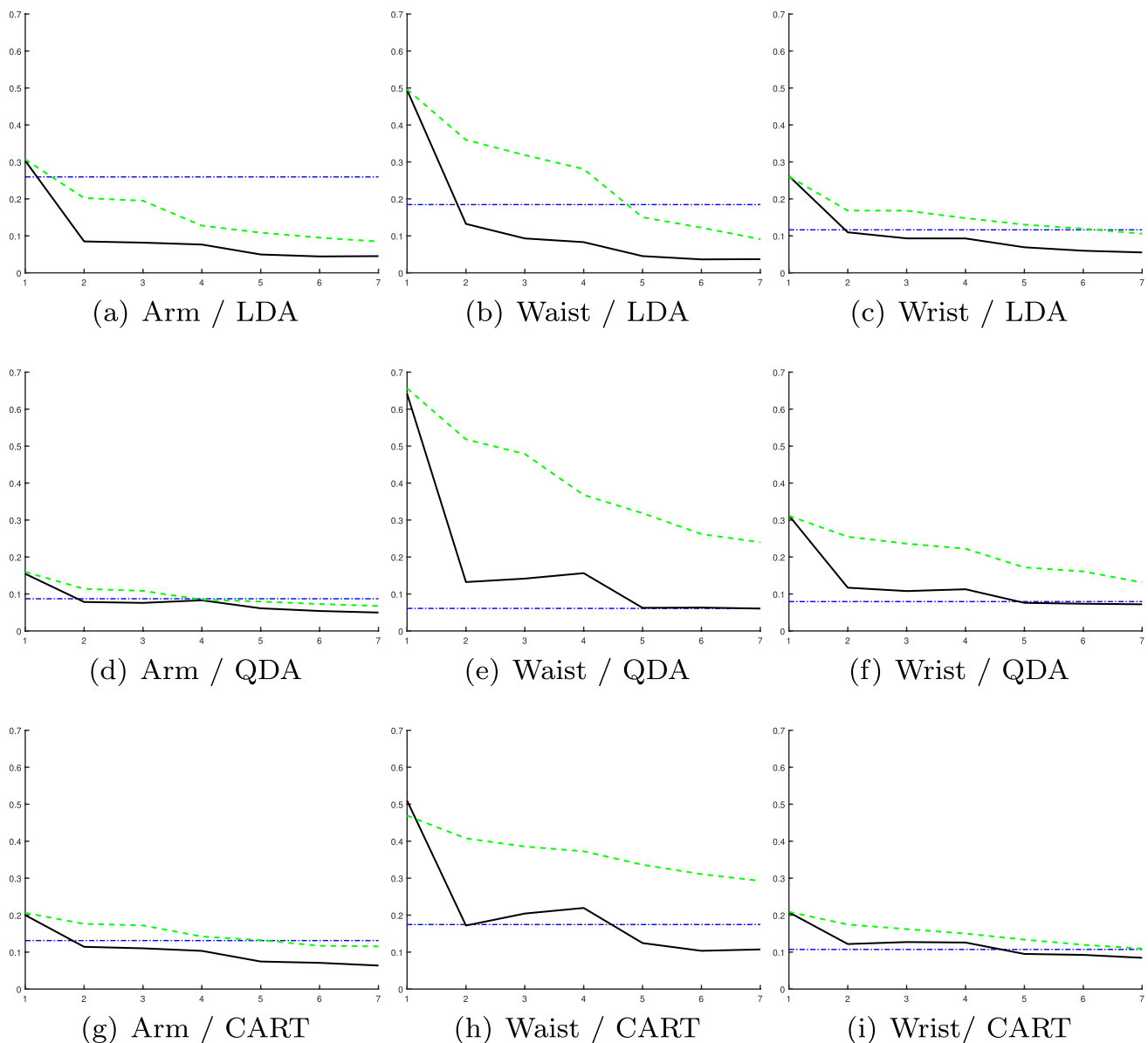
Fig. 5 Results from scenario 1. Adding of new base models to Learn++ decreases the error rate when balanced accuracy is used as a performance metrics. The error rate is shown in *y*-axis and *x*-axis is related to the number of used base models. $x = 1$ shows the error rate using user-independent base models (three from arm, wrist and waist body positions), and starting from $x = 2$ it is shown how personalization effects to error rate. User-independent results are shown using dash-dotted horizontal line, non-weighted using dashed line, and the proposed method using solid line. The proposed method out-performs the user-independent and non-weighted personalization methods

