Fahad, L. G., Khan, A. U. & Rajarajan, M. (2015). Activity recognition in smart homes with self verification of assignments. Neurocomputing, 149(C), pp. 1286-1298. doi: 10.1016/j.neucom.2014.08.069

Activity recognition in smart homes with self verification of assignments

Labiba Gillani Fahad[a,*], Asifullah Khan[b], Muttukrishnan Rajarajan[a]

[a]*School of Engineering and Mathematical Sciences, City University London, Northampton Square, EC1V 0HB, London, United Kingdom.*
[b]*Pattern Recognition Lab, Pakistan Institute of Engineering and Applied Sciences, Nilore, Islamabad, Pakistan.*

Abstract

Activity recognition in smart homes provides valuable benefits in the field of health and elderly care by remote monitoring of patients. In health care, capabilities of both performing the correct recognition and reducing the wrong assignments are of high importance. The novelty of the proposed activity recognition approach lies in being able to assign a category to the incoming activity, while measuring the confidence score of the assigned category that reduces the false positives in the assignments. Multiple sensors deployed at different locations of a smart home are used for activity observations. For multi-class activity classification, we propose a binary solution using support vector machines, which simplifies the problem to correct/incorrect assignments. We obtain the confidence score of each assignment by estimating the activity distribution within each class such that the assignments with low confidence are separated for further investigation by a human operator. The proposed approach is evaluated using a comprehensive performance evaluation metrics. Experimental results obtained from nine publicly available smart home datasets demonstrate a better performance of the proposed approach compared to the state of the art.

Keywords: Activity recognition, Assisted living, Clustering, Performance evaluation metrics, Classification, Reliability.

1. Introduction

The development of effective, long term and technologically driven solutions in health care improves the living standards of elderly people and patients with chronic physical (low mobility level) and cognitive (Alzheimers disease) impairments [1, 2, 3]. In recent years, different strategies such as telemedicine or telemonitoring have been applied for remote observations [4]. With the further advancements in technology, the concept of a smart home equipped with sensors and actuators that enables people to live independently at home under a continuous monitoring, has got more attention [4, 5]. Activity recognition is a fundamental task in smart homes through which the performed activities such as hand washing, meal preparation, eating, sleeping, appropriate usage of medicines and prescribed physical exercises, can be identified and tracked. A long term analysis of the performed activities can provide important information to doctors about their patient's medical condition that is helpful in the timely prevention of many associated risks.

Low level observations of user and the context information such as location and human-object interactions, is gathered through multiple non-intrusive sensors within the environment. The obtained sensor data is partitioned into multiple segments in order to map them to the activity descriptions known as activity segmentation, where a segment is a consecutive sequence of time instants during which an activity is performed [6]. Activity segmentation is performed using different techniques, sliding windows [7], relative weighting of objects in adjacent activities [8] or pattern mining [9], just to name a few. Segmented activity instances are classified in activity classes using different learning models such as Hidden Markov Model (HMM) [10], Conditional Random Fields (CRF) [11], Naive Bayes (NB) [12], Support Vector Machine (SVM) [13], Artificial Neural Network (ANN) [14, 15], and Decision Tree (DT) [16]. In activity classification, a false assignment could occur due to the unreliable nature of sensor data [17], incorrect execution of an activity [18], similar activities due to overlapping in features [19] or inability of a learning algorithm to assign the correct label [20]. In health care systems, the reliability of activity recognition models is extremely important, therefore along with the correct recognition of activities, a model should also be capable of detecting and avoiding false assignments [20]. Most of the activity recognition approaches while focusing on the segmentation and recognition may ignore false assignments [10, 13, 21]. Additionally the existing approaches that exploit the temporal pattern for recognition assume that the activities follow a certain predefined sequence of events, however a fixed sequence may not always be the case, since a single activity can be performed in different ways by different users [22]. The variation in activity instances may also be observed even in the case of single user repeating the activity a number of times. Therefore, it is also important to consider intra-class and inter-subject differences along with inter-class differences.

In this paper, we propose an intelligent approach namely, *Activity Recognition in Smart Homes with the capability of Self Verification* (ARSH-SV) to recognize the pre-segmented activities of daily life. The proposed approach is able to measure the reliability of the assigned label using a confidence score

*Corresponding author: phone: +44 (0) 20 7040 4073
Email addresses: `labiba.gillani.2@city.ac.uk` (Labiba Gillani Fahad), `asif@pieas.edu.pk` (Asifullah Khan), `r.muttukrishnan@city.ac.uk` (Muttukrishnan Rajarajan)

thus highlighting the activities recognized with less confidence. ARSH-SV exploits the properties of both learning and statistical methods. The activity recognition is performed by learning the differences between the correct and incorrect assignments. Since the correct assignments are far less than the incorrect assignments and each activity class has varying number of activity instances, we apply the learning method SVM. SVM has a better generalization ability for imbalance problems due to considering only the support vectors and is also computationally efficient. We find the confidence score of the assigned label by finding the underlying distribution of the data through sub-clustering within each activity class. In the case of clusters with limited number of instances, the resampling method bootstrap is applied to improve data representation in the training. The validation of the approach using nine smart homes datasets and through a comprehensive performance metrics shows that compared to the existing approaches, ARSH-SV improves the activity recognition by successfully reducing the false assignments.

The rest of the paper is organized as follows: Section 2 discusses the related work on activity recognition. In Sec. 3, the proposed approach is presented. Next, we discuss the datasets used, the evaluation criteria and the analysis of results in Sec. 4. Finally, Sec. 5 draws conclusions.

2. Related work

Activities of daily life can be categorized into (i) physical activities such as sitting, standing, walking, running or falling, and (ii) general activities such as cooking, eating, sleeping, cleaning or grooming. The sensing technology to capture human activity observations is based on either wearable sensors mainly used in physical activities, such as accelerometer and gyroscope [16], or environment interactive sensors used in general activities monitoring, such as light, temperature, motion, pressure and binary contact switch sensors [11]. While the proposed approach is focused more on the general activities category, for a review of the state of the art, we also briefly discuss the existing approaches applied for physical activities.

