

RESEARCH

Open Access



CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments

Sunder Ali Khowaja, Bernardo Nugroho Yahya and Seok-Lyong Lee 

*Correspondence:

sllee@hufs.ac.kr

Department of Industrial
and Management
Engineering, Hankuk
University of Foreign Studies,
Global Campus, Yongin-Si,
South Korea

Abstract

The existing action recognition systems mainly focus on generalized methods to categorize human actions. However, the generalized systems cannot attain the same level of recognition performance for new users mainly due to the high variance in terms of human behavior and the way of performing actions, i.e. activity handling. The use of personalized models based on similarity was introduced to overcome the activity handling problem, but the improvement was found to be limited as the similarity was based on physiognomies rather than the behavior. Moreover, human interaction with contextual information has not been studied extensively in the domain of action recognition. Such interactions can provide an edge for both recognizing high-level activities and improving the personalization effect. In this paper, we propose the context-aware personalized human activity recognition (CAPHAR) framework which computes the class association rules between low-level actions/sensor activations and the contextual information to recognize high-level activities. The personalization in CAPHAR leverages the individual behavior process using a similarity metric to reduce the effect of the activity handling problem. The experimental results on the “daily lifelog” dataset show that CAPHAR can achieve at most 23.73% better accuracy for new users in comparison to the existing classification methods.

Keywords: Personalized activity recognition, Associative learning, Human behavior process, Machine learning, Pervasive sensors, Wearable sensors

Introduction

The effectiveness of modern context-aware computing systems is heavily based on its ability to recognize human behavior and physical activities. These computing systems cover a wide range of application areas such as smart healthcare, smart homes, elder care, and e-health [1]. With the aid of wearable and pervasive sensors, the human activity recognition area is getting more and more popular and has led to the development of numerous IoT and context-aware based systems. For instance, MyBehavior [2] provides personalized health feedback by taking into account human behavior (diet) and physical activities. There are generally two ways to acquire the sensor readings for behavior modeling and activity recognition. One way is to use the wearable and virtual sensors which

provide the inertial measurements such as accelerometer, gyroscope sensors, and social interaction data through smartphones and mobility patterns, respectively. Another way is to use the sensor activations which require a specific infrastructure for the humans to directly interact with the sensors such as smart switches, infrared sensors, and pressure sensors. Existing studies also use the interaction between the environment and the people, i.e. contextual information, to model the human behavior and recognize complex physical activities [3, 4]. In the broad sense, contextual information can be viewed in two perspectives; location, identity, activity, time and user & role, process & task [5]. In this study, the contextual information is referred to as the data obtained by the secondary sensors associated with the performed action such as the time of action, the location where the action is performed, or the interrelationship of action and the object.

Existing studies use the context information as one of the features for training the action classifier to improve the accuracy of the recognition system, but they are limited to low-level actions. In our study, a low-level action refers to the basic human action performed in daily routine such as standing, sitting, walking, running, lying, and ascending/descending stairs. However, these are the fundamental ones and therefore can be recognized with quite a reasonable accuracy by using only inertial sensors [6]. Contextual information has also been used to derive high-level activities as well as to model human behavior with an assumption that prior knowledge is available for the interaction between human action and the specified context [7]. High-level activity refers to the higher level of abstraction for the performed action such as eating, personal grooming, desk work, and more. The high-level activities can be inferred by combining low-level actions and contextual information. Single or multiple contexts can be used depending on the available infrastructure and the sensor characteristics. For instance, time/duration of the activity can be obtained using the sensor characteristics such as an accelerometer, whereas the indoor location and object sensors need a fixed infrastructure. The high-level activity recognition can be performed either using direct inferencing by inference rules [8], and ontological reasoning [9], or integrating the contextual information in the machine learning framework, i.e. adding the context as one of the feature vectors [10]. The human behavior in this study refers to the way or routine in which the actions are performed while interacting with the available contexts. Existing studies have used human behavior modeling extensively for determining individual behavior, health anomalies, and human identification [2, 11]. Human behavior modeling can be categorized into two schools of thought, i.e. social and applied sciences. The social sciences more deviate towards the impact caused by a certain behavior on society while the applied science field leverages the behavior to attain automation or derive recommendations. Previous studies have also combined the use of physical activity recognition and human behavior models to provide recommendations for healthy lifestyles, social interactions, and human identification [2, 11].

The physical activity recognition systems heavily rely on the use of machine learning techniques or statistical inferences such as class association rules (CARs). The integration of inference rules (CARs) in a machine learning framework is referred to as associative learning. The reason for using CARs is that they provide a model that is simple and is proven to be effective in terms of accuracy and interpretability [12]. The CARs fall in the category of association rules which consider the rule consequent as a fixed class

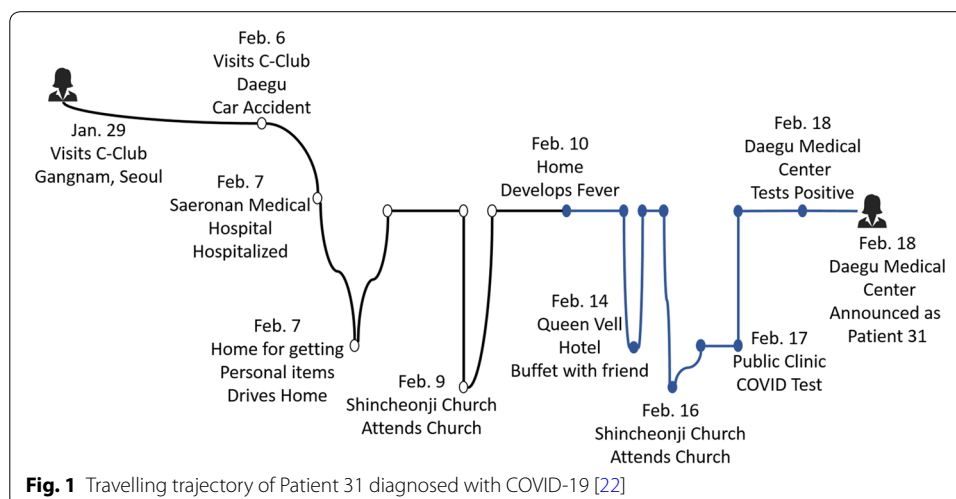
attribute [4]. The left side of the CARs could be multiple itemsets which are based on the conjunction of attribute-value pairs whereas the right side of the CARs corresponds to the target concept which refers to the activity class label in this study. These rules allow us to extract interesting patterns based on the attribute-value pairs concerning the target concept. The CARs in existing studies have been used with the sensor activations, i.e. an attribute-value pair, and without considering the contextual information [4, 12–15]. Furthermore, the CARs have not been exploited for personalized recognition services. The goal of human activity recognition (HAR) systems is to learn from annotations and acquired sensor readings to infer the performed activity. Most of the machine learning algorithms take into consideration the feature vectors which are either provided by the physical or contextual sensors. The problem is that the learning algorithm does not distinguish between the base and the contextual information; it explicitly trains the model. Therefore, if contextual information appears out of order, it will simply look up the weights assigned to a feature vector instead of considering the relationship between the sensor and the secondary information. Moreover, the shallow learning algorithms are either too complex such as random forests, and support vector machines, or considered to be “black boxes” like neural networks. Considering that the CARs are highly interpretable, they have competitive strength in the field of personalized activity recognition by providing inferential insights regarding human behavior. Hence, it is necessary to integrate the contextual information with the attribute-value pair of low-level actions to model the CARs for recognizing personalized activities.

There are two issues that we need to address on developing personalized models; activity handling problem and human behavior process. First, the activity handling problem, as aforementioned, is a phenomenon in which the subjects may perform actions differently with respect to multiple contexts (e.g., location, time, user, and identity information). The existing studies mainly focus on developing subject-independent models for activity recognition and are regarded as a one-fits-all activity model. However, the subject-independent models do not exhibit the same performance when applied to data from a new test subject due to high variation in terms of human behavior and the mannerisms for performing a specific action. The study [16] implies that the subjects are idiosyncratic; they interact differently with the context as per their style, behavior, and ability to perform different actions. It has also been proved that the subjects may perform actions differently with respect to the location and time of the day [2, 11, 17]. For instance, the “meal preparation” activity at home can be executed differently when performed at the office, or “housework” activity may vary in execution when performed in the morning time or at night. In this regard, researchers often move towards building personalized or subject-dependent activity models. Very few studies apply the personalization aspect due to the challenge of collecting and annotating large amounts of user-specific data [1].

Human behavior processes have been modeled in existing studies using Kalman Filters, crowd simulation, and soft computing techniques [18, 19]. The researchers in the field of process mining have also tried to model human behavior as a process for analyzing the interesting behavior patterns [20]. Process mining techniques represent the sensor measurements or samples from an indoor location system in process workflows for modeling human behavior [21]. Most of the works in process mining regarding human

behavior focus on improving the process workflows. Assuming that the users may have similarities in their behavior patterns, this study uses it for computing the similarity between the existing pool of models to select and test it against the data of a new user. However, assuming that some of the data from new users are available, a semi-population calibration approach can be used to test one of the trained models against the new user data. It maps a new individual to a pool of existing users to solve the “cold-start” problem. The cold-start problem concerns the personalized activity recognition for new users having less or no amount of annotated data. The mapping of the personalized models onto the new users in existing studies was performed using the gender and demographic details such as age, color, and region. However, gender and demographic details have limited knowledge to be leveraged, and they do not provide sufficient information to solve the activity handling problem. We assume that the behavioral characteristics of human activities can be modeled best using their respective behavior processes. Although many studies use physical activities to derive or identify the behavior process models, the reverse approach has not been explored for solving the cold-start problem in physical activity recognition studies. Intuitively, it makes sense as the similarity in behavior is more likely to map the action characteristics of one individual to another in comparison to age and gender. An implication of such a study is evident by the recent report from Reuters Graphics [22] which provided an insight into the trajectory of patient 31 as shown in Fig. 1. Provided with the limited information of the physical activity, time, and location, the system could have chosen a similar individual based on the behavioral patterns and predicted the location of the next activity, or as a precautionary measure, if an individual seems probably affected, the system could generate a notification to the locations where the subject tends to socialize or have a meeting. This shows the importance of using behavior patterns for recognizing human activities and the relevancy of this study with current situation.

In this study, we address two issues for personalized activity recognition systems. The first is the integration of the contextual information from multiple sensor modalities to improve high-level activity recognition. The second is the use of human behavior processes to solve the cold-start problem by mapping the new user



onto the existing pool of subjects. In this regard, we propose the context-aware personalized human activity recognition (CAPHAR) framework using associative learning and human behavior process models. The CARs in associative learning take into account the contextual information from different sensors such as location, object, and time along with the low-level action from IMUs or sensor activations to find the frequent patterns. These frequent patterns are then used to classify the high-level activities. The method also provides a generalized way of recognizing high-level activities from either low-level actions or sensor activations which has not been directly addressed in the existing studies. We demonstrate the effectiveness of associative learning using two real-world human activity datasets, i.e., “daily life-logging” [17] and “Activity Recognition in the Home” [23]. The former consists of low-level actions from IMUs, location, and time information, whereas the latter consists of sensor activations and the time information. We assume that the evaluation of these two datasets will prove the real-world applicability of our proposed method.

Furthermore, we generate the individual behavior process models of each user using process mining techniques from the recognized high-level activities. We propose a way to measure the similarity between the behavior model of a new user and the existing pool of behavior process models to perform personalized high-level activity recognition. The proposed framework is shown to work in an environment where we have a limited amount of annotated data from a new user. The contributions of our study are summarized as follows:

- We propose the CAPHAR framework for personalized HAR based on human behavior and contextual information.
- The CAPHAR framework can handle various sensing inputs such as low-level actions or sensor activations in a single framework.
- We propose the use of a semi-population calibration approach by computing the similarity between the human behavior process models for solving the cold-start problem.
- The classification accuracy is improved by leveraging human behavior using associative learning.

The remainder of the paper is structured as follows: Sect. “[Related works](#)” presents a summary of the related works. Section “[Proposed method](#)” explains the proposed methodology to recognize low-level actions using traditional machine learning methods, to recognize high-level activities using associative learning, to generate process models from the recognized high-level activities, and to compute the similarity for mapping the personalized models on to new users. Section “[Experiments and results](#)” briefly describes the daily life-logging dataset and activity recognition in the home database. The results are also presented in Sect. “[Experiments and results](#)” which first presents activity recognition results and then depicts the usefulness of the proposed work for personalized activity recognition. We conclude our study in Sect. “[Conclusion](#)” along with discussion and some future directions.

Related works

Association rule mining and Class association rules for activity recognition

The association rule is formally defined as $X \Rightarrow Y$, where X and Y are disjoint sets and are referred to as consequent and antecedent, respectively. One of the most used association rule mining algorithms is Apriori [24] which finds the frequent itemsets based on the user-specified minimum support and confidence. Another one is the FP-growth algorithm [25] which uses a tree-like structure to mine the frequent patterns. Association rule mining has been widely used on historical transaction data for predictive analysis [8].

Association rule mining has also been used as a feature extraction mechanism for many classification models. Chien and Chen [24] used the association rules as features to train the classification model for stock trading. Pach et al. [25] introduced the concept of using association rules with fuzzy logic to classify numerical data. Qodmanan et al. [26] and Yan et al. [27] used association rules in conjunction with genetic algorithms to optimize the value for minimum support and confidence.