the static user-independent model is much smaller. Apparently, the the data from wrist-position is easier to classify than the data from other two positions, and therefore, the static user-independent model performs so well in this context, that the results cannot be improved much. When all three context are studied, on average, in scenario 1 the error rate using user-independent model is 13.3%, using non-weighted personalization method 13.8%, and using the proposed method 6.4%. According to the paired *t* test, this improvement is statistically significant in both cases.

Moreover, the context-aware personalization method is in each case better than the non-weighted personalization method. In addition, F1 score supports this finding (Fig. 6). In fact, in scenario 1, F1 score obtained using the proposed method is in each case much higher than the F1 scores obtained using user-independent and non-weighted personalization method. Moreover, according to Table 1, the standard deviation of balanced accuracies between the study subjects is much lower using the proposed method than using the other two rival methods.
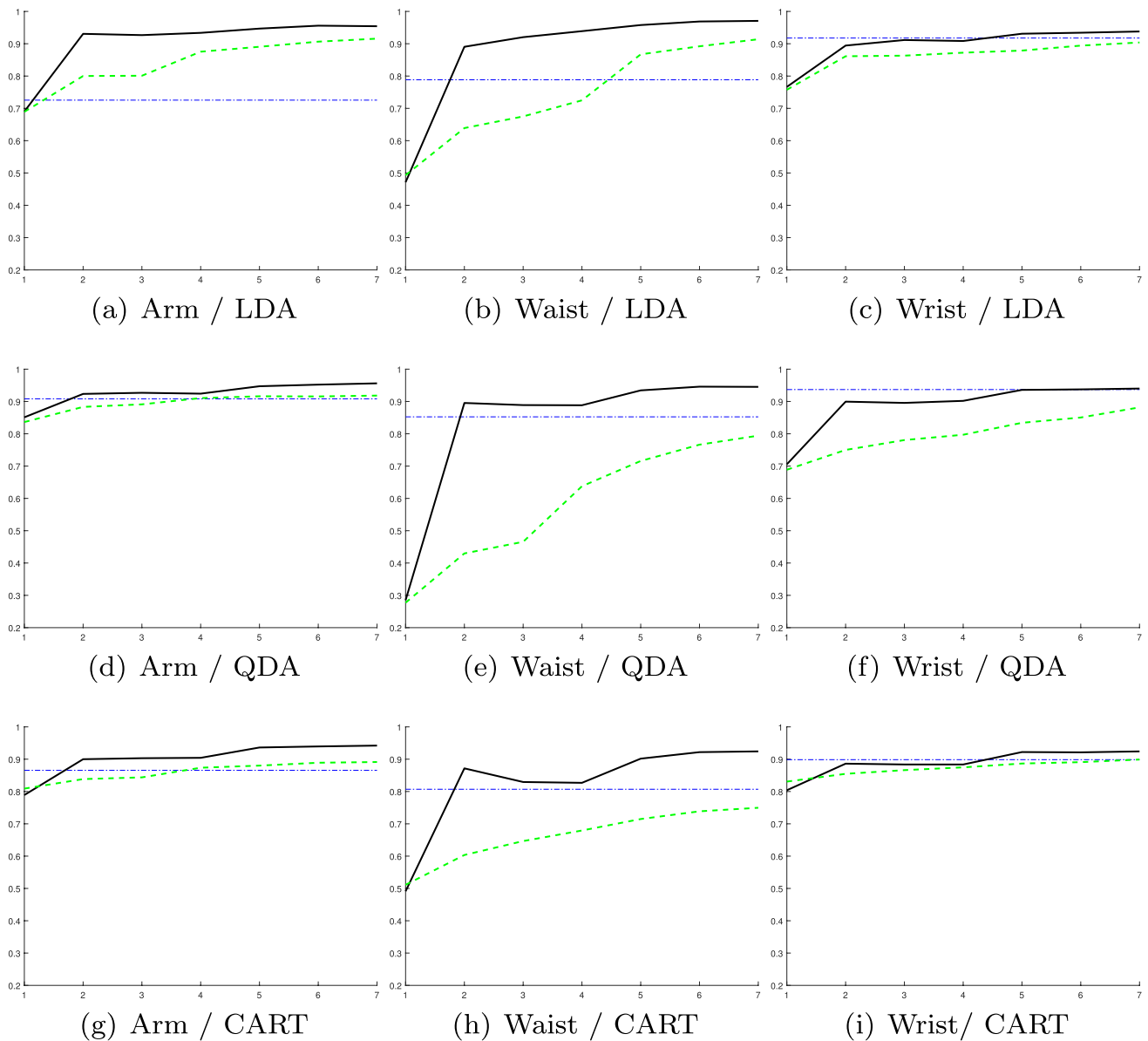
Fig. 6 Results from scenario 1. Adding of new base models to Learn++ improves F1 score. Score is shown in *y*-axis and *x*-axis is related to the number of used base models. *x* = 1 shows the error rate using user-independent base models (three from arm, wrist and waist body positions), and starting from *x* = 2 it is shown how personali- zation effects to error rate. User-independent results are shown using dash-dotted horizontal line, non-weighted using dashed line, and the proposed method using solid line. The proposed method out-performs the user-independent and non-weighted personalization methods