In physical activities, the information through wearable sensors such as movement patterns extracted from acceleration data [1] obtained from accelerometers are exploited. DT is applied to classify twenty physical activities [16]. An ANN based approach [14] using acceleration features first separates the static (standing, sitting) and dynamic (walking, running) activities and then classifies the activities in each class, where Principal Component Analysis (PCA) is used to obtain the well performing features. ANN is also compared with auto generated and domain knowledge based DTs for activity classification, where auto generated DT shows better accuracy while ANN suffers from over fitting [23]. DT is then combined with ANN in a hybrid classifier model for activity recognition, which merges the prior knowledge of activities with the non-linear classification properties of ANN [24]. Probabilistic Neural Network (PNN) and Fuzzy Clustering based incremental learning method can also be applied for activity recognition [25]. The three classifiers: DT, ANN and SVM, are learned for activity recognition [26], where SVM shows a stable performance compared to others. Finally, an interesting approach on eye movement based activity recognition applies SVM in order to classify six static activities such as browsing a web and reading a printed paper [27].

General activities are recognized by gathering the location information and the user interactions with multiple objects within the environment [10, 28, 29, 30]. A dense sensing platform is used, such as contact switch sensors to monitor the opening and closing of doors, motion sensors to detect the user presence at a particular location, or pressure sensors to indicate the usage of objects, bed or sofa [12, 31, 32]. Switch sensors deployed in multiple objects in a home such as doors, windows, cupboards and refrigerator, can be used in the NB classifier based recognition approaches [12, 30], where NB identifies the activity corresponding to the sensor values with the highest probability. PNN classifier [15, 33] derived from Bayesian and Fisher discriminant analysis (FDA) can be applied to estimate the likelihood of a sample being part of a learned activity class. A cluster based classification approach [34] groups the similar activities into clusters, while learning is performed within each cluster. The Evidence Theoretic KNN (ET-KNN) is applied to recognize the activities [34, 35], where neighborhood of each pattern to be classified is considered as an evidence supporting certain hypothesis associated with the class membership of that pattern. The class with the maximum supporting evidence is assigned to the pattern, while the parameters are optimized by the error minimizing function in [36]. In order to exploit the semantic information of domain knowledge, sensor data and activities, context lattices are applied for activity recognition (CL-AR) [37]. HMM is applied and compared with CRF [11] and in order to get a more generalized activity recognition approach, NB, HMM and CRF are compared for activities within a dataset and by combining the common activities of multiple datasets with different environmental settings [30]. HMM requires a large set of training samples and unlike CRF it may not be able to capture long range dependencies of observations [19]. Long range dependencies between the observations within activity segments can be modeled by integrating sequential pattern mining, used to characterize the time spans during an activity execution, with Hidden Semi-Markov Model (HSMM) for the activity recognition (AR-SPM) [6]. Switching-HSMM defines two layers to recognize the daily activities and to identify anomalies [32]. In Hierarchical-HMM (HHMM) two layers are defined [38], where one layer presents the activities, while the second corresponds to the clusters of actions in an activity. HHMM proves to be more effective than HSMM and HMM.

Activities can be recognized by Frequent Pattern (FP) mining followed by the Emerging Pattern (EP): a discriminative pattern used for classification between patterns of activities [39, 40]. Discontinuous FPs are mined to cluster similar activity patterns into groups and then HMM is applied for the recognition [10]. FP mining is used to find the repetitive patterns and Latent Dirichlet Allocation is applied to cluster the co-occurring sequential patterns in order to recognize the activities (ADR-

SPLDA) [21]. The pattern mining and sequence alignment methods can be used to select the representative patterns of activities, which are then matched with the observed sequences to recognize the activities [41]. The Inter-transaction Association Rule (IAR) mining finds the frequent events, while anomalies are identified by using emergent IAR that highlights the abrupt changing points in the dataset [42]. An Active Learning approach in the presence of Overlapping activities (AALO) performs location based frequent item set mining to find the activity patterns and then density based clustering is applied to form the activity clusters [29]. Soft-SVM (C-SVM), Linear Discriminant Analysis (LDA) and CRF classification approaches are compared in handling the issue of imbalanced data in recognizing daily activities [43].

Feature selection techniques have been applied in the activity recognition problem to select significant and discriminant subset of features [13, 27, 31, 44]. Minimum Redundancy Maximum Relevance (mRMR) can be used to select the best feature subset for target classes, then SVM is used for the classification of activities [27]. The mRMR and Information Gain (IG) methods are applied for feature selection and then four classifiers SVM, LogitBoost, Bayes Belief Networks and ANN are compared for recognition of activities [31], where LogitBoost proves to be better among the four due to its ability of generating a stronger classifier by combining the several weak classifiers. In a PCA-SVM approach [13], PCA is used to extract the significant features and the multi-class SVM (one-versus-one) is applied for classification of activities. Using SVM for classification, the comparison of the feature selection methods, filter and wrapper based on single and sequential feature selection shows that wrapper method based on sequential feature selection performs better [44], since it also considers the redundancy of features during the selection process, unlike the other two.

Most of the existing activity recognition approaches focus on the segmentation and recognition of the activities, while a few deal with the uncertainty and noise in the data at sensor level [17, 28, 45, 46]. An evidential fusion approach exploits the combination rule of Dempster Shafer Theory (DST) to support conflict resolution in activity recognition (EFA-AR) by combining the sensed information and commonsense knowledge [45]. The two data fusion and reasoning approaches DST and Dynamic Bayesian Network are compared and their applicabilities are discussed under activity recognition scenario in [47]. The temporal information of domain knowledge is incorporated in the DST of evidence, where the activity start time and duration is used in the Evidence Decision Network (EDN) to recognize the daily activities [28, 48]. A three layered framework for monitoring the human activities in a smart home, exploits DST, the prior use of domain knowledge and the activity patterns in order to manage the uncertainty of sensor data [46]. In a data fusion approach for activity recognition, ontology is applied to represent the hierarchical interrelationships between sensors, contexts and activities, while DST is used for the reasoning under uncertainty [17].

The existing activity recognition approaches address the problem of uncertainty in observations at data fusion level. In the proposed approach, we minimize the uncertainty of observations at the decision level through the verification of assigned labels while utilizing the environment interactive sensors for the recognition of general activities.