Very few studies have applied the association rules for mining and matching frequent patterns to recognize activities. Gu et al. [28] and Palmes et al. [15] extracted the repetitive patterns from sensor activations which were frequent for one activity and infrequent for the others. They transformed the sensor activation data into different window sequences and computed the score for all individual activities. These scores were used as weights for categorizing activities from sensor activations. Rashidi et al. [29] proposed an activity recognition and tracking algorithm based on unsupervised learning. They mined the frequent patterns and clustered these patterns for all the available activities. However, they ignored the fact that the same sensors can be activated by different activities. Luhr et al. [30] used frequent pattern mining for activity recognition using sensor activation sequences, which at many times change in realistic environments. Yassine et al. [31] used association rule mining for analyzing the energy consumption patterns from electronic appliances which are directly related to human actions. Sfar et al. [32] proposed a causal association rule mining to detect and recognize anomalous activities in smart home environments. Atzmueller et al. [12] proposed a class association rule-based mining algorithm (CARMA) for analyzing different rule sets to select the final classification model. Their work used the inertial measurement units from mobile phones for recognizing human activities. Liu et al. [13, 14] used the association rules to mine the frequent patterns and used these patterns as discriminative features. Multi-task learning was employed to improve recognition performance. Liu et al. [33] focused on extracting mid-level features by mining frequent patterns using association rules. They suggested that the extracted temporal patterns can help improve activity recognition performance. Marimuthu et al. [34] proposed the activity recognition method using an adaptive neuro-fuzzy inference system with frequent pattern mining. The frequent pattern mining in this study was used to reduce the number of membership functions and the rules. Our method is inspired by the works of [4, 14, 33], i.e. proposing a modified Apriori algorithm to mine the frequent patterns. The main focus of the works mentioned above was to recognize overlapping activities efficiently. Most of the existing studies use the Opportunity dataset in which the activities were performed in a controlled environment and have uniform labels. The existing methods mine the frequent

patterns between the activity (ground truth) and the sensors involved while performing the activity (sensor activations). The proposed work is unique in a sense that it integrates both the sensor activations and low-level actions from IMUs to mine the patterns rather than limiting the method to a single modality. With the advances in IoT and micro-electromechanical systems (MEMs), the use of contextual sensors has also increased drastically. Therefore, the need of a framework which can undertake the information from more than one context is needed which is ignored by many existing studies. Moreover, none of the above-mentioned studies leverages human behavior for personalized human activity recognition. We extended their work for accommodating contextual information along with the generalization on the data acquired from sensor activations or inertial measurement units to mine the frequent patterns for activity recognition.

High-level activity recognition using contextual information

A very few studies focus on recognizing high-level activities as most of the existing works focused on classifying low-level actions. Recently a survey [35] highlighted how the characteristics such as time, location, and objects, are used as contexts for recognizing such high-level activities. Khowaja et al. [9] exploited ontological reasoning combined with data-driven methods to derive high-level activities. Gong et al. [36] introduced an idea of using pattern mining with the location as a context for high-level activity recognition. They used a single context, i.e. location, for recognizing high-level activities. Villalonga et al. [7] proposed a knowledge-driven approach for recognizing high-level activities using multiple contexts such as location and emotion. Cao et al. [37] proposed a group-based context-aware human action recognition (GCHAR) where they use the transitional logic and the previous state of the human to predict the action. The transitional logic was referred to as the context-aware component in their study. They also focused on improving the action recognition accuracy by leveraging the contextual information in the form of transitional logic, however, they only recognized four low-level actions, i.e. (grouping the similar actions in one) using IMUs. Other studies try to leverage the contextual information for constructing a better classifier model by embedding the contextual information in the feature space. Filippopolitis et al. [10] presented a location enhanced activity recognition where the location is considered as one of the features for training the classification model. Zhang et al. [38] used the subtasks as contextual information to recognize high-level activities. A sequence chunking strategy was proposed by segmenting and labeling different chunks of high-level activities. Lee et al. [39] conducted an exploratory study on the context affecting human activity recognition and in their study the context referred to as the housing environments. Aminikhanghahi and Cook [40] proposed the use of time context to segment a high-level activity into its corresponding subtasks to improve the recognition performance. Zhang et al. [41] proposed an ontological knowledge-based system for human activity recognition in smart homes. The main focus of the work was to deal with heterogeneous data sources and interoperability, therefore, multiple contexts have been used in collaboration with the sensor activations deployed at smart homes. Civitarese et al. [42] also proposed an ontological framework for human activity recognition for classifying interleaved and concurrent activities. The method was tested on the CASAS dataset which has homogeneous activity labels. Moreover, the personalized characteristics has somewhat been reduced

by removing the occurrences of noisy motion sensors as they may refer to a habit or a behavioral trait.

The above-mentioned studies focus on integrating the contextual information within the learning framework without considering the interrelationship of multiple contexts. The studies are also restricted to a single context and a single source of physical sensor readings (either sensor activations or inertial measurements) for recognizing high-level activities. Additionally, such studies do not consider personalization and activity handling problem due to the varying behavior which can hinder in generalizing the recognition performance. For instance, one subject performs “personal grooming” activity at the location “home” while the other subject performs it at the location “office.” Similarly, the activity “desk work” can be performed at both locations “home” and “office.” Using locations merely for constructing inference rules by assuming that certain activities are always performed at specific locations is another way of avoiding activity handling problems.

Personalized activity recognition

In realistic scenarios, the activities performed by one subject can be different from others in terms of activity label and behavior. For instance, while walking on the street, one subject may label the instance as socializing activity whereas the other subject may label it as a transportation activity. The activity handling problem arises due to diverse interactions with the contexts, which can be solved to an extent by introducing the aspect of personalization suggesting that the classification model should be separately trained for each subject. Many researchers agree upon having a personalized rather than the one-fits-all model [1].

Zhang et al. [43] proposed a probabilistic learning method for training activity models from both incomplete and complete data. Stikic et al. [44] tried to reduce the number of annotations using graph-based and multi-instance learning-based label propagation. Maekawa et al. [45] proposed a method where the model is trained on new users as well as other user’s data and termed it as supportive users. Bianchi et al. [46] proposed a personalized activity recognition method by collecting the data from individuals and training a light-weight model for each individual from a few days of data. The personalization effect was added irrespective of any physiognomies or behavior as the model is trained separately for each individual. Siirtola et al. [47] performed personalization with incremental learning approaches. The undertaken dataset had 9 subjects having the same action labels. As the data was collected in a controlled environment, i.e. activity handling problem was avoided, there was no improvement when applied the personalization effect. Burns and Whyne [48] proposed the use of personalized deep features, personalized engineered features, and personalized triplet networks to embed the personalization effect in their system. However, the study neither used any contextual representation nor accounted for human behavior. Other researchers focused on mapping the training data from the pool of existing users onto the new ones. The mapping is generally based on gender and demographic details such as age, height, weight, and gender [1]. The activity labels are the same for new and old users which is again an unrealistic scenario. For instance, one user can perform an activity labeled as taking medication or playing soccer whereas another user may not perform such activities at all. Thus the

activity labels may not be the same for all the users. In this case, the performance of personalized models also decreases as they are grouped based on their physicality rather than the similarity in their routines and the way they handle their activities.

Vaizman et al. [49], surveyed the contextual information which can be used for user behavior modeling to develop a specific healthcare application. They also suggested that for large-scale applications, the context-recognition component should be made unassertive without making the users alter their behavior. Shen et al. [50] used the behavioral characteristics of smartphone sensors to recognize human activities. They experimented on both the general and the personalized models using the smartphone usage behavior for recognizing low-level actions and suggested that the personalized models perform better than its counterpart. Mafrur et al. [11], modeled human behavior using life-logging data from smartphones to identify the human instead of activity recognition. This study supports our assumption that a general activity recognition would fail in real-world applications as human behavior is unique and different from other users. Jalali et al. [51], proposed the use of human behavior for recognizing multiple events. Furthermore, the sequential and parallel relations of these events were recognized using frequent co-occurrence patterns. Soleimani and Nazerfard [52] proposed the use of subject adaptor generative adversarial networks (SA-GAN) for cross-subject transfer learning. The method was tested on the Opportunity dataset which has homogeneous labels and the behavior of the subjects is the same as all the activities are performed in a controlled environment. It is apparent that human behavior analysis has not been used previously for personalized high-level activity recognition. To the best of our knowledge, this study is the first to use the behavior process models for computing the similarity between a new user and the existing pool of subjects.

The benefit of using associative learning is that it automatically maps the relationship of actions or sensor activations with the contextual information for each user, hence, reducing the problem of activity handling to some extent. Unlike the existing methods which use demographics or physiognomies of the users, we employ a process modeling approach to measure the similarity between different users to solve the activity handling problem. Furthermore, the studies only consider the datasets which exhibit a uniform set of labels for all the subjects which is not the case with real-world scenarios. For instance, "Taking medication" activity is not common with all the subjects, therefore, a limited number of users might perform this activity. In such cases, the system fails to recognize the activity due to the variations in activity labels. In this study, we propose the similarity metric based on the behavior process models to map the subject in such a way that the variation in activity labels can be reduced and the activity recognition performance can be enhanced.

Some of the existing works are consolidated and summarized in Table 1 with respect to the personalization, activity handling, data source, classification approach, low-level actions, high-level activities, and use of contextual information. As it can be noticed that none of the existing works focuses on the activity handling which represents a real-world problem as each individual performs an action in their own way. Similarly, there are very few works proposing a generalized framework that can be applied to both the IMUs and sensor activations. With respect to the consolidated review of existing works, we are focusing on the maximum number of activities, i.e. combined low-level actions

Table 1 Summary and comparison of related works

Study	Pers	AH	Data	CA	LLA	HLA	CI
[4]	✗	✗	SA	CAR	✗	8	✗
[53]	✗	✗	SA	CAR	✗	4	✗
[28]	✗	✗	SA	CAR	✗	4	✗
[29]	✗	✗	SA	Clustering, ARM	✗	8	Time
[30]	✗	✗	SA	ARM	✗	13	Time
[43]	✓	✗	IMU	EM, PL	✗	9	Time
[54]	✗	✗	IMU, SA	MLR, SI	4	6	Location
[16]	✓	✗	IMU, SA	OSVM, PCA	5	✗	✗
[45]	✓	✗	IMU	HMM	2	12	✗
[1]	✓	✗	IMU	HBN, SVM	7	✗	✗
[50]	✓	✗	IMU	RF	5	✗	✗
[10]	✗	✗	IMU	KNN, LR, RF, SVM	✗	8	Location
[36]	✗	✗	IMU	XGBoost	8	✗	Location
[9]	✗	✗	IMU	HCA, ELM	6	10	Location
[38]	✗	✗	IMU	CNN	✗	18	Time
[40]	✗	✗	SA	RF	✗	10	Time
Proposed	✓	✓	IMU, SA	CAR, AL, ECOC, RNN	7	8–13	Multiple contexts

Pers, Personalization; AH, Activity Handling; CA, Classification Approach; LLA, Low-Level Actions; HLA, High-Level Activities; CI, Context Information; SA, Sensor activations; IMU, Inertial measurement unit; CAR, Class Association Rules; ARM, Association rule mining; SVM, Support Vector Machines; PCA, Principal Component Analysis; EM, Expectation Minimization; PL, Probabilistic Learning; MLR, Multiple Logistic Regression; SI, Statistical Inferencing; OSVM, Online SVM; HMM, Hidden Markov Models; CNN, Convolutional Neural Networks; HBN, Hybrid Bayesian Networks; KNN, K-Nearest Neighbor; LR, Logistic Regression; RF, Random Forest; XGBoost, Extreme Gradient Boosting; HCA, Hierarchical Classification Approach; ELM, Extreme Learning Machines; CAR, Class association rules; AL, Associative Learning; ECOC, Error-Correcting Output Codes; RNN, Recurrent Neural Networks

and high-level activities, to evaluate the performance of the CAPHAR framework. Class association rules have been used previously for 4–8 activities using only sensor activations without the consideration of contextual information. This work not only extends the number of activities to recognize but also adds multiple contexts along with the generalization of both sensor modalities, i.e. IMUs and sensor activities. Moreover, the existing studies use demographics and physiognomies for personalization whereas in this study we use the human behavior process to model the personalization effect which helps to reduce the activity handling problem.