In Table 3, it is shown how weighting gives different weights to the base models in scenario 1. Table 3 shows that in each context, with every tested base classifier, the user-independent base models trained using data from the same body position as the test data gets bigger weight in the classification process than user-independent models from other positions. This shows that context-awareness works, and model can recognize from which body position the test data originates. In fact, as a secondary result, it can be noted that the proposed method can be used to detect the body

position of the sensor, and therefore, it can recognize the context. Moreover, Table 3 shows that the importance of user-independent models decreases when model is updated and more personal models are added to the ensemble.

Results from scenario 2 (see Figs. 7, 8) show that, again, initial user-independent base models for non-weighted and proposed personalization do not perform well. However, what also can be noted is that static user-independent model trained using data from arm, waist, and wrist positions performs poorly when test data is obtained from context

Table 1 Average performance of different approaches in scenario 1

	User-Independent	Non-weighted	Proposed
Arm/LDA	26.0 (7.2)	8.5 (7.2)	4.5 (2.8)
Waist/LDA	18.5 (8.7)	9.1 (9.0)	3.7 (3.2)
Wrist/LDA	11.6 (8.3)	10.6 (6.9)	5.5 (4.5)
Mean/LDA	18.7 (8.1)	9.4 (7.8)	4.6 (3.6)
Arm/QDA	8.7 (6.9)	8.7 (6.8)	5.0 (5.0)
Waist/QDA	6.1 (3.5)	24.0 (11.8)	6.1 (3.9)
Wrist/QDA	7.9 (4.1)	13.1 (6.6)	7.2 (4.9)
Mean/QDA	7.6 (5.1)	14.6 (8.7)	6.6 (4.6)
Arm/CART	13.1 (8.7)	11.6 (9.4)	6.4 (6.5)
Waist/CART	17.5 (8.5)	29.3 (19.6)	10.7 (6.1)
Wrist/CART	10.7 (6.1)	11.0 (7.3)	8.5 (5.2)
Mean/CART	13.8 (7.9)	17.3 (13.2)	8.5 (6.0)
Mean/All	13.3 (7.2)	13.8 (10.2)	6.4 (4.8)

Numbers show the percentual error rate of different approaches in different contexts when the performance is measured using balanced accuracy. Standard deviation between the study subjects in parentheses

which is not included to the training data set. Moreover, the results from this scenario underlines the effectiveness of the proposed method in situations where context changes. On average, in S2, the error rate of the user-independent model, using balanced accuracy as the performance metric, is 36.6%, non-weighted personalization method 36.9%, and the proposed method 14.1% (see Table 2). The proposed method performs especially well when LDA is used as a base classifier, in this case the error rate is only 9.2% when balanced accuracy is used as a performance metric. In addition, F1 score shows that accurate personalized models can be obtained also when the models are adapted to previously unknown context (Fig. 8). However, while balanced accuracy and F1 score of the proposed method show a huge

improvement compared to other methods, still there is a lot variation in the results between the study subjects (Table 2). On the other hand, still the variation between the subject is smaller using the proposed method than using the rival methods.