3. Activity recognition with self verification

ARSH-SV assigns the label of an activity class to the pre-segmented activity instance, while separating more reliable assignments from the less reliable based on the confidence score. The block diagram of ARSH-SV is shown in Fig. 1. The approach is discussed in the following sub-sections and summarized in Algorithm 1.

3.1. Activity representation

Let $\mathbf{A} = \{A_1, ..., A_k, ..., A_K\}$ be a set of K activity classes and $\mathbf{I}_k = \left\{I_{1k}, ..., I_{jk}, ..., I_{Jk}\right\}$ be a set of J pre-segmented activity instances of an activity class A_k, in the training data. Each I_{jk} is observed by R binary sensors installed at multiple locations in a smart home. I_{jk} is represented by a feature set $\mathbf{F}_{jk} = \left\{f^r_{jk}\right\}^R_{r=1}$ of R features. The number of features in the feature set is equal to the number of installed sensors. Each feature represents the number of times a sensor is activated in an activity instance, while for a non-activated sensor a zero is assigned. We normalize the feature set $\mathbf{F}_{jk}$ such that $0 \leq \hat{f}^r_{jk} \leq 1$ given by

$$\hat{\mathbf{F}}_{jk} = \left\{\hat{f}^r_{jk}\right\}^R_{r=1} = \left\{\frac{f^r_{jk} - \min_r f^r_{jk}}{\max_r f^r_{jk} - \min_r f^r_{jk}}\right\}^R_{r=1}. \tag{1}$$

The normalized feature sets are used in the learning model for the label assignment and in the measurement of the confidence of that label.

In order to identify the intra-class variations in the performance and the location, we group the similar activity instances within an activity class A_k into sub-clusters using the Lloyd's clustering algorithm [49]. Each sub-cluster S_{uk} aims at minimizing the error objective function that measures the distance of the activity instances from their respective cluster centers given as

$$\sum_{u=1}^{U_k} \sum_{j=1}^{J} ||I_{jk} - C_{uk}||^2, \tag{2}$$

where $\mathbf{C}_{uk}$ is R dimensional cluster-center for each S_{uk}. We get $u = 1 \cdots U_k$ sub-clusters for A_k, after removing any empty clusters in case of less intra-class variations.

In order to measure the inter-class variations, we represent A_k by Mean Feature Representation (MFR). We obtain a set of mean features $\mathbf{M}_k = \{m^r_k\}^R_{r=1}$ using all the activity instances $\{I_{jk}\}^J_{j=1}$ of A_k in the training data such that

$$m^r_k = \frac{1}{J} \sum_{j=1}^{J} \hat{f}^r_{jk}, \tag{3}$$

where m^r_k is the mean representation of each feature in $\mathbf{M}_k$ and $k = 1 \cdots K$.

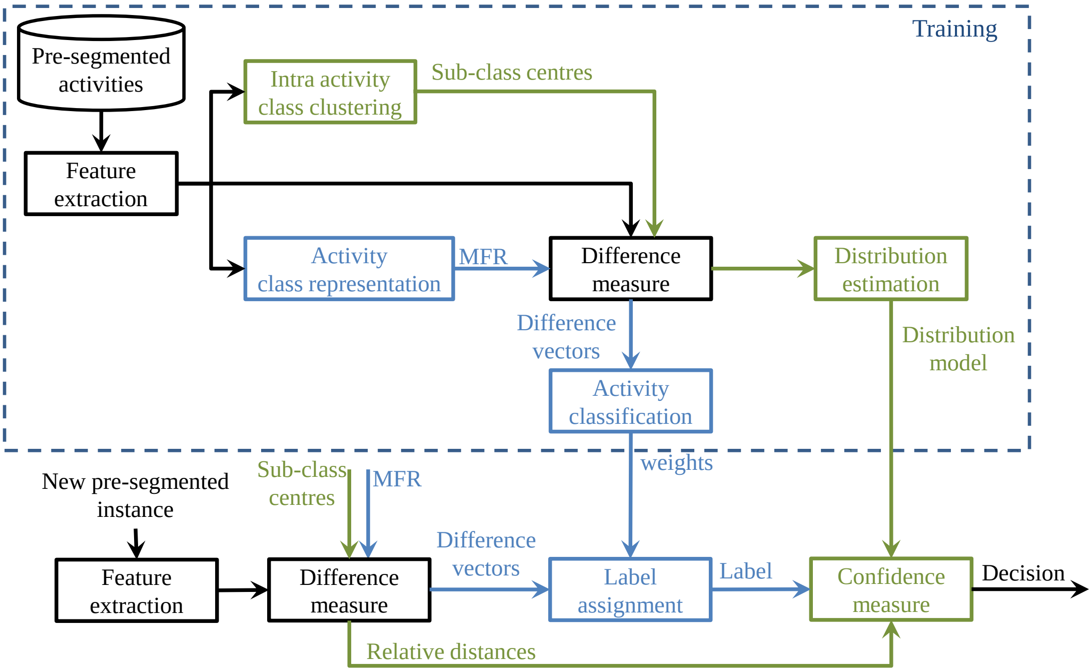


Figure 1: Block diagram of the proposed approach. Black-blue colors represent label assignment and Black-green colors represent confidence measure.

3.2. Activity classification

In order to learn the distances between correct and incorrect assignments, we calculate for each activity instance I_{jk}, a set $\mathbf{\Delta}_{jk} = \{\mathbf{\Omega}^d_{jk}\}^K_{d=1}$ of difference vectors where $\mathbf{\Omega}^d_{jk}$ is a difference vector between the feature set $\hat{\mathbf{F}}_{jk}$ and $\mathbf{M}_d$, where $d = 1 \cdots K$ given as

$$\mathbf{\Delta}_{jk} = \{\mathbf{\Omega}^d_{jk}\}^K_{d=1} = \{\hat{\mathbf{F}}_{jk} - \mathbf{M}_d\}^K_{d=1}. \tag{4}$$

$\mathbf{\Delta}_{jk}$ contains one difference vector $\mathbf{\Omega}^k_{jk}$ from the MFR of the same class ($k = d$) and $K - 1$ difference vectors $\{\mathbf{\Omega}^d_{jk}\}_{k \neq d}$ from MFR of other classes. We then solve the binary class problem such that for each $\mathbf{\Omega}^d_{jk}$ a binary target T_{jk} is assigned as

$$T_{jk} = \begin{cases} 1, & \text{if } k = d; \\ -1, & \text{otherwise.} \end{cases} \tag{5}$$

If each activity class has J instances then by using Eq. (5), we get $J \times K$ number of $\mathbf{\Omega}^k_{jk}$ with $T_{jk} = 1$ and $J \times K \times (K - 1)$ number of $\mathbf{\Omega}^d_{jk}$ with $T_{jk} = -1$. Fewer positive compared to the negative classes represent a class imbalance problem by $1 : K$.