Proposed method

Figure 2 illustrates the CAPHAR framework, specifying the inter-dependencies among the building blocks. The method for high-level activity recognition is based on two modules: low-level action recognition (in the case of data from IMUs) and high-level activity recognition. We consider the data from IMU, i.e., accelerometer (Acc) and Gyroscope (Gyr) sensors for low-level action recognition (LLAR) module. The method extracts the features from raw sensor data and applies machine learning methods to build a classification model. The output of this module (classified action) will be the input to the high-level activity recognition (HLAR) module. The pipeline of the LLAR module is quite similar to conventional learning, therefore, we adopt the feature extraction and classification method proposed in existing works. As suggested in Sect. “[Association rule mining and Class association rules for activity recognition](#)”, the associative learning method

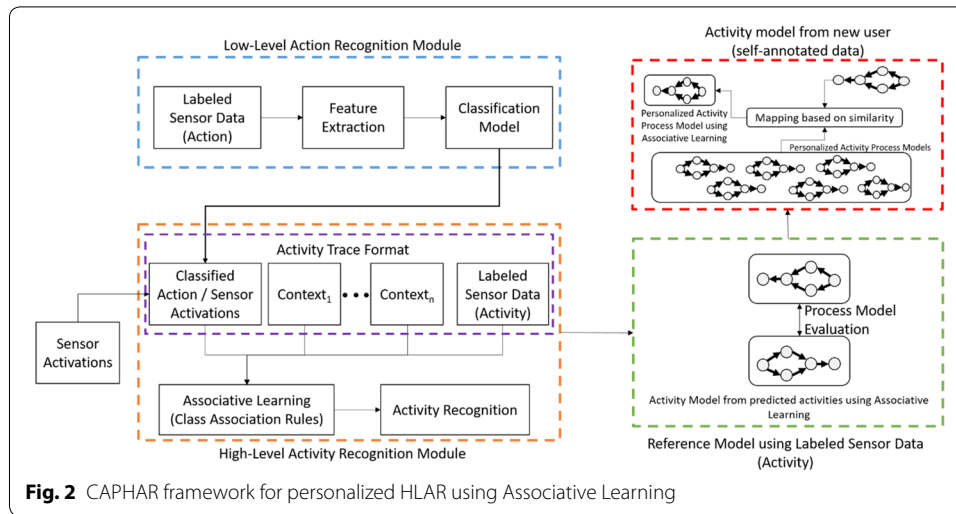


Table 2 Nomenclature of the variables used in this study

Variable	Description	Variable	Description
N	Number of samples for training	AS	Set of available activations
pr_i	Probability vector of i -th instance	I	Itemset
q_i	Correct label for i -th instance	T	Set of Time encoded values
ℓ	Number of classes	L	Set of locations
A	Set of low-level actions	HA	Set of high-level activities
S	Set of sensor activations	tr_k	k -th activity trace
C	Set of available contexts	$Supp$	Support
$Conf$	Confidence	car	Class association rule
z	Unlabeled itemset	E	Set of events
$PT\Upsilon$	Event trace	\mathcal{L}	Event Log
P_l	Set of places	$\mathcal{T}\tau$	Set of transitions
\mathcal{F}	Set of directed arcs	TF	Trace fitness
Mod	Process model	$\hat{\mathcal{L}}$	Arbitrary event log
Ge	Generalization	Sim	Similarity

in HLAR is inspired by the works proposed in [4, 14, 33], however, we have modified the algorithm and method to work with multiple contexts as well as to integrate the use of IMUs and sensor activations. Furthermore, the use of behavior models for computing the similarity among the pool of subjects for the semi-population calibration approach is the main highlight of the CAPHAR framework. Table 2 presents the nomenclature of the variables used in this study. Each of the variables is briefly explained in the text at the point of their usage.

The HLAR module takes into account either the classified action or sensor activations, along with the available contextual information to perform activity recognition using the associative learning method. To perform the personalization, we first generate the behavior process models from the predicted activities using associative learning. These models are evaluated against the reference activity models generated from the annotated data. For a new test user, the behavior process model is mapped on to the existing pool

of models based on the similarity metric. The most similar model will be elected to recognize personalized high-level activities from the new test user's data. In the below sub-sections, we explain the particulars of each building block for LLAR and HLAR modules with the details for process mining and similarity approaches.

Low-level action recognition module

The LLAR module follows a typical machine learning pipeline to recognize low-level actions. We present the details of the employed dataset in a later section. The data from inertial sensors are usually acquired with a sampling rate of 50 Hz. The data is first pre-processed to reduce the noise in the data stream. We employ the low-pass butter-worth and median filter with a 10 Hz cutoff frequency to filter the raw data. The signals are then sampled with sliding windows of 1.5 s and overlapping of 30% between the sliding windows. Once the data is pre-processed, we extract features from each sliding window.

Feature extraction

We extract 34 statistical features from Acc and Gyr sensor readings from each sliding window, i.e. 17 features from each of the sensors. The fast Fourier transform (FFT) coefficients are based on the frequency components from 1 to 50 Hz, which are extracted from each of the sensors and are added to the feature space, accordingly. Thus, the statistical features (34) and the FFT features (100) results in a total of 134 features. We present the list of extracted features in Table 3. These features are commonly used for low-level action recognition studies [55].

Classification model

The classification in our approach is to recognize the discrete category of new sensor readings by learning from the training samples. Many generic classification methods have been proposed and used for LLAR such as SVM, neural networks (NN), decision trees, and gradient boosting. Recently, deep learning algorithms have left a positive impact in the field of action recognition, especially the convolutional neural network (CNN) [3] and recurrent neural network (RNN) [56]. These methods have shown promising results, but they require a large dataset for training which sometimes may not be available in real-life scenarios. As the LLAR module is of vital importance to the pipeline of HLAR, we will evaluate different kinds of classification methods for attaining high accuracy. In this regard, we mainly adopt two methods for classification which have shown promising results in the literature, i.e. error-correcting output codes (ECOC) [6] and long short-term memory networks (LSTM) [56]. We use ECOC with only a single base classifier without any mini-batch to perform the training. The base classifier is chosen empirically as shown in Sect. “[Low-level action recognition results](#)”.

Table 3 List of extracted features

Statistical features	Mean, Standard Deviation, Energy, Mean-Crossing Rate, Maximum Value, Minimum value, First Quartile, Second Quartile, Third Quartile, Amplitude of the power spectral density (PSD) 4 features, shape of PSD 4 features
Frequency domain features	Fast Fourier Transform (FFT) coefficients.

We also evaluate the LLAR module using LSTM networks, a variant of RNN. The difference with existing studies is that we use the loss-function based on the F1-score instead of the cross-entropy. The reason for employing F1-score loss is that most of the publicly available datasets are highly imbalanced in terms of samples for a particular action label. Thus, using cross-entropy, L1 or L2 loss may result in overfitting for specific action classes with a large number of observations. The F1-loss for the set of N samples is defined in Eq. 1.

$$F1_{loss} = \mathbb{E}_{l \in [1, \ell]} \left\{ 1 - \frac{2 \sum_{i=1}^N pr_i^\circ q_i}{\sum_{i=1}^N pr_i + \sum_{i=1}^N q_i} \right\} \quad (1)$$

where pr_i refers to the probability vector for each instance, the symbol $^\circ$ denotes element-wise multiplication, ℓ represents the number of action classes, and q_i is the binary vector signifying the correct action class l , i.e. $q_{il} = 1 \wedge (q_{ik} = 0)_{k \neq l} | q_i \in \mathbb{R}^\ell$. For example, let us consider we get the probability vector for an instance for action classes: sitting, standing, walking, and running as $[0.3 \ 0.3 \ 0.25 \ 0.15]$, respectively. If the true label for the given instance is standing, the binary vector q_i will be represented as $[0 \ 1 \ 0 \ 0]$ for the respective action classes.

High-level activity recognition module

Our HLAR module considers either the low-level action or sensor activations along with the contextual information. HLAR faces one dependency which is the availability of information from either of the modality for carrying out the further process. As shown in Fig. 2, if the data from IMUs is available, we use the classified action; otherwise, the sensor activations will be considered to recognize high-level activities. In this section, we define the concept of building an activity classifier using CARs.

To move further, we first define the *activity trace* which is the set of low-level actions or sensor activations and the contextual information during an activity. For instance, the high-level activity “socializing” may comprise multiple low-level actions such as walking, climbing (down), and standing, at location “office” on a specific time which can be regarded as the activity trace of “socializing.” An example of an activity trace is shown in Fig. 3. The first example shows the transformation of the raw data from IMUs to the activity trace format in compliance to Fig. 2 (HLAR Module). The second example considers the sensor activations as an input. The datasets which we chose for evaluation of associative learning provide time and location as their contexts. Therefore, in Fig. 3, *context1* and *context2* refer to time code and location, respectively. The goal of associative learning in our study is to construct a classifier based on CARs from the traces of each activity.

Definition 1 (Activation) Let AS be a set of available activations. Then, AS is defined in Eq. (2).

$$AS = [A \vee S \mid A \oplus S = 1] \quad (2)$$

where A is a set of low-level actions defined as $A = \{a_1, a_2, \dots, a_g\}$, and S refers to a set of sensor activations defined as $S = \{s_1, s_2, \dots, s_h\}$. The indexes g and h represent

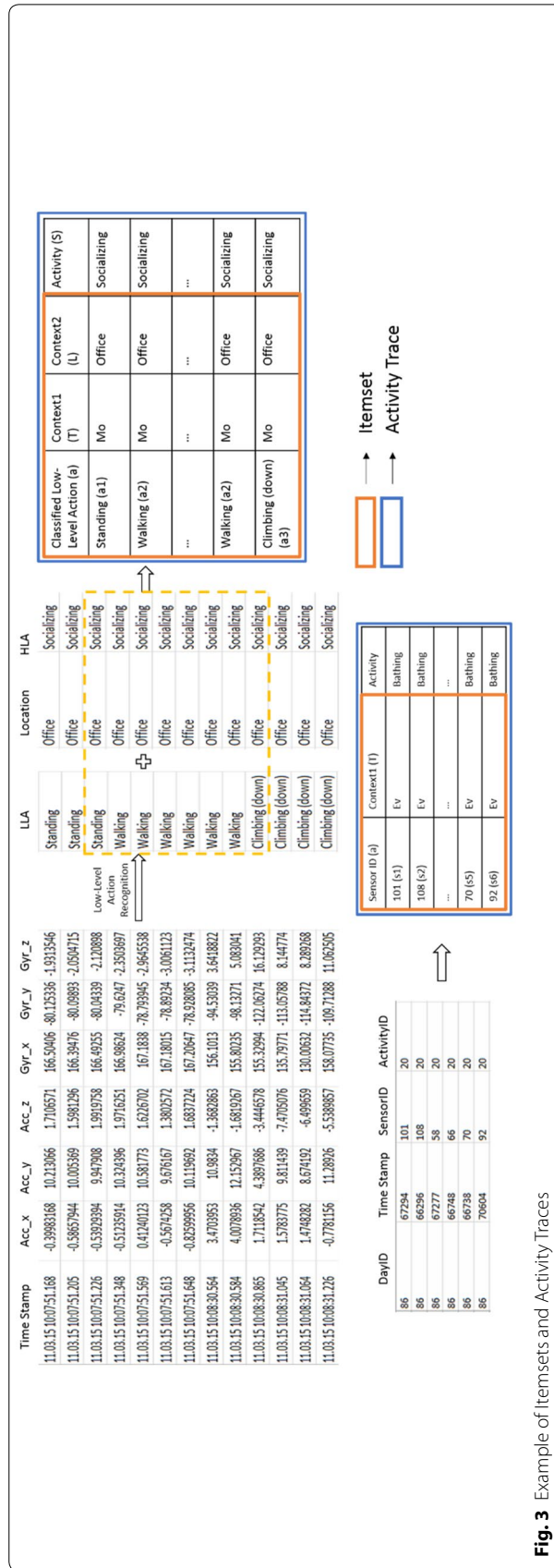


Fig. 3 Example of Itemsets and Activity Traces

the number of low-level actions and the number of sensors available for activations, respectively. An operator \oplus refers to an exclusive-OR (XOR) operation suggesting that the **activation** will simply return either the set of low-level actions or sensor activations, based on the availability of the data. We define the set of available contexts as $C = \{c_m | m = 1, \dots, M\}$, where m is the number of available contexts. We constitute the following definition for the itemset which combines the available activations with different contexts.

Definition 2 (Itemset) Let I be a set of itemsets. Then, an itemset $i_k \in I$ is defined in Eq. (3)

$$i_k = \{AS_k, C_k\} \quad (3)$$

where $AS_k \subseteq AS$ and C_k are the k -th action and context sets, respectively. For example, AS may comprise of four low-level actions i.e., “standing, walking, sitting, and lying”, whereas AS_k may only include “sitting and standing” amongst four actions. As per the example shown in Fig. 3, we use two contexts i.e., *Time* (T) and *Location* (L) which are defined as $L = \{l_1, l_2, \dots, l_b\}$, and $T = \{Mo, Af, Ev, Ni\}$ where L represents a set of locations and T is a set of time-coded values. In our study, we mapped the timestamps to four encoded values as follows: *Mo*: 05.00 a.m.–12.00 p.m., *Af*: 12.01 p.m.–04.30 p.m., *Ev*: 04.31 p.m.–07.30 p.m., and *Ni*: 07.31 p.m.–12.00 a.m. In the case of activities performed in overlapping time codes or different locations, we consider the one with a maximum number of instances. For example, if the user has performed an activity from 5.30 to 7.45, we compute the number of instances for *Ev* and *Ni* and choose the one having maximum instances i.e., *Ev*. In this regard, we can replace C_k by T_k and L_k in definition 2. We define the set of high-level activities (HLA) in Eq. (4)

$$HA = \{ha_r | r = 1, \dots, R\} \quad (4)$$

where R refers to the number of HLAs. The activity trace in the form of a transaction is defined in the following definition.