By comparing the performance of the different classifiers, it can be noted that the proposed methods proposed better results than the other two methods no matter which base classifier is used. However, especially well the proposed method performs when base models are LDA-based. In fact, the best performance in both scenarios, and in fact, in every studied context are obtained using the proposed context-aware method and LDA as a base classifier. Moreover, according to the results, CART is less accurate than QDA and LDA. The difference is not that big in scenario 1 but in scenario 2 the performance of CART is bad. In addition, in S1 there is no big difference between the static user-independent model and the proposed method when using QDA. However, this is not the case in S2 where the proposed method performs much better than the user-independent model as in this case user-independent model does not include any training data from the target context.

The comparison of the results from scenarios 1 and 2 shows that the proposed method performs especially well in scenario 2, where the aim is to adapt to a totally new context. The reason for this is that in scenario 2, the initial user-independent base models are more inaccurate than in scenario 1. This means that in scenario 2 it is more important to give much bigger weights to personal base classifiers trained using data from new context (and smaller weights to user-independent base models trained using data from other contexts) than in scenario 1 in order to obtain reliably results. In fact, the more different the data obtained from the next context is compared to the data from obtained from known contexts is, the more important the weighting is.
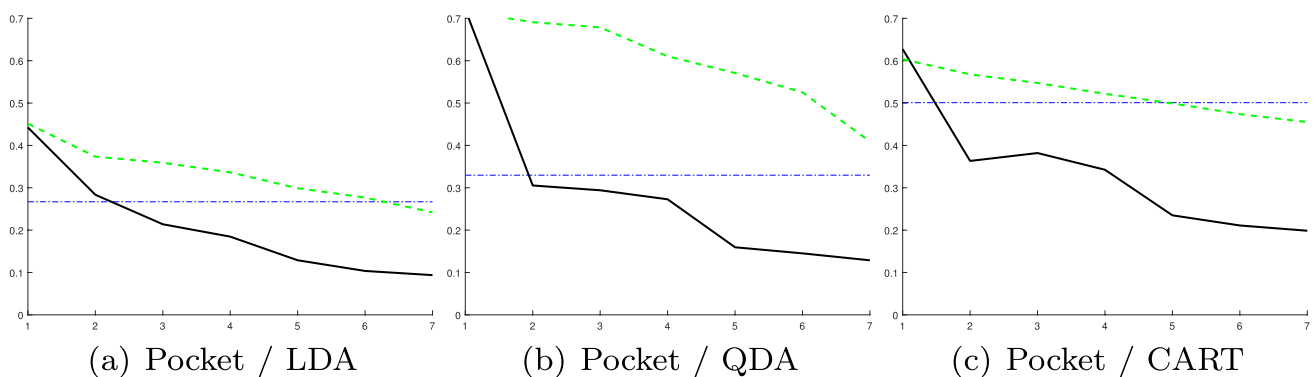

(a) Pocket / LDA (b) Pocket / QDA (c) Pocket / CART

Fig. 7 Results from scenario 2. Adding of new base models to Learn++ decreases the error rate when balanced accuracy is used as a performance metrics. The error rate is shown in y-axis and x-axis is related to the number of used base models. $x = 1$ shows the error rate using user-independent base models (three from arm, wrist and waist body positions), and starting from $x = 2$ it is shown how personalization effects to error rate. User-independent results are shown using dash-dotted horizontal line, non-weighted using dashed line, and the proposed method using solid line. The proposed method out-performs the user-independent and non-weighted personalization methods
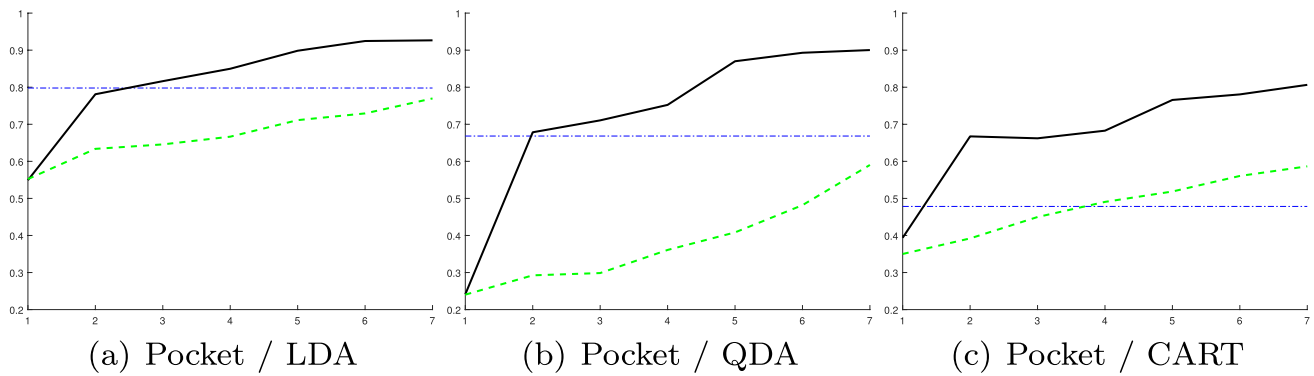
(a) Pocket / LDA   (b) Pocket / QDA   (c) Pocket / CART