We use the learning method SVM for the activity classification. SVM is a binary classifier and finds the most optimal hyperplane to discriminate the data points of two classes with maximum margin [50]. SVM supports the efficient learning of linear and non-linear functions through kernel trick in a variety of classification problems [51, 52].

Consider the training sample set $(\mathbf{\Omega}_i, T_i)$, where for simplicity $i = jk$ with an upper limit of $J \times K^2$. $\mathbf{\Omega}_i$ is R dimensional feature vector, T_i is the target and can have the values of -1 and $+1$. SVM learns on the correct and incorrect differences. A hyperplane is defined by $\mathbf{w}_i.\mathbf{\Omega}_i + b = 0$, where $\mathbf{w}_i$ is R dimensional weight vector, $\mathbf{\Omega_i}$ is R dimensional data point on the hyperplane and b is the bias. In the case of linearly separable data, the separating hyperplane is defined by:

$$T_i(\mathbf{w}_i.\mathbf{\Omega}_i + b) \geq 1, \tag{6}$$

where $\mathbf{\Omega}_i$ nearest to the boundary are support vectors and defined by $T_i(\mathbf{w}_i.\mathbf{\Omega_i} + b) = 1$. The optimal hyperplane can be constructed by solving the optimization problem $min_{\mathbf{w}_i} \left\{\frac{1}{2}\|\mathbf{w}_i\|^2\right\}$ under the constraints of Eq. 6. As in our problem the data is non-linearly separable, a slack variable ξ_i is introduced for the non-linear support vector machine

$$T_i(\mathbf{w}_i.\mathbf{\Omega}_i + b) > 1 - \xi_i. \tag{7}$$

A penalty term $C \sum_{i=1}^{J \times K^2} \xi_i$ is usually added, when ξ_i is very large for a solution. The optimization problem is then given by:

$$min_i \left[\frac{1}{2}\|\mathbf{w}_i\|^2 + C \sum_{i=1}^{J \times K^2} \xi_i\right], \tag{8}$$

under the constraints of Eq. 7, where C is a positive regularization constant. It defines the trade-off between a large margin and misclassification error. ξ_i controls the distance of $\mathbf{\Omega}_i$ from the decision boundary.

For non-linear problems, the training data is mapped from an input space to a high dimensional feature space $\mathbf{H}$ through a mapping function ϕ, which fits a hyperplane s.t $\mathbf{\Omega}_i \rightarrow \mathbf{H}$. The input data point in the high dimensional space $\mathbf{H}$ is presented by $\phi(\mathbf{\Omega}_i)$. The computational complexity of SVM in the feature space is reduced by using a positive definite kernel function (Ker)

$$\phi(\mathbf{\Omega}).\phi(\mathbf{\Omega}_i) = Ker(\mathbf{\Omega}.\mathbf{\Omega_i}), \tag{9}$$

which leads to a final decision function:

$$f(\mathbf{\Omega}) = \ sign\ (\sum_{i=1}^{J\times K} \alpha_i T_i Ker(\mathbf{\Omega}.\mathbf{\Omega}_i) + b), \tag{10}$$

where α_i represents the Lagrange multiplier and a Linear, a Radial Basis or a Polynomial kernel function can be used. The kernel function must satisfy the Mercer condition [50].

3.3. Distribution estimation

In order to measure the confidence score of the recognized activity instance, we estimate the distribution of data from the sub-clusters defined within each activity class. For the activity instances within each sub-cluster S_{uk} represented by $I_{jk}^{(u)}$, we measure their relative distances $D(\cdot)$ from the cluster-center $\mathbf{C}_{uk}$. $D(\cdot)$ receives a feature pair as input and returns the dissimilarity δ_{jk} between $\hat{\mathbf{F}}_{jk}^{(u)}$ and $\mathbf{C}_{uk}$

$$\delta_{jk}^{u} = D(\hat{\mathbf{F}}_{jk}^{(u)}, \mathbf{C}_{uk}) = ||\hat{\mathbf{F}}_{jk}^{(u)} - \mathbf{C}_{uk}||, \tag{11}$$

where $||.||$ is the Euclidean norm. The obtained set of dissimilarities $\mathbf{V}_k^u = \{\delta_{jk}^u\}_{j=1}^{J_{uk}}$ for each sub-cluster S_{uk} within the activity class A_k is used to find the distribution of activity instances from $\mathbf{C}_{uk}$. Assuming a Gaussian distribution [18], we apply the bootstrap method [53] to obtain the mean μ_{uk} and the standard deviation σ_{uk} of the distribution in the set $\mathbf{V}_k^u$. Bootstrap is a computational method to increase the accuracy of an estimate. It is particularly useful in cases, where sufficient amount of data is not available or the distribution of underlying data is unknown. Bootstrap generates Z number of independent and identically distributed bootstrap estimates by resampling from the original sample, where Z can be a large value. From the set of dissimilarities $\mathbf{V}_k^u$ within the cluster S_{uk}, a bootstrap resample $\mathbf{V}_k^{u(z)} = \{\delta_{jk}^{u(z)}\}_{j=1}^{J_{uk}}$ is drawn at random by reassigning and recomputing the values of original sample such that the values in the rearranged sample may appear zero, once or multiple times. For each resample $S_{uk}^{(z)}$, the mean $\mu_{uk}^{(z)}$ and the standard deviation $\sigma_{uk}^{(z)}$ is calculated as

$$\left\{\mu_{uk}^{(z)}, \sigma_{uk}^{(z)}\right\} = \left\{\frac{1}{J_{uk}}\sum_{j=1}^{J_{uk}} \delta_{jk}^{u(z)}, \sqrt{\frac{1}{J_{uk}}\sum_{j=1}^{J_{uk}} (\delta_{jk}^{u(z)} - \mu_{uk}^{(z)})^2}\right\}. \tag{12}$$