Definition 3 (Activity trace) Let Tr be a set of activity traces. Then, an activity trace $tr_k \in Tr$ is defined in Eq. (5)

$$tr_k = \{ha_k, i_k\} \quad (5)$$

where is the k -th trace, ha_k and i_k are activity and an itemset assigned to the k -th trace tr_k , respectively. An example $tr_1 = \{PersonalGrooming, [\{sitting, standing\}, Home, Mo]\}$ represents that $ha_1 = PersonalGrooming$ and $i_1 = [\{sitting, standing\}, Home, Mo]$. Then, we can define the support and confidence of Tr given antecedent I^* in Eq. 6, 7 where $I^* \subseteq I$

$$Supp(I^* \Rightarrow ha_r) = \frac{\sum_k f_{I^*, ha_r}(tr_k)}{|Tr|}, Conf(I^* \Rightarrow ha_r) = \frac{\sum_k f_{I^*, ha_r}(tr_k)}{\sum_k f_{I^*}(tr_k)} \quad (6)$$

where

$$f_{I^*, ha_r}(tr_k) = \begin{cases} 1, & I^* \subset tr_k \wedge ha_r \in tr_k \\ 0, & \text{otherwise} \end{cases} \text{ and } f_{I^*}(tr_k) = \begin{cases} 1, & I^* \subset tr_k \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

In the above equation, $|Tr|$ is the total number of activity traces in the training set, $\sum_k f_{I^*}(tr_k)$ refer to the number of activity traces having I^* and $\sum_k f_{I^*, ha_r}(tr_k)$ is the number of activity traces containing I^* and ha_r . We provide the examples of activity traces in Table 4 using low-level actions with the contextual information. From Table 4 we can compute $Supp$ and $Conf$ for $([\{sitting\}, Af, Office] \Rightarrow PersonalGrooming)$. There is only one instance with itemsets $([\{sitting\}, Af, Office]$ and $\{Personal\ Grooming\}$, resulting in $\sum_k f_{I^*, ha_r}(tr_k) = 1$. In Table 4 the itemset $[\{sitting\}, Af, Office]$ appears twice, thus $\sum_k f_{I^*}(tr_k) = 2$. We can get easily: $Supp$ and $Conf$ for $([\{sitting\}, Af, Office] \Rightarrow Personal\ Grooming)$ which are $1/4$ and $1/2$, respectively.

The activity trace may include the itemsets from I which are unique for different HLAs, but the real-life scenarios are much complicated which gives rise to activity handling problems as different people perform the activities differently [57, 58]. For example, in the case of sensor activations, activity ha_1 triggers the sensors $\{s_2, s_4, s_4, s_4\}$ and activity ha_5 triggers the sensors $\{s_2, s_2, s_2, s_4\}$ with the same location and time code. Both the HLAs trigger the same set of sensors $\{s_2, s_4\}$, therefore they are quite challenging to differentiate. However, both the activities have different frequencies of sensor activations, and this information can be used to improve the recognition process. In this regard, we use the frequencies of actions or sensors triggered during a certain activity. This will allow us to find the temporary patterns for the sensor activations or actions which might improve the performance of associative learning, intuitively. By doing so, we are incorporating the frequency of actions or sensors into the confidence calculation for an association rule. We modify the set of actions and sensors as, $A = \{(a_1, |a_1|), (a_2, |a_2|), \dots, (a_g, |a_g|)\}$ and $S = \{(s_1, |s_1|), (s_2, |s_2|), \dots, (s_h, |s_h|)\}$, respectively. The notation $|\{\cdot\}|$ refers to the frequency of corresponding actions or sensor activations. The same earlier example will result in $\{(s_2, 3), (s_4, 1)\}$ and $\{(s_2, 1), (s_4, 3)\}$ for activities ha_1 and ha_5 , respectively. The embedded frequency information provides better discriminative capability compared to the previous scenario where we only considered

Table 4 Example of activity traces

Activity trace No.	Itemset			Activity
	Low-level actions (AS_k)	Context1 (T)	Context2 (L)	
tr ₁	Sitting, standing	Mo	Home	Personal grooming
tr ₂	Sitting	Af	Office	Personal grooming
tr ₃	Sitting	Af	Office	Desk work
tr ₄	Standing, walking	Ni	Streets	Socializing

the sensor activations. For further analysis and computation, we use the modified set of actions and sensor activations which embeds the frequency information.

Another advantage of using the frequency information is that it helps to overcome the intra-activity variations in the activity traces. Assuming the two activities, i.e. “grooming” and “toileting” as shown in Fig. 4. The “toileting” activity triggers the sensor placed at location “bathroom,” while “grooming” activity is performed at different locations including “bathroom.” The “toileting” activity triggers the sensor at the location “bathroom” more often than the activity “grooming.” This suggests that ignoring frequency information may misclassify these activities due to the intra-activity variation. We extend the method of mining CARs from [4].

Classification of activities using CARs

As shown in the previous section, each HLA has its own activity traces. In this regard, we create CARs in which the activity trace itself is the antecedent and the HLA is the consequent as suggested in [4, 59]. The classifier using associative learning can be built by using all class association rules which meet user-specified minimum *Supp* and *Conf*. There are mainly two steps for building such a classifier [4]: (1) eliminating the rules which are subordinated to other rules, and (2) removing the rules from the first step that do not contribute to the improvement of classification accuracy. For instance, given two rules, if rule 2 is subordinate to rule 1 and both the rules have the same consequent, then rule 2 will be considered redundant with rule 1. Consequently, the frequent pattern (antecedent) of rule 2 is considered as the subset to that of the rule 1. Some of the rules proposed in [4] are also applied for the elimination which is as follows:

- Rule1 has higher *Confidence* than rule2
- The *Confidence* of both rules is the same, but rule 1 has higher *Support*.
- The *Confidence* and *Support* for both rules are the same, but rule 1 was generated before rule 2.

In such cases, the rule 2 will be eliminated. While classifying a new instance, the rule whose activity trace satisfies the instance will be considered for the classification. We create a group of CARs and represent it as $car(ha_r)$. For the classification, we apply the group of CARs to aggregate the *gain* against the HLA for classifying a test instance. Using multiple association rules has proven to achieve better classification performance as compared to the single association rule [24]. In order to create the list of CARs, we define $car(ha_r)$ that stores activity traces associated with different activities in Eq. (8)



Fig. 4 Example of intra-activity variation

$$car(ha_r) = \{z | z \in Z, Conf(z \Rightarrow ha_r) \geq Conf_{min}, Supp(z \Rightarrow ha_r) \geq Supp_{min}\} \quad (8)$$

Algorithm 1. Classification of test instance using associative learning
Input: <i>unlabeled itemset z and car for each HLA</i>
Output: <i>class label of z</i>
1: For each $ha_r \in HA$ do
2: $var = \phi, gain(z, ha_r) = 0$
3: For each z which is available in $car(ha_r)$ do
4: If $z \subseteq Z$ and $!set(z).issubset(set(var))$ then
5: $gain(z, ha_r) += \frac{conf(z \Rightarrow ha_r)}{conf(z \Rightarrow ha_r) + 1}$
6: $var.append(z)$
7: End If
8: End For
9: End For
10: Return ha_r with maximum $gain(z, ha_r)$

where z is the unlabeled itemset. The notations $Conf_{min}$ and $Supp_{min}$ refer to the user-specified minimum confidence and support, respectively. We arrange the itemsets according to their length in descending order. For example, if rule1 is the superset of rule 2, then rule 2 would not be considered for the computation of $gain(z, ha_r)$ which is the exact purpose of line 4 in Algorithm 1. The $gain(z, ha_r)$ is aggregated based on the CARs (line 5, Algorithm 1), and the HLA having the maximum gain will be returned as the classification result (line 10, Algorithm 1). Algorithm 1 only focuses on the association rules for a certain high-level activity, therefore, the complexity of the proposed algorithm is $O(n)$. Furthermore, the association rules that do not meet the requirement of minimum threshold in terms of confidence are pruned which results in a smaller search space.

Human behavior process modeling

In past years, human behavior process modeling from activities of daily life (ADL) has been a hot research topic. Various studies try to mine the patterns based on the annotated activity logs. Some of them focus on anomalous behavior detection whereas others emphasize predicting activities [60]. There are various methods and techniques to perform behavior process modeling. We provide notations and a brief description of the elements considered for constructing a behavior process model. Generally, a process model consists of events, traces, and logs. It is to be noted that the trace considered in the process model is quite different from the activity trace. Let E , PTr , and \mathcal{L} represent the set of events, traces of events, and logs, respectively. The set E comprises of a finite set of events, and the trace PTr is the sequence of events in E such that $PTr = \langle e_1, e_2, \dots, e_N \rangle$ where N is the size of the trace $|PTr|$. The \mathcal{L} comprises of traces such that $\mathcal{L} = \{PTr_1, PTr_2, \dots, PTr_M\}$. In our study, an event is an activity (HLA) and the PTr is the sequence of the HLAs performed. An example of \mathcal{L} can be given as $\{\langle Housework, Eating \setminus Drinking, PersonalGrooming^2 \rangle, \langle PersonalGrooming, DeskWork, Socializing, Transportation \rangle\}$. In this example each HLA is an event and each sequence of events enclosed by $\langle . \rangle$ is a trace of events. The given example consists of three traces: two instances of $\langle Housework, Eating \setminus Drinking, PersonalGrooming \rangle$

whose frequency is denoted by the number in the superscript, and one instance of $\langle \text{PersonalGrooming}, \text{DeskWork}, \text{Socializing}, \text{Transportation} \rangle$. Using the events, traces of events and logs, the process model can be created using Petri net. For more details of process mining building blocks and creation of process model using Petri net, refer to [61]

The reasons for using such process modeling methods are twofold: the first is to use the activity process models in conjunction with associative learning to improve recognition, efficiency, and reliability, and the second is to use the activity process models as a means for computing the similarity between two users. The similarity-based on behavior process models will not only help to map the new user on to the pool of existing models but also to reduce the effect of activity handling issues. We used the inductive visual miner (IVM) [62] to construct the process model from the activities classified using the proposed associative learning. The reason for choosing IVM is the guarantee of soundness in comparison to other methods. The soundness refers to the absence of deadlock in the process model [62]. We evaluated the effectiveness of this implication by generating the process model from classified activities using a part of individual data and compared it with the model generated from the annotated data. The process model is first discovered with Petri nets which represent the flow of a process using modeling formalisms. A triplet can represent a Petri net (P_t, T_t, \mathcal{F}) , where P_t and T_t refer to the set of places and transitions, respectively, such that their interaction results in \emptyset . The term \mathcal{F} is regarded as a set of directed arcs and can be defined as $\mathcal{F} = (P_t \times T_t) \cup (T_t \times P_t)$. Once the Petri net model is created, we then apply the IVM to generate the final process model which is represented as a process tree. For the Petri net, we assume the standard semantics as proposed in [61].

The evaluation was based on two parameters, i.e., trace fitness and generalization [63]. Trace fitness measures the extent of the model for reproducing the traces from the event log. It tries to align as many events as possible from the traces (also called the alignment measure). If the alignment is not perfect, the events may be skipped or inserted without their presence in the log, but this adds a penalty to the fitness score. The computation of trace fitness score is presented in Eq. (9)

$$TF(Mod, \mathcal{L}) = 1 - \frac{cost(Mod, \mathcal{L})}{\min(cost(Mod, \hat{\mathcal{L}}))} \quad (9)$$

where Mod refers to the process model created using Petri net followed by IVM. The variable $\hat{\mathcal{L}}$ is an arbitrary event log such that $\hat{\mathcal{L}} \subseteq \mathcal{L}$. Trace fitness heavily depends on the ratio for the cost of aligning \mathcal{L} with the Mod to the minimal cost of aligning the model on the $\hat{\mathcal{L}}$. The minimal cost is also responsible for normalizing the value between 0 and 1. The generalization refers to the frequency of each event in the model visited to reproduce the event log. It indicates incorrect behavior if an event is visited more often than it is anticipated. The generalization is considered to be bad if some events of the model are visited very infrequently. The computation of the generalization is given in Eq. (10)

$$Ge(HA) = 1 - \frac{\sum_{|HA|} (\sqrt{|HA|})^{-1}}{HA} \quad (10)$$

We represent the number of executed activities in the event log by HA' . The value of Ge is in the range of 0 to 1, where 0 indicates very bad generalization and 1 indicates the optimal generalization.

User similarity based on behavior process models for personalization

The studies carried out for personalized activity recognition have always thrived for solving the cold start problem. It is common in all activity recognition systems that the performance on the data acquired from a new user is not generalized to the same extent due to the small amount of labeled data. To cope with this problem, researchers use the *calibration* approach which is to map the characteristics of the new user data on to the pool of existing subjects. By doing so, the increase in recognition performance has been achieved. Existing studies map the characteristics of a new user based on gender, age, and physiognomies such as height, weight, and age. These physiognomies fail to capture the variations introduced by the difference in activity handling. In this regard, we measured the similarity of a new subject's annotated activity log with that of the existing behavior process models to find the most relevant mapping for personalized activity recognition. The similarity is computed based on the formulation shown in Eq. (11)

$$Sim(U_1, U_2) = \alpha * TF(Mod_{U1}, \mathcal{L}_{U2}) + \beta * Ge(HA) + \gamma * normalized\ number\ of\ activities(U_1, U_2) \quad (11)$$

where α , β , and γ are the user-defined weights for trace fitness, generalization, and number of normalized activities, such that $\alpha + \beta + \gamma = 1$, respectively. For computing similarity between two subjects i.e. U_1 and U_2 , we compute the trace fitness for the model generated by activities of U_1 and the log of U_2 . We normalized the number of activities for two subjects with respect to the maximum number of activities available either in the log or the model. To compute the normalized number of activities, we used the computation as: $\frac{avg(\#ofact_{U1}, \#ofact_{U2})}{\max(\#ofact_{U1}, \#ofact_{U2})}$. Once we find the process model having the maximum similarity from the existing pool of subjects, we can apply the association rules to perform activity recognition.