Fig. 8 Results from scenario 2. Adding of new base models to Learn++ improves F1 score. Score is shown in *y*-axis and *x*-axis is related to the number of used base models. $x = 1$ shows the error rate using user-independent base models (three from arm, wrist and waist body positions), and starting from $x = 2$ it is shown how personali-zation effects to error rate. User-independent results are shown using dash-dotted horizontal line, non-weighted using dashed line, and the proposed method using solid line. The proposed method out-performs the user-independent and non-weighted personalization methods

Table 2 Average performance of different approaches in scenario 2

	User-Independent	Non-weighted	Proposed
Pocket/LDA	26.7 (11.0)	24.2 (13.2)	9.4 (9.9)
Pocket/QDA	33.0 (12.8)	40.9 (11.8)	12.9 (7.8)
Pocket/CART	50.1 (15.3)	45.5 (11.9)	19.9 (10.5)
Mean	36.6 (13.2)	36.9 (12.3)	14.1 (9.5)

Numbers show the error rate percentage of different approaches in different contexts when the performance is measured using balanced accuracy. Standard deviation between the study subjects in parentheses

Table 3 Percentages showing the weights of different types base models in the ensemble in S1

Target/update round	Type of base model			
	U-I arm	U-I waist	U-I wrist	Personal
Wrist/1st update	32/28/31	14/14/15	54/62/54	–/–/–
Wrist/2nd update	19/17/20	8/10/10	34/38/37	40/36/33
Waist/1st update	16/13/18	67/73/63	16/13/15	–/–/–
Waist/2nd update	8/9/11	38/46/42	10/9/13	44/36/34
Arm/1st update	60/60/67	13/12/24	27/29/33	–/–/–
Arm/2nd update	39/41/36	8/6/12	18/17/19	35/36/33

The numbers show how much weight different base classifiers (LDA / QDA / CART) give to different types of base models. Base models trained using data from the target position have bigger weight in the classification process than models from other body positions. Moreover, weights of the user-independent (U-I) models decrease when ensemble is updated and personal base models are added to ensemble

The results discussed in this section show that the proposed method for personalizing and adapting human activity recognition models significantly improves the recognition rates in scenarios where the context changes. When the results of Learn++ ensemble are combined through weighted majority voting, with the help of user labelled observations, the method can give bigger weights to base models that are performing well in the context in question. Moreover, the results show that the benefit of using the proposed method is the biggest in contexts where static user-independent model is inaccurate.

6 Conclusions

This study presented a novel context-aware method to personalize human activity recognition by learning from streaming data. It is based on incremental learning method called Learn++, and what makes it novel is that it works in several contexts, and adapts to changes in context and user. This is done by weighting the most suitable base models based on their performance on the latest user-labeled data, and therefore, it gives bigger weights to those base models which perform the best in the current context. Thus, the weighting ensures that the best base models for the current situation and context always have bigger weight in the prediction process than other base models. This novel proposed method was compared to non-weighted personalization method and to static user-independent model. The experiments were performed in two scenarios: (S1) adapting model to a known context, and (S2) adapting model to a unknown context. In both scenarios, the models had to also adapt to the data of previously unknown person, as the initial user-independent dataset did not include any data from the studied user. Moreover, three different base classifiers (LDA, QDA, and CART) were compared. The results show the superiority of the proposed method. It performs better than the other two methods in both scenarios and no matter which base classifier is used. In S1, the error rate using static user-independent model was 13.3%, using