Finally, the true mean μ_{uk} and true standard deviation σ_{uk} is measured from the bootstrapped means $\{\mu_{uk}^{(z)}\}_{z=1}^{Z}$ and standard deviation $\{\sigma_{uk}^{(z)}\}_{z=1}^{Z}$ such that

$$\{\mu_{uk}, \sigma_{uk}\} = \left\{\frac{\sum_{z=1}^{Z}\mu_{uk}^{(z)}}{Z}, \frac{\sum_{z=1}^{Z}\sigma_{uk}^{(z)}}{Z}\right\}. \tag{13}$$

The estimated $\{\mu_{uk}, \sigma_{uk}\}_{u=1}^{U_k}$ for each A_k are then used in measuring the confidence score of the assigned label in the test data.

Algorithm 1 Activity recognition with self verification

A_k : k^{th} activity class;
I_{jk} : j^{th} activity instance of k^{th} class;
K : total number of activity classes;
J : total activity instances per class;
$\hat{\mathbf{F}}_{jk}$: normalized feature set of I_{jk};
$\mathbf{M}_k$: mean feature representation of A_k;
S_{uk} : u^{th} sub-cluster of A_k;
$\mathbf{C}_{uk}$: center of S_{uk};
U_k : total number of sub-clusters in A_k;
μ_{uk} : mean of data distribution in S_{uk}
σ_{uk} : standard deviation data distribution in S_{uk}
$\mathbf{\Omega}_{jk}$: difference vectors;
δ_{jk}^u : dissimilarity $I_{jk}^{(u)}$ from $\mathbf{C}_{uk}$

TRAINING
Input: $J \times K$ normalized feature sets $\hat{\mathbf{F}}_{jk} = \{\hat{f}_{jk}^r\}_{r=1}^R$
for $k = 1$ to K **do**
  for $r = 1$ to R **do**
    $m_k^r = \frac{1}{J}\sum_{j=1}^{J} \hat{f}_{jk}^r$
  end for
  $M_k = \{m_k^r\}_{r=1}^R$
end for
for $k = 1$ to K **do**
  for $j = 1$ to J **do**
    for $d = 1$ to K **do**
      $\mathbf{\Omega}_{jk}^d = \hat{F}_{jk} - M_d$
    end for
    $\mathbf{\Delta}_{jk} = \{\mathbf{\Omega}_{jk}^d\}_{d=1}^K$
  end for
end for
Assign labels T_{jk}^d using Eq.5
Input: $(\mathbf{\Omega}_{jk}^d, T_{jk}^d)$ to learning model SVM
Output: weight vector $\mathbf{w}_i$
for $k = 1$ to K **do**
  for $u = 1$ to U_k **do**
    for $j = 1 to J_{uk}$ **do**
      $\delta_{jk}^u = ||\hat{\mathbf{F}}_{jk}^{(u)} - \mathbf{C}_{uk}||$
    end for
    $\mathbf{V}_k^u = \{\delta_{jk}^u\}_{j=1}^{J_{uk}}$
  end for
end for
Obtain $\{\mu_{uk}, \sigma_{uk}\}_{u=1}^{U_k}$ for A_k using Eq.12 and Eq.13

χ : pre-segmented new activity instance;
k^* : label of recognized activity class;
u^* : label of the sub-cluster in A_{k^*};
$\Gamma_{\chi k^*}^{u^*}$: confidence score of χ_{k^*};
LABEL ASSIGNMENT
Extract feature set $\hat{\mathbf{F}}_{(\chi)}$
for $k = 1$ to K **do**
  $\mathbf{\Omega}_{(\chi)k} = \hat{\mathbf{F}}_{(\chi)} - \mathbf{M}_k$
end for
for $k = 1$ to K **do**
  $P_{(\chi)k} = \mathbf{\Omega}_{(\chi)k} \times \mathbf{w}_i$
end for
$k^* = \arg\max_k P_{(\chi)k}$
Activity instance gets the label as χ_{k^*}
for $u = 1$ to U_{k^*} **do**
  $\delta_{(\chi)k^*}^u = ||\hat{\mathbf{F}}_{(\chi)} - \mathbf{C}_{uk^*}||$
end for
$u^* = \arg\min_u \delta_{(\chi)k^*}^u$
Obtain $\Gamma_{(\chi k^*}^{u^*}$ of χ_{k^*} using Eq. 17
if $\Gamma_{(\chi)k^*}^{u^*} \geq 3\sigma_u^* k^*$ **then**
  Flag χ_{k^*} with low confidence score
end if

3.4. Label assignment

Let χ be a newly detected activity instance. Our objective is to find the correct label of χ from the existing activity classes in $\mathbf{A}$. We extract the normalized feature set $\hat{\mathbf{F}}_{(\chi)}$ using Eq. 1. $\hat{\mathbf{F}}_{(\chi)}$ can have similarity with multiple activity classes with different degree of relevance. Using the feature set $\hat{\mathbf{F}}_{(\chi)}$ (Eq. 1) and the mean feature sets $\{\mathbf{M}_k\}_{k=1}^{K}$ (Eq. 3), we find the set of difference vectors $\mathbf{\Delta}_{(\chi)} = \{\mathbf{\Omega}_{(\chi)}^{k}\}_{k=1}^{K}$ using Eq. 4. We find the cross product $P_{(\chi)k}$ between each difference vector $\mathbf{\Omega}_{(\chi)k}$ and the weight vector $\mathbf{w}_i$ obtained from SVM

$$P_{(\chi)k} = \mathbf{\Omega}_{(\chi)k} \times \mathbf{w}_i. \tag{14}$$

$P_{(\chi)k}$ generates a similarity score between χ and A_k. The higher the value of $P_{(\chi)k}$ the closer χ is to A_k. We find the activity class with the highest value of $P_{(\chi)k}$ given by

$$k^* = \arg\max_k P_{(\chi)k}, \tag{15}$$

where $k^* \in 1, \cdots, K$ is the ID of the recognized activity class. The new activity instance is assigned to A_{k^*} represented as χ_{k^*}.