Experiments and results

In this section, we present the details of the dataset employed in our study and the challenges associated with it. The experimental results obtained for LLAR and HLAR are discussed in detail. Additionally, we also present the results for the effectiveness of the behavior process model generation and personalization based on the proposed similarity metric as discussed in the former section. For the experiments and analysis, we use two datasets. The first is the daily lifelog dataset which we refer to as dataset1 and the second is activity recognition in the home dataset which we refer to as dataset2 in our study. One of the most widely used activity recognition dataset is Opportunity [64], however, we do not consider this dataset due to the following reasons:

- The recognition accuracy has already surpassed over 98% [13].

- It deals with only 5 high-level activities.
- The dataset is acquired in a controlled environment, therefore, including personalization will adhere little to no effect.
- This study is focused on the activity handling problem which cannot be addressed through this dataset due to the limited variations in performing activities.

The employed dataset1 has been made publicly available by the University of Mannheim [17]. This dataset consists of data from 7 subjects (age 23.1 ± 1.81). Each subject data has been recorded for 12 h/day on average for 2 weeks. There are seven low-level actions, i.e., climbing (up), climbing (down), sitting, standing, walking, lying, and running. Additionally, this dataset offers 13 high-level activities namely deskwork, eating/drinking, housework, meal preparation, movement, personal grooming, relaxing, shopping, socializing, sport, transportation, sleeping, and taking medication. The data was self-recorded using their current location, low-level action, and high-level activity. The subjects were only given an initial guideline and information on pre-defined labels, but they were not administered when performing activities. Due to the lack of supervision, subjects performed the activities differently as compared to the others which introduce the activity handling problem. Existing studies that use physiognomies might fail to perform well on this dataset as all the subjects are male with a similar age range and physical appearances.

The dataset2 has been made publicly available by the Massachusetts Institute of Technology [23]. This dataset consists of data from 2 subjects (age 30 and 80). Each subject data has been recorded for 16 days. This dataset comprises of sensor activations which were fitted to different appliances, furniture, and containers used in everyday life activities. The first subject performs 13 activities, i.e., cleaning, doing laundry, preparing a beverage, grooming, dressing, going out to work, bathing, preparing a snack, washing dishes, preparing breakfast, toileting, preparing lunch, and preparing dinner. On the other hand, the second subject performs 9 activities which include watching TV, preparing a snack, washing dishes, preparing breakfast, toileting, taking medication, listening to music, preparing lunch, and preparing dinner. As this dataset is a lot in contrast with the daily lifelog, it will prove the applicability of our proposed method to generalize on both datasets having different characteristics.

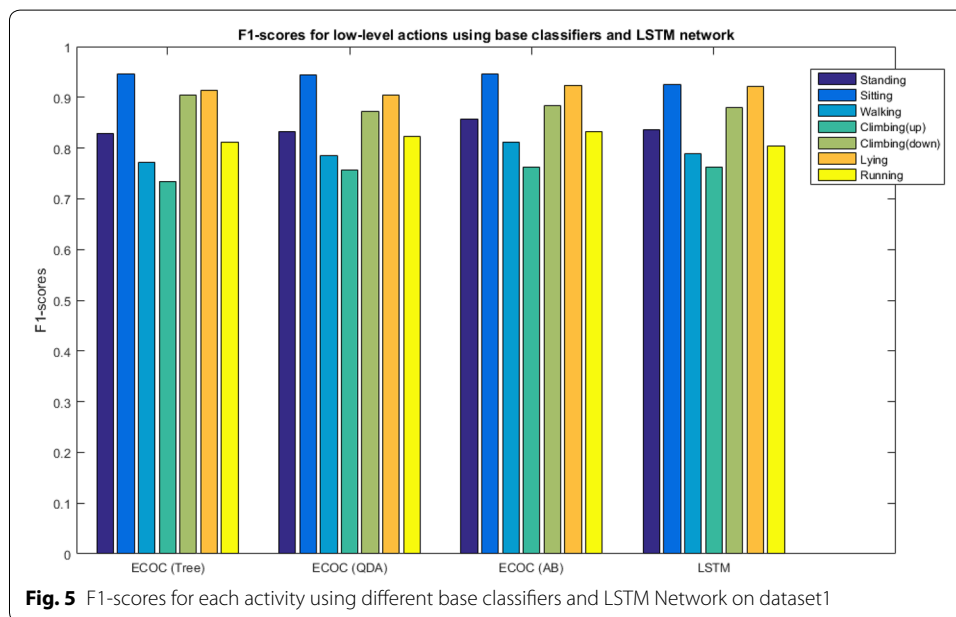
Low-level action recognition results

We perform the analysis of LLAR on dataset1 only as the dataset2 does not provide the information for low-level actions. As explained in the former section, we adopted an existing method [6] with a single base classifier for recognizing low-level actions along with the LSTM network for carrying out a fair comparison. We used the following model parameters for LSTM. The network model comprises of 2 layers with 256 units in each layer. The drop-out value was set to 0.5 for both hidden layers. We used an ADAM optimizer [65], with a learning rate of 0.001 decaying to 0.005. As the number of instances for the low-level actions is highly imbalanced, we compute not only the accuracy but also the F1-scores which is commonly used for evaluating action recognition methods. We have performed a leave-one-subject-out (LOSO) analysis since the low-level actions do not exhibit the activity handling problem in

Table 5 Low-level action results using base classifiers and LSTM network

Action	ECOC (Tree)	ECOC (QDA)	ECOC (AB)	LSTM
Standing	91.06%	91.18%	93.60%	92.67%
Sitting	94.45%	93.56%	95.97%	93.84%
Walking	94.49%	95.02%	93.50%	89.45%
Climbing (up)	82.93%	83.74%	87.12%	86.97%
Climbing (down)	91.80%	90.99%	92.45%	92.07%
Lying	96.65%	94.38%	97.67%	97.34%
Running	84.57%	86.84%	88.61%	84.16%
Average	90.85%	90.81%	92.70%	90.93%

ECOC, Error Correcting Output Codes; QDA, Quadratic Discriminant Analysis; AB, Adaptive Boosting; LSTM, Long Short-Term Memory



this dataset. Table 5 presents the accuracies for each action using different base classifiers and the LSTM network. Figure 5 shows the F1-scores for each action obtained using the respective methods. It is quite apparent that our existing method using only an adaptive boosting (AdaBoost) classifier performs better than LSTM networks. The graph shows that the variations in accuracy are uniform, for instance, the actions “lying” and “sitting” perform comparatively better in comparison to other actions which is obvious due to their non-confusing characteristics. We observe that “standing” action is confused with “walking” action sometimes. Similarly, “running” action is also confused with “walking” action. As the self-annotation was not administered, we assume that the users stop moving or stand for the time being while not annotating it as “standing”. In action recognition, we call it transitional actions such as stand-walk-stand, and stand-sit-stand. Another reason for the low-accuracy of “Climbing (up)” and “running” action is that the number of observations for these actions is less as compared to the other actions. It is obvious that in real life people do not run or

Table 6 Comparative analysis for High-level activity recognition

Subjects	RF	HMM	ECOC (AB)	LSTM	AL	# of Act
Subject 1	50.00%	51.02%	56.25%	62.96%	64.91%	12
Subject 2	48.96%	56.14%	59.60%	64.87%	69.39%	13
Subject 3	51.25%	49.78%	55.27%	59.95%	67.71%	13
Subject 4	53.56%	51.87%	61.84%	67.14%	73.91%	12
Subject 5	39.78%	34.25%	45.52%	46.20%	54.52%	8
Subject 6	52.37%	58.96%	67.30%	66.52%	72.00%	11
Subject 7	51.90%	59.00%	66.67%	65.36%	72.54%	10

RF, Random Forest; HMM, Hidden Markov Model; ECOC, Error Correcting Output Codes; AB, AdaBoost; LSTM, Long-Short Term Memory Networks; AL, Associative Learning; # of Act, Number of Activities

climb stairs more often in their daily lives which makes this dataset more challenging in terms of low-level action recognition as well.

Although the accuracy of all the methods is relatively closer, it is necessary to select the classifier with the best recognition performance as our HLAR is dependent on the accurate classification of low-level actions. It is also necessary for a good classifier to recognize not a certain subset but all the actions accurately. In comparison to the ECOC (AB), LSTM's performance for the walking and running action is not so good, even the performance for the running action is lower than the other two classifiers. We consider the average accuracy of 92.70% is promising enough given the above observations drawn from the results. Moreover, the testing time for ECOC using AdaBoost and LSTM networks is approximately 0.2 s and 0.48 s, respectively which shows that the ECOC with AdaBoost is better in terms of execution time as well as performance for LLAR.

High-level activity recognition results

To evaluate the personalized HLAR results using our proposed method, we use a leave-one day out (LODO) validation. We use a single day of each subject for testing while the remaining days for the computation of CARs. We repeat this experiment the times same as the number of days for each subject and report the average results. For computing Car (S_r), we set the value of $Conf_{min}$ and to 60% and 20%, respectively. In order to make a fair comparison with existing methods, we tested many discriminative classifiers such as decision trees, support vector machines, and Naïve Bayes. We also compared our method with hidden Markov models (HMM) [66] as they take into account the temporal relationships between the activities and have been used extensively in activity recognition. We combined the features extracted from Acc and Gyr sensor measurements along with a low-level action label, location, and time code to train the discriminative and generative classifiers for dataset1. Similarly, we used the sensor activations along with the time code to train the classifiers for dataset2.

Although the location information is available in dataset2 it is redundant to the sensor activations as the location indicates the placement of sensors not the presence of the user. We report the accuracies in Table 6 using HMM, random forest (RF), ECOC using AdaBoost as a base classifier, and LSTM network with the same configuration used for LLAR except that we added one more hidden layer with drop out ratio of 0.2. The reason for reporting the accuracies from these methods is that they achieved comparatively

Table 7 Average precision values for each activity performed by each subject on Dataset1

Activity	S1	S2	S3	S4	S5	S6	S7
Desk work	77.78%	85.61%	90.43%	72.22%	50.00%	89.99%	75.00%
Eating/drinking	44.44%	63.89%	72.50%	33.33%	N.P	84.62%	61.11%
House work	66.67%	37.50%	60.00%	76.67%	60.00%	68.00%	83.33%
Meal preparation	N.P	62.50%	46.67%	43.33%	40.00%	48.00%	55.55%
Movement	82.41%	65.39%	72.00%	96.38%	76.00%	60.00%	100.00%
Personal grooming	73.49%	58.33%	44.00%	64.58%	53.33%	60.67%	77.78%
Relaxing	83.33%	40.95%	60.00%	60.42%	N.P	73.33%	79.17%
Shopping	66.67%	62.50%	70.00%	54.17%	N.P	68.57%	N.P
Sleeping	00.00%	75.00%	80.00%	83.33%	N.P	N.P	100.00%
Socializing	43.06%	75.00%	66.00%	85.00%	45.00%	77.80%	66.67%
Sports	50.00%	72.50%	80.00%	58.33%	20.00%	73.33%	N.P
Taking medication	66.67%	85.00%	80.00%	N.P	N.P	N.P	N.P
Transportation	80.00%	75.00%	60.00%	83.33%	60.00%	86.19%	70.08%

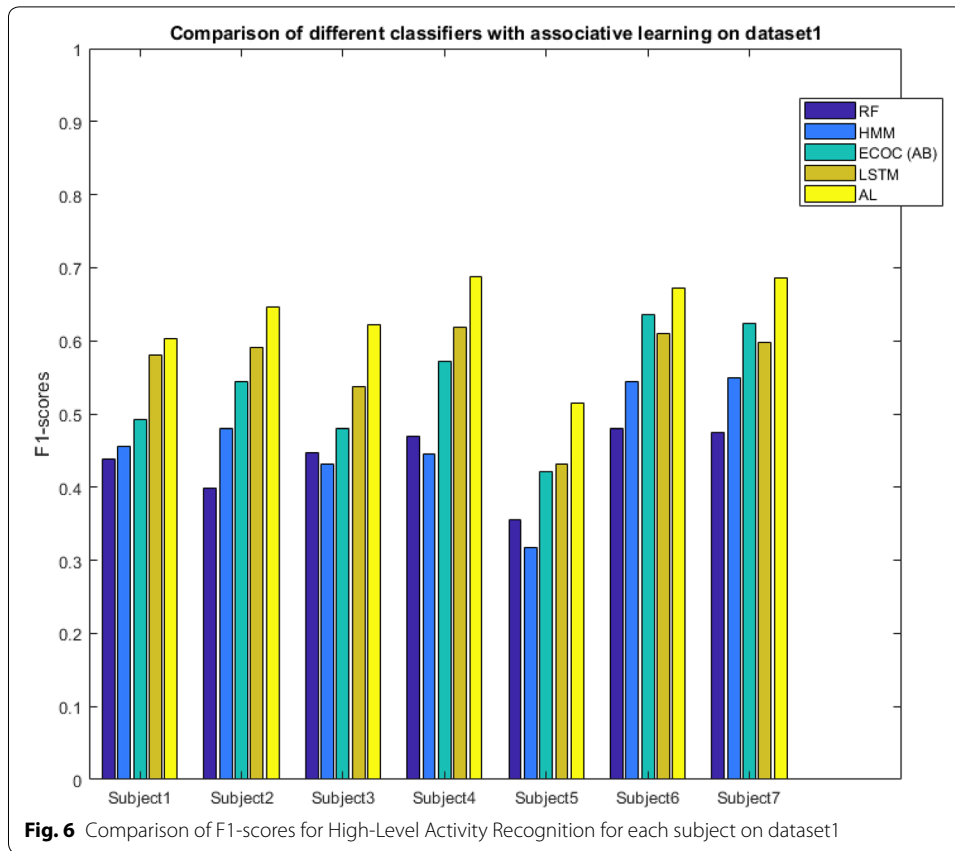
*S, Subject, N.P, Not performed

better accuracy than the other discriminative classification methods. We also present the average precision of each activity for each subject using the proposed AL method in Table 7.