non-weighted personalization method 13.8%, and using the proposed method 6.4% when balanced accuracy is used as a performance metric. Moreover, the results show that weighting works: the proposed methods gives bigger weights to base classifiers that are trained using data from the same context as the test data than to base classifiers trained using data from other contexts. The performance of the proposed method is also good in S2. Tn S2, the error rate of user-independent model was 36.6%, non-weighted personalization method 36.9%, and the proposed method 14.1%. In addition, F1 score supports this finding in both scenarios. Thus, as the main result, it was noted that it is possible to build personalized models that are context-aware, and can adapt to different contexts and users by weighting base models that are the most relevant to the studied context. Moreover, as a secondary result, it was noted that the proposed method can be used to detect the context, and therefore, the body position of the sensor.

Future work includes more experiments with multiple datasets. In fact, the dataset used in this this study is quite short, and therefore, it does not contain as much variation as a long real-life dataset includes. Thus, to show the true potential of the proposed method, a long real-life dataset from several study subject should be collected for the experiments. In addition, long real-life dataset could be used to study how to avoid uncontrolled growth of the ensemble size. This could be done for instance by using only the latest base models in the recognition process. Moreover, different ensemble methods should be experimented and compared to find the one that produces the best results. In addition, in this article three base classifiers were compared, future work includes experimenting with other base classifiers as well, including SVM and Naive Bayes.

The main weakness of the proposed method is that when new base classifiers are trained, the training data needs to include measurements from each of the studied activity. This means that if the study subject does not perform all the activities, new base classifiers cannot be trained, and therefore, the model cannot be personalized. Therefore, one important aspect of the future work is to solve this problem and find a way to personalize model without a need to perform all the activities.

References

Abdallah ZS, Gaber MM, Srinivasan B, Krishnaswamy S (2015) Adaptive mobile activity recognition system with evolving data streams. Neurocomputing 150:304–317

Akbari A, Jafari R (2020) Personalizing activity recognition models with quantifying different types of uncertainty using wearable sensors. IEEE Trans Biomed Eng 67(9):2530–2541

Albert M, Toledo S, Shapiro M, Kording K (2012) Using mobile phones for activity recognition in Parkinson's patients. Front Neurol 3(158):1–7

Almaslukh B, Artoli AM, Al-Muhtadi J (2018) A robust deep learning approach for position-independent smartphone-based human activity recognition. Sensors 18(11):3726

Bishop CM (2006) Pattern recognition and machine learning (information science and statistics). Springer, New York Inc, Secaucus

Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput Surv 46(3):1–33

Cvetković B, Kaluža B, Gams M, Luštrek M (2015) Adapting activity recognition to a person with multi-classifier adaptive training. J Ambient Intell Smart Environ 7(2):171–185

Fallahzadeh R, Ghasemzadeh H (2017) Personalization without user interruption: boosting activity recognition in new subjects using unlabeled data. In: Proceedings of the 8th international conference on cyber-physical systems. ACM, pp 293–302

Ferrari A, Micucci D, Mobilio M, Napoletano P (2020) On the personalization of classification models for human activity recognition. IEEE Access 8:32066–32079

Figueira C, Matias R, Gamboa H (2016) Body location independent activity monitoring. In: BIOSIGNALS, pp 190–197

Fujinami K, Kouchi S (2012) Recognizing a mobile phone's storing position as a context of a device and a user. In: International conference on mobile and ubiquitous systems: computing, networking, and services. Springer, pp 76–88

Garcia-Ceja E, Brena R (2015) Building personalized activity recognition models with scarce labeled data based on class similarities. In: García-Chamizo JM, Fortino G, Ochoa SF (eds) Ubiquitous computing and ambient intelligence. sensing, processing, and using environmental information. Springer International Publishing, Cham, pp 265–276

Gepperth A, Hammer B (2016) Incremental learning algorithms and applications. In: European symposium on artificial neural networks (ESANN), pp 357–368