In order to find the confidence score of the assigned label, we exploit the prior knowledge of $\{\mu_{uk^*}, \sigma_{uk^*}\}_{u=1}^{U_k^*}$ of the distribution of activity instances in the sub-clusters within the assigned activity class A_{k^*}. We find the dissimilarity measures $\{\delta_{(\chi)k^*}^{u}\}_{u=1}^{U_{k^*}}$ between χ_{k^*} and $\{C_{uk^*}\}_{u=1}^{U_{k^*}}$ of A_{k^*} (Eq. 11). Next, we assign χ_{k^*} to the sub-cluster with the least dissimilarity as

$$u^* = \arg\min_u \delta_{(\chi)k^*}^{u}. \tag{16}$$

Using $\{\mu_{u^*k^*}, \sigma_{u^*k^*}\}$ of the selected cluster $S_{u^*k^*}$ in the Gaussian distribution, we measure the confidence score $\Gamma_{(\chi)k^*}^{u^*}$ as

$$\Gamma_{(\chi)k^*}^{u^*} = g(\delta_{(\chi)k^*}^{u^*}, \mu_{u^*k^*}, \sigma_{u^*k^*}) = \frac{1}{\mu_{u^*k^*}\sqrt{2\pi}} e^{\frac{-(\delta_{(\chi)k^*}^{u^*} - (\mu_{u^*k^*})^2}{2(\sigma_{u^*k^*})^2}}. \tag{17}$$

If $\Gamma_{(\chi)k^*}^{u^*}$ lies within $3\sigma_{u^*k^*}$ or 99% of the distribution of $S_{u^*k^*}$, we assign a high confidence to the label, otherwise a low confidence score is assigned. We highlight the labels with the low confidence and separate those instance for further analysis by a human operator.

4. Evaluation and discussion

The objective of this evaluation is to analyze the effectiveness of ARSH-SV in recognizing the general activities and measuring the confidence of the recognitions in a smart home. ARSH-SV is evaluated through a comprehensive performance evaluation metrics containing seven performance measures using nine publicly available smart home datasets. The results are compared with the state-of-the-art activity recognition approaches [13, 15, 30, 35] using three-fold cross validation. The recognition rate (number of correctly recognized instances out of all instances in the test set) and F1-score (Eq. 23) are also compared with [21, 29, 43, 54] and [6, 37, 38, 45, 48], respectively. We perform the activity level performance analysis through confusion matrices and F1-score. The results are compared with [6, 28, 29, 48].

4.1. Performance evaluation metrics

The performance evaluation metrics is comprised of Precision, Recall, Specificity, False Positive Rate (FPR), Matthews Correlation Coefficient (MCC), F1-score and Accuracy, obtained from True Positives (TPs), False Negatives (FNs), True Negatives (TNs) and False Positives (FPs) in the recognized activity instances. For an activity class A_k, *TP* are the number of instances correctly assigned as A_k, while *FN* are the number of instances of A_k incorrectly recognized as any other class (missed assignments of A_k). *TN* are the number of instances correctly recognized as not from A_k, while *FP* are the instances of other classes recognized as A_k.

Precision is the ratio of correctly labeled instances out of the total recognized instances of a class. A high precision value indicates the presence of a higher number of correctly assigned labels out of the total assigned labels to an activity class.

$$Precision = \frac{TP}{TP + FP} \times 100. \tag{18}$$

Recall is the percentage of correctly labeled instances from the total instances of that class. Recall represents the ability of a classifier to return the most of the correct labels out of the total correct labels. It is also known as 'Sensitivity'. Precision and recall together shows the ability of an approach in retrieving the correct label with consistent performance.

$$Recall = \frac{TP}{TP + FN} \times 100. \tag{19}$$

Specificity, also known as 'True Negative Rate (TNR)', is the percentage of correctly assigned labels to other classes out of the total labels assigned to other classes.

$$Specificity = \frac{TN}{TN + FP} \times 100. \tag{20}$$

FPR is the percentage of incorrectly recognized instances. FPR is 100−Specificity.

$$FPR = \frac{FP}{TN + FP} \times 100. \tag{21}$$

MCC measures the overall performance of a classifier, which takes into account TPs, TNs, FPs and FNs. The range of MCC is between $[-1\ ,\ 1]$. A value of 1 indicates a perfect prediction of the classifier, 0 indicates a random prediction, where as −1 shows the conflict between the prediction and observation. We have used MCC in the performance evaluation, since it is considered a balanced measure in the case of activity classes with different number of instances.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{22}$$

F1-score combines precision and recall of the system into a single measure, which is the weighted average of precision and recall. Its value is between $[0\ ,\ 1]$. F1-score closer to 1 shows the best performance, whereas 0 indicates the worst performance.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{23}$$

Table 1: The description of nine datasets used for the evaluation of ARSH-SV. Key: R1 - Resident 1, R2 - Resident 2, 1+pet - One resident and a pet, 2+pet - Two residents and a pet.

S.no	Datasets	Description	Participants	Activity classes	Activity instances	Name of activities
1	Aruba1	Daily life, 2010-2011	1	11	6477	Bed to Toilet, Eating, Enter Home, House Keeping, Leave Home, Meal Preparation, Relax, Resperate, Sleeping, Wash Dishes, and Work.
2	Kasteren	Kasteren	1	7	245	Breakfast, Dinner, Drink, Leave Home, Shower, Sleep, and Toileting.
3	Kyoto1	ADL Activities	20	5	120	Cleaning, Cooking, Eating, Phone Call, and Wash Hands.
4	Kyoto2	ADL Activities with Errors	20	5	100	Cleaning, Cooking, Eating, Phone Call, and Wash Hands.
5	Kyoto8	Daily life, Summer 2009	2	11	1290	Cleaning, Cooking, Grooming, R1 Shower, R1 Sleep, R1 Wakeup, R1 Work, R2 Shower, R2 Sleep, R2 Wakeup, and R2 Work.
6	Tulum1	Daily life, 2009	2	10	1513	Cook Breakfast, Cook Lunch, Enter Home, Group Meeting, Leave Home, R1 Eat Breakfast, R1 Snack, R2 Eat Breakfast, Wash Dishes, and Watch TV.
7	Tulum2	Daily life, 2009-2010	2	16	12637	Bathing, Bed to Toilet, Eating, Enter Home, Leave Home, Meal Preparation, Personal Hygiene, R1 Sleeping, R2 Sleeping, Wash Dishes, Watch TV, Work Bedroom1, Work Bedroom2, Work Living Room, Work Table, and Yoga.
8	Milan	Daily life, 2009	1+pet	15	2310	Bed to Toilet, Chores, Desk Activity, Dining Room Activity, Evening Medications, Guest Bathroom, Kitchen Activity, Leave Home, Master Bathroom, Master Bedroom Activity, Meditate, Morning Medications, Read, Sleep, and Watch TV.
9	Cairo	Daily life, 2009	2+pet	13	600	Bed to Toilet, Breakfast, Dinner, Laundry, Leave Home, Lunch, Night Wandering, R1 Sleep, R1 Wake, R1 Work in Office, R2 Sleep, R2 Take Medicine, and R2 Wake.