The precision values highlight the false-positive rate (FPR) suggesting that the FPR would be low if the precision is high and vice versa. Holistically, the dataset1 have many instances for the activity ‘DeskWork’ and ‘Movement’. It is reflected by the precision values for these two activities. However, ‘TakingMedication’, ‘Sleeping’, and ‘Shopping’ have fewer occurrences as compared to the other activities. The advantage of using the proposed AL is that even the activities such as ‘TakingMedication’, ‘Sports’, and ‘Sleeping’, occur only twice or thrice for data recording of 12 days, it can generate the corresponding rules and classify the activities with considerable precision. The only instance where the proposed AL failed to recognize ‘Sleeping’ activity was for subject 1 as the activity occurred at only two instances and was performed on a train during transportation. The frequency of the posture for the ‘Sleeping’ activity was also the same as the ‘Transportation’ activity, therefore, the rule was eliminated. Many times the ‘HouseWork’ activity was confused with ‘MealPreparation’ and ‘PersonalGrooming’, the ‘Eating/Drinking’ activity was confused with ‘Relaxing’ and ‘DeskWork’ when performing the activities at location ‘Home’, and the ‘Socializing’ activity was confused with ‘Movement’, ‘Sports’, and ‘Shopping’ when performing the activities at location ‘Street/Road/Pasture’, respectively. The precision values often resulted in “0” for some of the less frequent activities when using other classifiers.

The precision is a good metric but not a complete one as it does not reflect the ratio of correctly positive predicted values to the actual positive ones, i.e. “False Negatives”. In this regard, we show the F1 scores for each subject in Fig. 6 which consider both the false positive and false negative, accordingly.

We present the day-wise accuracies for both the subjects from dataset2 in Table 8. Unlike the results from dataset1, ECOC (AB) achieves second-best results in comparison to LSTM networks. One disadvantage of using LSTM on dataset2 is less



number of observations as compared to dataset1. Although for some days the results are better for ECOC (AB), the results from associative learning are the best amongst all other classifiers for most of the days and overall average accuracy. Similar to the dataset1, we provide the precision values for each activity performed by each subject on dataset2 in Table 9. Most of the instances for both the subjects belong to the activity 'Toileting' which is reflected by its high precision. For subject 1, the activities 'Preparing Dinner', 'Washing Dishes', and 'Cleaning' have the least number of instances which is apparent by their low precision, however, the proposed AL can still classify these activities whereas other classifiers such as RF, HMM, and LSTM yield "0" precision for the said activities. The three activities 'Preparing Breakfast', 'Preparing Lunch', and 'Preparing Dinner' is mostly confused due to the similar sensor activities and frequency. Subject 2 has more uniformly distributed instances for all the activities except 'Toileting' which results in high precision values for most. In addition to the precision values, we show the F1 scores for each subject in Fig. 7.

It is evident from the results that the proposed method not only achieves the best results for the data from inertial measurement units but also achieves the best results for the sensor activations which shows the generalizing ability of CARs on different data modalities. The results show that the classifiers assuming the data to be independent and identically distributed (IID) do not perform well as they do not take into account the relationship of activities being performed with respect to a certain context. The

Table 8 Day wise comparison of accuracies on Dataset2

Days	RF	HMM	ECOC (AB)	LSTM	AL	# of Act
	Subject 1					13
1	48.00%	39.00%	60.50%	54.25%	66.75%	
2	51.33%	55.50%	64.67%	58.00%	71.33%	
3	78.00%	75.00%	91.00%	82.00%	88.00%	
4	55.14%	59.45%	69.43%	62.29%	76.57%	
5	40.67%	29.00%	48.00%	48.00%	56.33%	
6	62.29%	37.18%	75.57%	69.43%	73.71%	
7	52.23%	57.09%	61.64%	57.09%	71.73%	
8	43.44%	41.63%	64.67%	62.29%	70.22%	
9	46.76%	39.36%	62.76%	59.11%	70.22%	
10	48.00%	32.17%	69.43%	66.29%	76.57%	
11	52.23%	38.71%	76.57%	63.29%	69.43%	
12	51.33%	42.00%	58.00%	51.33%	64.67%	
13	40.86%	38.71%	55.14%	49.00%	62.29%	
14	59.11%	36.89%	69.22%	64.67%	75.78%	
15	40.31%	36.62%	48.00%	55.69%	59.54%	
16	32.00%	27.00%	53.00%	48.00%	58.00%	
Avg	50.10%	42.83%	64.22%	59.42%	69.44%	
	Subject 2					9
1	46.15%	38.46%	53.85%	53.85%	53.85%	
2	50.00%	37.50%	62.50%	50.00%	75.00%	
3	38.46%	30.77%	61.54%	53.85%	61.54%	
4	50.00%	41.67%	66.67%	58.33%	75.00%	
5	66.67%	58.33%	83.33%	75.00%	91.67%	
6	42.86%	35.71%	57.14%	50.00%	64.28%	
7	50.00%	42.86%	64.28%	57.14%	64.28%	
8	40.00%	40.00%	60.00%	50.00%	60.00%	
9	50.00%	40.00%	70.00%	60.00%	80.00%	
10	55.56%	33.33%	66.67%	55.56%	66.67%	
11	66.67%	66.67%	100.0%	83.33%	100.0%	
12	66.67%	55.56%	77.78%	66.67%	88.89%	
13	68.75%	62.50%	81.25%	75.00%	87.50%	
14	45.45%	45.45%	63.64%	54.54%	72.73%	
15	50.00%	37.50%	75.00%	62.50%	87.50%	
16	50.00%	50.00%	66.67%	50.00%	66.67%	
Avg	52.33%	44.77%	69.39%	66.43%	74.72%	

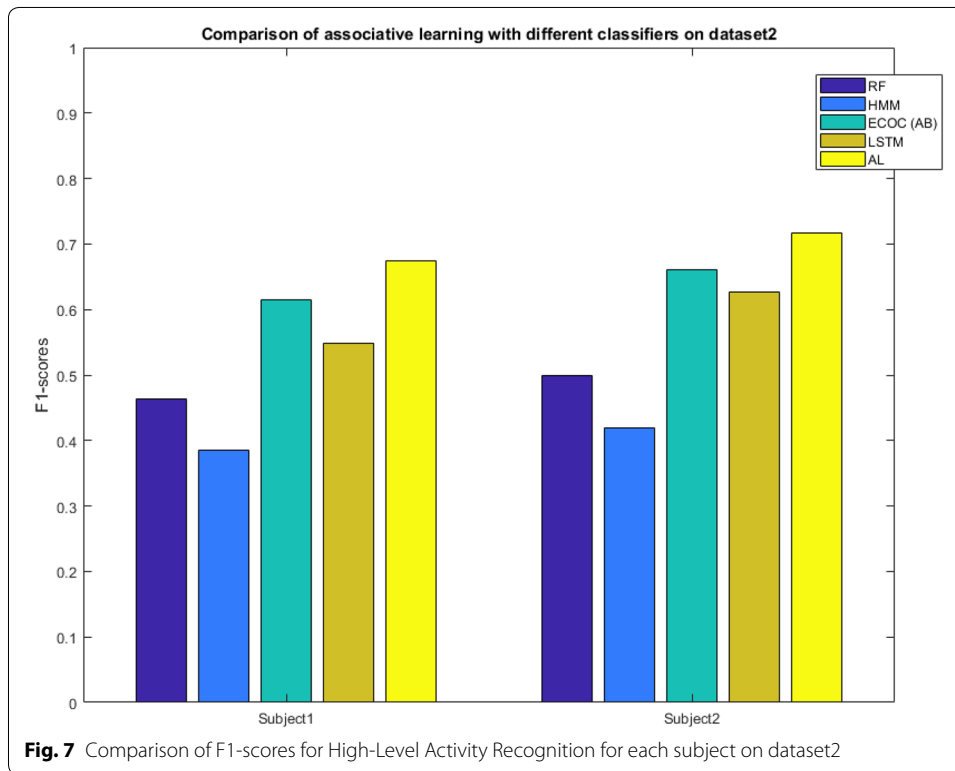
RF, Random Forest; HMM, Hidden Markov Model; ECOC, Error Correcting Output Codes; AB, AdaBoost; LSTM, Long-Short Term Memory Networks; AL, Associative Learning; # of Act, Number of Activities

proposed AL classifier models the activity-context relationship well which results in better recognition performance. For dataset1, the accuracies are not so high and one of the major reasons is the inhomogeneity in the activity labels. As we performed LODO analysis, an activity label that is performed on the testing day may not be available in the training days which results in misclassifications. However, the proposed AL is better at recognizing activities with variations in terms of the interaction with the contexts, i.e. time and location, which is supported by the better accuracy, precision, and F1-scores.

Table 9 Average precision values for each activity performed by each subject on Dataset2

Activity	Activity code	S1	S2
Going out to work	5	80.95%	N.P
Toileting	15	84.72%	82.29%
Bathing	20	59.10%	N.P
Grooming	25	81.77%	N.P
Dressing	30	75.64%	N.P
Taking Medication	40	N.P	68.75%
Preparing breakfast	60	50.00%	85.00%
Preparing lunch	65	46.67%	76.92%
Preparing dinner	70	37.50%	83.33%
Preparing snack	75	80.00%	54.54%
Preparing beverage	80	81.81%	N.P
Washing dishes	85	57.14%	73.08%
Cleaning	100	22.22%	N.P
Doing laundry	105	71.90%	N.P
Watching TV	135	N.P	95.83%
Listening to music	140	N.P	91.67%

*S, Subject, N.P, Not performed



Behavior process modeling of human activities

To evaluate the associative learning method for the behavior process modeling of human activities, we divide the datasets for each subject into two. We generate a reference model from the first ten days and 14 days of annotated activities for each

Table 10 Comparative results for process modeling of human activities on dataset1

Metric	S1	S2	S3	S4	S5	S6	S7
	RF						
Trace fitness	0.78	0.79	0.59	0.60	0.59	0.64	0.69
Generalization	0.65	0.65	0.48	0.57	0.34	0.60	0.61
	HMM						
Trace Fitness	0.80	0.80	0.64	0.67	0.45	0.64	0.68
Generalization	0.68	0.69	0.62	0.60	0.29	0.58	0.65
	ECOC (AB)						
Trace fitness	0.85	0.88	0.81	0.80	0.82	0.83	0.86
Generalization	0.77	0.75	0.69	0.70	0.59	0.71	0.70
	LSTM						
Trace fitness	0.87	0.90	0.82	0.79	0.82	0.82	0.84
Generalization	0.78	0.79	0.68	0.71	0.55	0.69	0.65
	AL						
Trace fitness	0.92	0.97	0.89	0.87	0.87	0.88	0.90
Generalization	0.84	0.87	0.81	0.80	0.71	0.80	0.83

Table 11 Comparative results for process modeling of human activities on dataset2

Metric	S1	S2
	RF	
Trace fitness	0.73	0.69
Generalization	0.69	0.67
	HMM	
Trace Fitness	0.64	0.61
Generalization	0.62	0.59
	ECOC (AB)	
Trace fitness	0.87	0.86
Generalization	0.87	0.82
	LSTM	
Trace fitness	0.79	0.76
Generalization	0.80	0.75
	AL	
Trace fitness	0.91	1.00
Generalization	0.90	0.94

subject from dataset1 and dataset2, respectively. We use the last 2 days' data from both the datasets to predict the activities using RF, HMM, ECOC with AdaBoost, LSTM, and AL. We then compute the trace fitness and generalization parameters for each subject. For the computation of the said parameters and generating process model from activity logs, we used Rapid Miner with RapidPROM extension [67]. An example of the process is provided in the Additional file 1 generated using 80% activities and 80% path. We present the results in Tables 10 and 11, accordingly. The results indicate that our proposed associative learning generates the best alignment of event logs which is reflected by the values of trace fitness. It was also observed that the predicted activity log generalizes quite well to the reference model specified by the generalization parameter.