Incel OD, Kose M, Ersoy C (2013) A review and taxonomy of activity recognition on mobile phones. BioNanoScience 3(2):145–171

Losing V, Hammer B, Wersing H (2018) Incremental on-line learning: a review and comparison of state of the art algorithms, vol 275. Elsevier, Amsterdam, pp 1261–1274

Mannini A, Intille SS (2019) Classifier personalization for activity recognition using wrist accelerometers. IEEE J Biomed Health Inform 23(4):1585–1594

Mo L, Feng Z, Qian J (2016) Human daily activity recognition with wearable sensors based on incremental learning. In: International conference on sensing technology, pp 1–5

Munguia Tapia E (2008) Using machine learning for real-time activity recognition and estimation of energy expenditure. Ph.D. thesis,

Ph.D. dissertation, Department of Architecture at the Massachusetts Institute of Technology, Cambridge, MA, USA

Ntalampiras S, Roveri M (2016) An incremental learning mechanism for human activity recognition. In: IEEE symposium series on computational intelligence, pp 1–6

Polikar R, Upda L, Upda SS, Honavar V (2001) Learn++: an incremental learning algorithm for supervised neural networks. IEEE Trans Syst Man Cybern Part C (Appl Rev) 31(4):497–508

Roggen D, Förster K, Calatroni A, Tröster G (2013) The adarc pattern analysis architecture for adaptive human activity recognition systems. J Ambient Intell Humaniz Comput 4(2):169–186

Sato K, Fujinami K (2017) Active learning-based classifier personalization: a case of on-body device localization. In: 2017 IEEE 6th global conference on consumer electronics (GCCE), pp 1–2

Shi D, Wang R, Wu Y, Mo X, Wei J (2017) A novel orientation-and location-independent activity recognition method. Pers Ubiquit Comput 21(3):427–441

Shoaib M, Bosch S, Incel OD, Scholten H, Havinga P (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors 14(6):10146–10176. http://www.mdpi.com/1424-8220/14/6/10146

Siirtola P, Röning J (2019) Incremental learning to personalize human activity recognition models: the importance of human AI collaboration. Sensors 19(23):5151

Siirtola P, Koskimäki H, Röning J (2011) Periodic quick test for classifying long-term activities. In: 2011 IEEE symposium on computational intelligence and data mining (CIDM). IEEE, pp 135–140

Siirtola P, Koskimäki H, Röning J (2016) Personal models for ehealth-improving user-dependent human activity recognition models using noise injection. In: IEEE symposium series on computational intelligence (SSCI). IEEE, pp 1–7

Siirtola P, Koskimäki H, Röning J (2018) Personalizing human activity recognition models using incremental learning. In: European symposium on artificial neural networks, computational intelligence and machine learning (ESANN), pp 627–632

Subasi A, Dammas DH, Alghamdi RD, Makawi RA, Albiety EA, Brahimi T, Sarirete A (2018) Sensor based human activity recognition using adaboost ensemble classifier. Procedia Comput Sci 140:104–111. https://doi.org/10.1016/j.procs.2018.10.298. Cyber physical systems and deep learning Chicago, Illinois November 5–7, 2018

Subasi A, Fllatah A, Alzobidi K, Brahimi T, Sarirete A (2019) Smartphone-based human activity recognition using bagging and boosting. Procedia Comput Sci 163:54–61

Sztyler T, Stuckenschmidt H (2016) On-body localization of wearable devices: an investigation of position-aware activity recognition. In: 2016 IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 1–9

Wang Z, Jiang M, Hu Y, Li H (2012) An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. IEEE Trans Inf Technol Biomed 16(4):691–699

Weiss G, Lockhart J (2012) The impact of personalization on smartphone-based activity recognition. In: AAAI workshop on activity context representation: techniques and languages, pp 94–104

Widhalm P, Leodolter M, Brändle N (2019) Into the wild—avoiding pitfalls in the evaluation of travel activity classifiers. Human activity sensing: corpus and applications. Springer International Publishing, Cham, pp 197–211

Yu T, Zhuang Y, Mengshoel OJ, Yagan O (2016) Hybridizing personal and impersonal machine learning models for activity recognition on mobile devices. In: MobiCASE, pp 117–126