Finally, the *Accuracy* of the system is measured, which takes into account not only the correctly assigned labels (TPs) but also the activities that are correctly rejected (TNs) out of the total assigned labels

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100. \qquad (24)$$

4.2. Datasets

Table 1 shows the summary of activities performed in each of the nine publicly available challenging smart home datasets: eight from CASAS [55] and one from Kasteren [11]. Datasets are selected based on the challenges that include the number of participants in each home, overlapping of features among activity instances of different classes, variations in the performance of the same activity instances by different users, addition of noise and errors in the the datasets because of the sensors and the participants, and the presence of non participating agents (pets) in the home that affect the sensor inputs during an activity instance. The number of available instances per activity class is also considered as a factor in the selection of the dataset, since fewer instances make it challenging for the learning methods to be trained while large number of instances increases the computational cost of a system.

Kasteren and *Aruba1* datasets include activities performed by a single resident in a smart home, where 7 and 11 types of different activities are performed, respectively. Total number of instances in Kasteren are 245 and in Aruba1 are 6477. In *Kyoto1* and *Kyoto2* datasets, 5 types of activities are performed by 20 participants. Each activity is performed once by each participant one after the other. The number of activity instances are 120 in Kyoto1 and 100 in Kyoto2. Types of activities in Kyoto2 are same as in Kyoto1 while some erroneous features are added in the activity instances. *Kyoto8*, *Tulum1* and *Tulum2* contain activity instances performed by two residents living together and performing activities independently without cooperation. These are typically large datasets containing 11, 10 and 16 types of activities, respectively. *Milan* and *Cairo* datasets contain 15 and 13 types of activities performed by single and two participants, respectively, along with a pet in the house.

In the above datasets, the binary sensors used are motion sensors, contact switch sensors, absent/present status of item sensors and door open/close status of cabinet sensors. In addition, the analog sensors to measure the temperature and the status of water and burner are used.

4.3. ARSH-SV results and evaluation

Table 2 shows the number of assigned labels separated based on confidence scores using ARSH-SV. Labels with low and high confidence scores are assigned to the activity instances depending upon their distances from the predicted sub-clusters within an assigned class. ARSH-SV correctly assigns low confidence to more than 90% of the incorrect labels. A few of correct labels far from the centers of the sub-clusters within an activity class are also assigned low confidence, which is useful in such systems that require high accuracy in outputs for automated decisions, such as in medical diagnostic systems. 90% to 95% of the correctly assigned labels are given high confidence score. However, we can see that between 1 to 15% of the activity instances with incorrect labels are also assigned high confidence. The overall accuracy of the confidence score assignment

Table 2: Confidence measure for assignments by ARSH-SV for nine smart home datasets. Key: LC - Low Confidence, HC - High Confidence, TP - True Positive, FP - False Positive, TN - True Negative, FN - False Negative, Acc - Accuracy.

Datasets	Total Assignments	Labels with LC	LC to incorrect labels - TP	LC to correct labels - FP	Labels with HC	HC to correct labels -TN	HC to incorrect labels - FN	Acc (%) (Eq.24)
Aruba1	6477	448	301	147	6029	5650	379	91.88
Kasteren	245	16	10	6	229	221	8	94.30
Kyoto1	120	14	12	2	106	104	2	96.67
Kyoto2	100	14	11	3	86	84	2	94.98
Kyoto8	1290	83	52	31	1207	1174	33	95.04
Tulum1	1513	13	8	5	1500	1494	6	99.28
Tulum2	12637	493	481	12	12144	10250	1894	84.92
Milan	2310	536	534	2	1774	1665	109	95.20
Cairo	600	29	20	9	571	545	26	94.17

remains more than 90% for eight datasets, while one dataset have accuracy of 84%. In Tulum1 dataset, we achieve the highest accuracy of up to 99%, where almost all incorrect labels are assigned low and all the correct labels are assigned the high confidence score. This is because of the presence of well separated activity classes with high inter-class differences and low intra-class differences among the activity instances. In Tulum2, it can be observed that low confidence is correctly assigned to 98% of the incorrect labels. However, a large number of incorrect labels (1894) also get high confidence score because of the presence of very high inter-class similarities. Activities categorized in different classes share similar features because of the same location and same sensor activations, such as bathing and personal hygiene, enter home and leave home, and watch tv and work living room (see Table 1), which result in the least accuracy of 84% in Tulum2 dataset.

Table 3 shows the performance of ARSH-SV compared with the existing learning approaches; PCA-SVM [13], ET-KNN [35], PNN [15] and NB [30], using defined performance evaluation metrics applied on the results obtained from nine datasets using three-fold cross validation. ARSH-SV, shows the highest accuracy (using Eq. 24) in activity recognition for all datasets. After ARSH-SV, the next highest accuracies are achieved by ET-KNN in four, PCA-SVM in three and PNN in two out of nine datasets. The accuracy of PCA-SVM and ET-KNN remains comparable to each other, while PNN has comparatively less accuracy in recognition than ARSH-SV in the datasets with activities having similar features. One such example is of Tulum2 dataset, where PNN obtained the lowest accuracy of 90.70%, while in Kyoto1 and Kyoto2 with activities having enough discriminative information, PNN performs better compared to PCA-SVM and ET-KNN. NB classifier achieves the least accuracy in all datasets and especially in Kasteren dataset, it gets the the lowest accuracy of 74.70%, which is 24.36%, 23.32%, 23.67% and 19.01% less than ARSH-SV, PCA-SVM, ET-KNN and PNN, respectively.