Table 12 Similarity matrix for all subjects

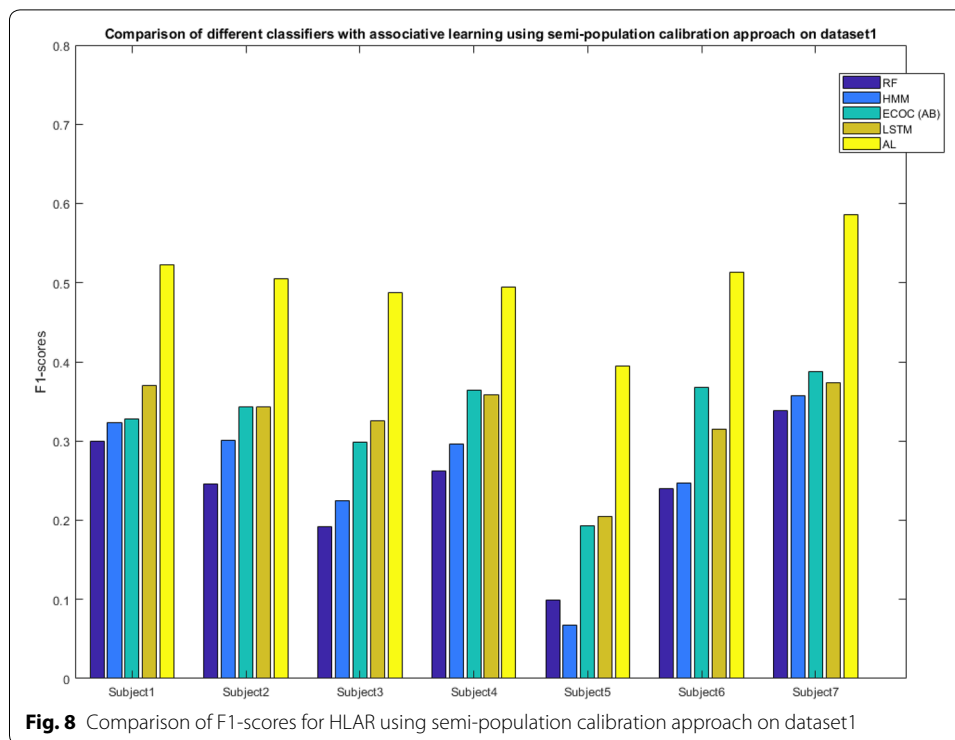
	S1	S2	S3	S4	S5	S6	S7
S1	–	0.6574	0.3676	0.7073	0.3008	0.5740	0.5960
S2	0.6932	–	0.3760	0.5949	0.3851	0.5339	0.6350
S3	0.5471	0.6369	–	0.4225	0.3396	0.5469	0.4235
S4	0.6768	0.5392	0.3822	–	0.4365	0.5600	0.4542
S5	0.5042	0.6109	0.4571	0.6308	–	0.4250	0.4497
S6	0.6950	0.6966	0.3807	0.7275	0.5409	–	0.6274
S7	0.6864	0.6997	0.4846	0.6129	0.6540	0.7307	–

Table 13 Comparative accuracy analysis for activity recognition using semi-population calibration approach

Subjects	RF	HMM	ECOC (AB)	LSTM	AL
Subject 1	34.37%	39.05%	38.30%	42.37%	56.95%
Subject 2	32.09%	37.53%	39.47%	39.36%	54.52%
Subject 3	27.52%	29.87%	35.72%	39.59%	52.17%
Subject 4	30.65%	32.78%	39.48%	38.41%	53.19%
Subject 5	19.87%	14.52%	25.25%	26.02%	44.25%
Subject 6	30.73%	30.96%	39.30%	36.25%	55.00%
Subject 7	36.18%	39.76%	42.67%	42.63%	61.45%

Personalized human activity recognition using CAPHAR

The experimental results for personalization have been performed on dataset1 only due to the number of subjects. For cross-subject analysis, we use a semi-population calibration approach as suggested in [1]. We leave one subject out as the test user and take 2 days of the test user's data to compute the similarity metric from the pool of the remaining subjects. We then conduct activity recognition using RF, HMM, ECOC with AdaBoost, LSTM, and AL to compute the accuracies and F1-scores for performance evaluation on the remaining days of test user's data. We develop the similarity matrix as shown in Table 12 to select the model closest to the test user. For the similarity computation, we set the value of α , β , and γ to be 0.4, 0.4, and 0.2 respectively. We highlight the best similarity of the test user with others in the bold face. It should be noted that the similarity matrix does not follow symmetry property i.e., $Sim(U_1, U_2) \neq Sim(U_2, U_1)$. The reason behind non-symmetrical results is the variation in the number of activities mostly. For instance, subject 1 might perform only 6 annotated activities in his first two days, while the subject 2 has a total of 13 activities. Similarly, subject 2 may perform 6 activities in his first 2 days while subject 1 has a total of 12 activities. The results also affect the TF and Ge values which result in the asymmetric similarity matrix. We observe that subject 3 has a different way of handling the activities as compared to other users which leads to low similarity values. We assume that the reason for subject 5 having low similarity value is less amount (days) of data and less number of activities as compared to other subjects. Based on the similarity values, we perform the activity recognition on the remaining days of data for each test user to evaluate the HLAR performance. The accuracies and F1-scores for each subject are shown in Table 13 and Fig. 8, respectively.



The results show the weakness of conventional discriminative and generative learning algorithms. Although the performance degrades when the model is fitted to a new subject which is obvious in activity recognition field, still the associative learning achieves better results with a good margin, relatively. The highest accuracy was achieved for subject seven whereas the lowest accuracy was recorded for subject 5 which we assume is the same reason for getting low similarity with all subjects in the similarity matrix. The graph in Fig. 8 demonstrates the effectiveness of associative learning in comparison to the other classification algorithms. Due to the capability of AL to model the relationship between the actions/activations and the interaction with contexts, it can cope with the activity handling problem. The strength of AL can be justified by comparing its performance with HMM and LSTM as both the techniques use sequential modeling of the activities. However, the sequence of actions is not enough to enhance the performance, the frequency of actions being performed or interacted with the corresponding context should also be taken into account as considered in AL, accordingly.

Comparison with State-of-the-Art methods

As stated, most of the works have used the opportunity dataset for proving the effectiveness of their work, however, the opportunity dataset has less number of subjects, high-level activities, does not exhibit activity handling problem, and is homogeneous in terms of activity labels. The proposed work deals with all the above-mentioned problems. To prove its effectiveness, we implement some state-of-the-art works proposed in existing studies and apply it on dataset1 to perform a fair comparison. We used the implementation of DeepSense [68], Deep Residual Bidirectional LSTM (DRBLSTM) [69], DeepConvLSTM [3], and Associative Learning [4], respectively. We evaluated the methods with

Table 14 Comparison with state-of-the-art methods on dataset1 with respect to accuracy

Subjects	DeepSense	DRBLSTM	DeepConvLSTM	Wen et al. (Loc)	Wen et al. (Time)	AL
Subject 1	58.73%	59.45%	62.04%	59.27%	56.36%	64.91%
Subject 2	59.26%	63.69%	67.24%	61.43%	60.37%	69.39%
Subject 3	58.94%	62.37%	66.15%	62.23%	60.97%	67.71%
Subject 4	63.87%	67.65%	69.43%	65.52%	61.26%	73.91%
Subject 5	45.44%	48.59%	54.36%	46.20%	44.76%	54.52%
Subject 6	64.91%	68.84%	72.45%	68.26%	63.72%	72.00%
Subject 7	65.54%	69.42%	72.67%	65.36%	63.68%	72.54%
Avg	59.53%	62.86%	66.33%	61.18%	58.73%	67.85%
Avg E.T.T	2.04 s	1.45 s	1.73 s	0.24 s	0.27 s	0.32 s

The best accuracy is represented by boldface

*E.T.T, Execution Time for Testing

Leave-one-day-out (LODO) protocol as performed in the earlier experiments. We used the raw sensor readings of accelerometer and gyroscope along with the location and time-encoded feature vectors as suggested in DeepSense, DRBLSTM, and DeepConvLSTM, instead of directly providing the low-level action label. The associative learning proposed in [4] used only a single context so we apply the method using the location context first and then the time-encoded values. The comparative results for each subject are shown in Table 14. The results show that the associative learning method in [4] works better with the location context in comparison to the time-encoded values. It is shown that the proposed method outperforms the existing ones on all the subjects except 6 and 7, however, the difference is quite low, i.e. 0.45 and 0.13, respectively. The proposed method also achieves better accuracy on average for all subjects in comparison to the existing works. An advantage of the proposed method over the existing ones is also the reduced number of parameters required in optimizing the training process concerning deep learning architectures. The reason for better performance is assumed to be the constrained search space and the pruning policy which our proposed method undergoes for generating association rules. It helps to avoid overfitting for a certain activity label which is important as the labels are not homogeneous and the activities are not always performed in similar way. It is also worth noticing that the execution time of the proposed method is 0.32 s which is quite faster than DeepSense, DRBLSTM, and DeepConvLSTM having 2.04, 1.45, and 1.73 s, respectively.

Conclusion

Recognizing low-level actions has been addressed by many existing studies and detecting only the basic actions such as standing, sitting, walking, running, and lying down are not of many benefits as the daily life activities are much more complex than the basic actions. In this paper, we proposed CAPHAR using the associative learning method based on class association rules with contextual information for personalized human activity recognition. The effectiveness of associative learning when dealing with the activity handling problem is evident with the results shown in Table 14. The proposed method not only achieves the best results but also shows the best trade-off in terms of accuracy and execution time. The personalization effect in CAPHAR has

been achieved by computing the similarity between existing and new users based on their behavior process models. We proved that leveraging association rules for classifying activities performs better than the conventional discriminative and generative classification methods. The results show that associative learning can achieve 18.16%, 16.28%, 8.93%, and 6.00% better accuracy using RF, HMM, ECOC with AdaBoost, and LSTM on dataset1, respectively. This shows that not only the CAPHAR can achieve better personalization results but also can cope with the activity handling problem as users do not perform the activity in the same manner. We also proved that associative learning could perform better on both datasets having different characteristics and modalities, i.e., low-level actions and sensor activations.

We also provided the results for the behavior process modeling of individual user activities and provided a solution to the cold start problem using the proposed similarity metric for personalization. The results for the behavior process modeling of individual users show that associative learning can predict the activities quite well, resulting in better trace fitness and generalization values in comparison to the other classification approaches. This leads to the realization of CAPHAR in the field of human behavior modeling which is more complex.

Furthermore, we trained activity models from different subjects and used these models to predict activities for a new user having a small amount of annotated activity data. The mapping of a new user was performed using the similarity metric based on trace fitness, generalization, and an average number of activities. Experimental results show that the associative learning method can achieve 23.73%, 21.86%, 16.76%, and 16.13% better accuracy in comparison to RF, HMM, ECOC with AdaBoost, and LSTM on dataset1, respectively. The results show the effectiveness of our personalized human activity recognition when provided with a new user having small amount of annotated activity data. We have also shown the qualitative results for F1-scores for LLAR, HLAR, and HLLAR using a semi-population calibration approach.

In this paper, we show a different way of recognizing activities by leveraging the association rules using LLA/SA with contextual information. Although the reported results are better in comparison and show promising aspects to cope with the activity handling problem, there is still room for improving the performance of activity recognition. The limitation of this study is that it uses at least 2 days of the test users' data to generate a process model for mapping. In real-life situations, an online learning approach may be required to deal with this issue so that the model is updated regularly by asking the user for the activity labels. Currently, we only used the associative learning method on two datasets that provide low-level actions/sensor activations data along with location and time information. We also intend to record a large database with more fine-grained activities to leap forward in terms of realization. The extension of the dataset will also allow us to integrate our similarity measure with demographic features to understand the various aspects of human activity and useful insights for variations in activity handling. Furthermore, the associative learning method can be used for large-scale activity recognition from videos by considering the associations of the contexts such as objects and scene (background) rather than the co-occurrences of Spatio-temporal features. It will be interesting to see the association of the action with respect to a certain context and to compare it with the

attention mechanism while dealing with the activity handling issue. The analysis will certainly highlight the benefits of the interactions with the context.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13673-020-00240-y>.

Additional file 1. An example of a process model generated from sensor logs using CAPHAR framework.

Acknowledgements

This research was supported by Hankuk University of Foreign Studies Research Fund, and also supported by National Research Foundation of Korea (NRF) of Ministry of Education (2018R1D1A1B07049113) and National Research Foundation of Korea (NRF) of Ministry of Education (2018R1D1A1B07047241).

Funding

Dr. Seok-Lyong Lee is the recipient of the grant-funded by the Basic Science Research Program through NRF (2018R1D1A1B07049113) and Dr. Bernardo Nugroho Yahya is the recipient of the grant funded by the Ministry of Education through NRF (2018R1D1A1B07047241).

Availability of data and materials

The datasets analyzed in this study are included in the published articles [17, 23] and are available at the following web-sites. Daily Life Log Dataset (https://sensor.informatik.uni-mannheim.de/#dataset_dailylog). Activity recognition in the Home setting Dataset (<https://courses.media.mit.edu/2004fall/mas622j/04.projects/home/>).

Competing interests

The authors declare that they have no competing interests.

Received: 7 March 2020 Accepted: 13 July 2020

Published online: 12 August 2020

References

- Hong J-H, Ramos J, Dey AK (2016) Toward personalized activity recognition systems with a semipopulation approach. *IEEE Trans Human Mach Syst* 46:101–112. <https://doi.org/10.1109/THMS.2015.2489688>
- Rabbi M, Aung MH, Zhang M, Choudhury T (2015) MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In: Proceedings of the 2015 ACM International Joint Conference on pervasive and ubiquitous computing—UbiComp'15. ACM Press, New York. pp 707–718
- Ordóñez F, Roggen D (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16:115. <https://doi.org/10.3390/s16010115>
- Wen J, Zhong M, Wang Z (2015) Activity recognition with weighted frequent patterns mining in smart environments. *Expert Syst Appl* 42:6423–6432. <https://doi.org/10.1016/j.eswa.2015.04.020>
- Kaltz J, Wolfgang JZ, Lohmann S (2005) Context-aware web engineering: modeling and applications. *RIA Revue d'Intelligence Artif Spec Issue Applying Context* 19:439–458
- Khowaja SA, Yahya BN, Lee S-L (2017) Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems. *Expert Syst Appl* 88:165–177. <https://doi.org/10.1016/j.eswa.2017.06.040>
- Villalonga C, Razzaq M, Khan W et al (2016) Ontology-based high-level context inference for human behavior identification. *Sensors* 16:1617. <https://doi.org/10.3390/s16101617>
- Rajasethupathy K, Scime A, Rajasethupathy KS, Murray GR (2009) Finding “persistent rules”: combining association and classification results. *Expert Syst Appl* 36:6019–6024. <https://doi.org/10.1016/j.eswa.2008.06.090>
- Khowaja SA, Prabono AG, Setiawan F et al (2018) Contextual activity based Healthcare Internet of Things, Services, and People (HIoTSP): an architectural framework for healthcare monitoring using wearable sensors. *Comput Networks*. <https://doi.org/10.1016/j.comnet.2018.09.003>
- Filippopolitis A, Oliff W, Takand B, Loukas G (2017) Location-enhanced activity recognition in indoor environments using off the shelf smart watch technology and BLE beacons. *Sensors* 17:1230. <https://doi.org/10.3390/s17061230>
- Mafrur R, Nugraha IGD, Choi D (2015) Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. *Human Centric Comput Inf Sci* 5:31. <https://doi.org/10.1186/s13673-015-0049-7>
- Atzmueller M, Hayat N, Trojahn M, Kroll D (2018) Explicative human activity recognition using adaptive association rule-based classification. In: IEEE International Conference on Future IoT Technologies (Future IoT). IEEE, New York. pp 1–6
- Liu Y, Nie L, Han L, et al (2015) Action2Activity: recognizing complex activities from sensor data. In: Twenty-Fourth International Joint Conference on artificial intelligence. pp 1617–1623
- Liu L, Cheng L, Liu Y, et al (2016) Recognizing complex activities by a probabilistic interval-based model. In: Thirtieth AAAI Conference on artificial intelligence. pp 1266–1272
- Palmes P, Pung HK, Gu T et al (2010) Object relevance weight pattern mining for activity recognition and segmentation. *Pervasive Mob Comput* 6:43–57. <https://doi.org/10.1016/j.pmcj.2009.10.004>
- Chen L, Nugent CD, Wang H (2012) A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans Knowl Data Eng* 24:961–974. <https://doi.org/10.1109/TKDE.2011.51>

17. Szttyler T, Carmona J, Völker J, Stuckenschmidt H (2016) Self-tracking reloaded: applying process mining to personalized health care from labeled sensor data. In: Transactions on perti nets and other models of concurrency. pp 160–180
18. Markovikj D, Gievaska S, Kosinski M, Stillwell DJ (2013) Mining Facebook data for predictive personality modeling. In: Seventh International AAAI Conference on weblogs and social media. pp 23–26
19. Friasmartinez E, Magoulas G, Chen S, Macredie R (2005) Modeling human behavior in user-adaptive systems: recent advances using soft computing techniques. *Expert Syst Appl* 29:320–329. <https://doi.org/10.1016/j.eswa.2005.04.005>
20. Prabono AG, Lee S-L, Yahya BN (2019) Context-based similarity measure on human behavior pattern analysis. *Soft Comput* 23:5455–5467. <https://doi.org/10.1007/s00500-018-3198-6>
21. Fernández-Llatas C, Benedi J-M, García-Gómez J, Traver V (2013) Process mining for individualized behavior modeling using wireless tracking in nursing homes. *Sensors* 13:15434–15451. <https://doi.org/10.3390/s131115434>
22. Hernandez M, Scarr S, Sharma M (2020) The Korean clusters: how coronavirus cases exploded in South Korean churches and hospitals. In: Reuters Graph. <https://graphics.reuters.com/CHINA-HEALTH-SOUTHKOREA-CLUSTERS/0100B5G335B/index.html>. Accessed 4 Mar 2020
23. Tapia EM, Intille SS, Larson K (2004) Activity recognition in the home using simple and ubiquitous sensors. In: International Conference on pervasive computing. pp 158–175
24. Chang Chien Y-W, Chen Y-L (2010) Mining associative classification rules with stock trading data—a GA-based method. *Knowl Based Syst* 23:605–614. <https://doi.org/10.1016/j.knosys.2010.04.007>
25. Pach F, Gyenesei A, Abonyi J (2008) Compact fuzzy association rule-based classifier. *Expert Syst Appl* 34:2406–2416. <https://doi.org/10.1016/j.eswa.2007.04.005>
26. Qodmanan HR, Nasiri M, Minaei-Bidgoli B (2011) Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence. *Expert Syst Appl* 38:288–298. <https://doi.org/10.1016/j.eswa.2010.06.060>
27. Yan X, Zhang C, Zhang S (2009) Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support. *Expert Syst Appl* 36:3066–3076. <https://doi.org/10.1016/j.eswa.2008.01.028>
28. Gu T, Chen S, Tao X, Lu J (2010) An unsupervised approach to activity recognition and segmentation based on object-use fingerprints. *Data Knowl Eng* 69:533–544. <https://doi.org/10.1016/j.datak.2010.01.004>
29. Rashidi P, Cook DJ, Holder LB, Schmitter-Edgecombe M (2011) Discovering activities to recognize and track in a smart environment. *IEEE Trans Knowl Data Eng* 23:527–539. <https://doi.org/10.1109/TKDE.2010.148>
30. Lühr S, West G, Venkatesh S (2007) Recognition of emergent human behaviour in a smart home: a data mining approach. *Pervasive Mob Comput* 3:95–116. <https://doi.org/10.1016/j.pmcj.2006.08.002>
31. Yassine A, Singh S, Alamri A (2017) Mining human activity patterns from smart home big data for health care applications. *IEEE Access* 5:13131–13141. <https://doi.org/10.1109/ACCESS.2017.2719921>
32. Hela S, Amel B, Badran R (2018) Early anomaly detection in smart home: a causal association rule-based approach. *Artif Intell Med* 91:57–71. <https://doi.org/10.1016/j.artmed.2018.06.001>
33. Liu Y, Nie L, Liu L, Rosenblum DS (2016) From action to activity: sensor-based activity recognition. *Neurocomputing* 181:108–115. <https://doi.org/10.1016/j.neucom.2015.08.096>
34. Marimuthu P, Perumal V, Vijayakumar V (2019) OAFPM: optimized ANFIS using frequent pattern mining for activity recognition. *J Supercomput* 75:5347–5366. <https://doi.org/10.1007/s11227-019-02802-z>
35. Ni Q, García Hernando A, de la Cruz I (2015) The Elderly's independent living in smart homes: a characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors* 15:11312–11362. <https://doi.org/10.3390/s150511312>
36. Gong H, Xing K, Du W (2018) A user activity pattern mining system based on human activity recognition and location service. In: IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, pp 1–2
37. Cao L, Wang Y, Zhang B et al (2018) GCHAR: an efficient Group-based Context—aware human activity recognition on smartphone. *J Parallel Distrib Comput* 118:67–80. <https://doi.org/10.1016/j.jpdc.2017.05.007>
38. Zhang W, Qin L, Zhong W, et al (2019) Framework of sequence chunking for human activity recognition using wearables. In: Proceedings of the 2019 International Conference on image, video and signal processing—IVSP 2019. ACM Press, New York. pp 93–98
39. Lee H, Ahn C, Choi N et al (2019) The effects of housing environments on the performance of activity-recognition systems using wi-fi channel state information: an exploratory study. *Sensors* 19:983. <https://doi.org/10.3390/s19050983>
40. Aminikhanghahi S, Cook DJ (2019) Enhancing activity recognition using CPD-based activity segmentation. *Pervasive Mob Comput* 53:75–89. <https://doi.org/10.1016/j.pmcj.2019.01.004>
41. Zhang Y, Tian G, Zhang S, Li C (2020) A knowledge-based approach for multiagent collaboration in smart home: from activity recognition to guidance service. *IEEE Trans Instrum Meas* 69:317–329. <https://doi.org/10.1109/TIM.2019.2895931>
42. Civitarese G, Bettini C, Szttyler T et al (2019) newNECTAR: collaborative active learning for knowledge-based probabilistic activity recognition. *Pervasive Mob Comput* 56:88–105. <https://doi.org/10.1016/j.pmcj.2019.04.006>
43. Zhang Shuai, McClean SI, Scotney BW (2012) Probabilistic learning from incomplete data for recognition of activities of daily living in smart homes. *IEEE Trans Inf Technol Biomed* 16:454–462. <https://doi.org/10.1109/TITB.2012.2188534>
44. Stikic M, Larlus D, Ebert S, Schiele B (2011) Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Trans Pattern Anal Mach Intell* 33:2521–2537. <https://doi.org/10.1109/TPAMI.2011.36>
45. Maekawa T, Kishino Y, Sakurai Y, Suyama T (2013) Activity recognition with hand-worn magnetic sensors. *Pers Ubiquitous Comput* 17:1085–1094. <https://doi.org/10.1007/s00779-012-0556-8>
46. Bianchi V, Bassoli M, Lombardo G et al (2019) IoT wearable sensor and deep learning: an integrated approach for personalized human activity recognition in a smart home environment. *IEEE Internet Things J* 6:8553–8562. <https://doi.org/10.1109/JIOT.2019.2920283>
47. Siirtola P, Koskimäki H, Rönning J (2019) Personalizing human activity recognition models using incremental learning. [arXiv:1905.12628](https://arxiv.org/abs/1905.12628)

48. Burns DM, Whyne CM (2020) Personalized activity recognition with deep triplet embeddings. [arXiv:2001.05517](https://arxiv.org/abs/2001.05517).
49. Vaizman Y, Ellis K, Lanckriet G (2017) Recognizing detailed human context in the wild from smartphones and smart-watches. *IEEE Pervasive Comput* 16:62–74. <https://doi.org/10.1109/MPRV.2017.3971131>
50. Shen C, Li Y, Chen Y et al (2018) Performance analysis of multi-motion sensor behavior for active smartphone authentication. *IEEE Trans Inf Forensics Secur* 13:48–62. <https://doi.org/10.1109/TIFS.2017.2737969>
51. Jalali L, Oh H, Moazeni R, Jain R (2016) Human behavior analysis from smartphone data streams. In: *International Workshop on human behavior understanding*. pp 68–85
52. Soleimani E, Nazerfard E (2019) Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. [arXiv:1903.12489](https://arxiv.org/abs/1903.12489)
53. Huang P-C, Lee S-S, Kuo Y-H, Lee K-R (2010) A flexible sequence alignment approach on pattern mining and matching for human activity recognition. *Expert Syst Appl* 37:298–306. <https://doi.org/10.1016/j.eswa.2009.05.057>
54. Riboni D, Bettini C (2011) COSAR: hybrid reasoning for context-aware activity recognition. *Pers Ubiquitous Comput* 15:271–289. <https://doi.org/10.1007/s00779-010-0331-7>
55. Wang Z, Wu D, Gravina R et al (2017) Kernel fusion based extreme learning machine for cross-location activity recognition. *Inf Fusion* 37:1–9. <https://doi.org/10.1016/j.inffus.2017.01.004>
56. Guan Y, Plötz T (2017) Ensembles of deep LSTM learners for activity recognition using wearables. *Proc ACM Interactive, Mobile Wearable Ubiquitous Technol* 1:1–28. <https://doi.org/10.1145/3090076>
57. Cook D (2012) Learning setting-generalized activity models for smart spaces. *IEEE Intell Syst* 27:32–38. <https://doi.org/10.1109/MIS.2010.112>
58. Hodges MR, Pollack ME (2007) An 'Object-Use Fingerprint': the use of electronic sensors for human identification. In: *UbiComp 2007: International Conference on ubiquitous computing*. Springer Berlin Heidelberg, Berlin, Heidelberg. pp 289–303
59. Chen Z, Chen G (2008) Building an associative classifier based on fuzzy association rules. *Int J Comput Intell Syst* 1:262–273. <https://doi.org/10.1080/18756891.2008.9727623>
60. Tax N, Sidorova N, van der Aalst WMP (2018) Discovering more precise process models from event logs by filtering out chaotic activities. *J Intell Inf Syst*. <https://doi.org/10.1007/s10844-018-0507-6>
61. Hwang I, Jang YJ (2017) Process mining to discover shoppers' pathways at a fashion retail store using a wifi-base indoor positioning system. *IEEE Trans Autom Sci Eng* 14:1786–1792. <https://doi.org/10.1109/TASE.2017.2692961>
62. Verbeek E, van der Aalst WMP (2000) Woflan 2.0 A Petri-Net-based workflow diagnosis tool. In: *International Conference on application and theory of Petri Nets*. pp 475–484
63. Buijs JCAM, van Dongen BF, van der Aalst WMP (2012) On the role of fitness, precision, generalization and simplicity in process discovery. In: *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. pp 305–322
64. Roggen D, Calatroni A, Rossi M, et al (2010) Collecting complex activity datasets in highly rich networked sensor environments. In: *2010 Seventh International Conference on Networked Sensing Systems (INSS)*. IEEE, New York. pp 233–240
65. Kingma DP, Ba LJ (2015) Adam: A Method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)*. pp 1–11
66. Trabelsi D, Mohammed S, Chamroukhi F et al (2013) An unsupervised approach for automatic activity recognition based on hidden Markov Model regression. *IEEE Trans Autom Sci Eng* 10:829–835. <https://doi.org/10.1109/TASE.2013.2256349>
67. van der Aalst WM, Bolt A, van Zelst SJ (2017) RapidProM: mine your processes and not just your data. [arXiv:1703.03740](https://arxiv.org/abs/1703.03740).
68. Yao S, Hu S, Zhao Y, et al (2017) DeepSense: a unified deep learning framework for time-series mobile sensing data processing. In: *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 351–360
69. Zhao Y, Yang R, Chevalier G et al (2018) Deep residual Bidir-LSTM for human activity recognition using wearable sensors. *Math Probl Eng* 2018:1–13. <https://doi.org/10.1155/2018/7316954>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)