In the next sub-sections, we perform a detailed comparison of the results presented in Table 3 by grouping the datasets with similar properties. We also discuss the confusion matrix of one of the datasets in each group. The extra column of *IA* (Irregular Activities) in a confusion matrix refers to the activity instances of a class that are recognized with less confidence and are not confused with other activity classes.

Single resident - Datasets

Single resident activities are performed in Kasteren and Aruba1 datasets. The number of instances in Kasteren are far less compared to Aruba1 dataset. Some of the activities such as meal preparation, dinner, drink and dish washing are performed in the same location and therefore share same sensors, which may result in less discriminative information, highlighting less inter-class variations.

In Kasteren dataset, ARSH-SV achieves precision, recall and specificity rates of 94.57%, 90.39% and 99.36%. Precision and recall of NB are up to 73% less than that of ARSH-SV, which shows a poor performance of NB in labeling the activity instances belonging to a class. In addition, the specificity reaches up to 13% less than ARSH-SV, which shows that the NB approach also do not label the negative instances correctly. MCC measure of ARSH-SV, PCA-SVM, ET-KNN, PNN and NB is 0.91, 0.85, 0.90, 0.64 and 0.08 respectively. In Aruba1 dataset, precision, recall and specificity of ARSH-SV are 90.12%, 91.71% and 99.40% respectively, which remain higher than PCA-SVM, ET-KNN, PNN and NB. Higher values of precision and recall show that ARSH-SV is better able to label the instances belonging to one activity class. Similarly, higher specificity shows that ARSH-SV is able to identify the instances not belonging to the target class. Finally, the higher MCC measure of 0.90 of ARSH-SV compared to PCA-SVM (0.78), ET-KNN (0.76), PNN (0.41) and NB (0.21) and the higher F1-score, confirms the correct label assignments with less false assignments in the case of less inter-class variations.

Table 4 shows the confusion matrix of activities in Aruba1 dataset. It can be observed that almost all the activities are recognized with high accuracy. The activity instances identified with low confidence are separated as irregular activities. In *House Keeping* activity, 69% of instances are identified with high confidence, while 24% are recognized with low confidence. It shows that 69 + 24 = 93% of HK instances are not confused with other classes, while indicating that 69% assignments are reliable. A few of the activities are transferring their errors to the other similar activities. An interesting example is that of *Leave Home* (LH) and *Enter Home* (EH). Since LH and

Table 3: Performance evaluation metrics on nine smart home datasets for ARSH-SV and the existing approaches [13], [35], [15] and [30] using three-fold cross validation. Precision, Recall, Specificity, FPR and Accuracy are in percentages (%), The range of F1-score is between [0 , 1], and range of MCC is between [−1 , 1]. Key: FPR - False Positive Rate, MCC - Matthews Correlation Coefficient. The highest values in the performance evaluation metrics are highlighted in bold.

S.no	Datasets	Classifiers	Precision(%)	Recall(%)	Specificity(%)	FPR(%)	MCC	F1-score	Accuracy(%)
1	Aruba1	ARSH-SV	**90.12**	**91.71**	**99.40**	**0.60**	**0.90**	**0.90**	**98.92**
		[13]	77.97	80.19	99.31	0.69	0.78	0.79	98.82
		[35]	80.31	77.53	99.07	0.93	0.76	0.75	98.39
		[15]	65.60	36.53	94.60	5.40	0.41	0.41	93.98
		[30]	27.08	30.76	92.50	7.50	0.21	0.22	86.81
2	Kasteren	ARSH-SV	**94.57**	90.39	**99.36**	**0.64**	**0.91**	**0.92**	**99.06**
		[13]	87.06	86.34	98.76	1.24	0.85	0.86	98.02
		[35]	92.07	**90.83**	98.98	1.02	0.90	0.90	98.37
		[15]	80.84	59.97	94.39	6.00	0.64	0.65	93.71
		[30]	16.38	16.67	85.84	14.00	0.08	0.10	74.70
3	Kyoto1	ARSH-SV	**98.43**	**98.33**	**99.58**	**0.42**	**0.98**	**0.98**	**99.33**
		[13]	94.87	94.17	98.54	1.46	0.93	0.94	97.67
		[35]	95.46	95.00	98.75	1.25	0.94	0.95	98.00
		[15]	96.30	95.83	98.96	1.04	0.95	0.96	98.33
		[30]	73.88	74.17	93.54	6.46	0.68	0.71	89.67
4	Kyoto2	ARSH-SV	**98.10**	**97.94**	**99.49**	**0.51**	**0.97**	**0.98**	**99.19**
		[13]	92.96	91.11	97.75	2.25	0.90	0.91	96.39
		[35]	90.06	90.00	97.50	2.50	0.87	0.90	96.00
		[15]	92.57	90.75	97.76	2.24	0.89	0.91	96.40
		[30]	60.83	60.00	90.02	9.98	0.51	0.55	83.96
5	Kyoto8	ARSH-SV	**94.57**	**97.27**	**99.76**	**0.24**	**0.95**	**0.95**	**99.53**
		[13]	57.44	63.29	96.78	3.22	0.56	0.58	94.81
		[35]	61.31	59.87	96.82	3.18	0.55	0.57	94.99
		[15]	44.75	40.97	95.94	4.06	0.38	0.41	91.03
		[30]	16.48	18.96	92.94	7.06	0.10	0.11	88.48
6	Tulum1	ARSH-SV	**95.91**	**96.31**	**99.96**	**0.04**	**0.96**	**0.96**	**99.92**
		[13]	67.28	62.79	97.68	2.32	0.62	0.64	96.21
		[35]	63.64	62.45	97.33	2.67	0.60	0.62	95.69
		[15]	54.19	36.99	93.64	6.36	0.37	0.38	91.45
		[30]	33.32	42.06	92.51	7.49	0.27	0.29	86.17
7	Tulum2	ARSH-SV	**72.07**	**70.74**	**99.01**	**0.99**	**0.70**	**0.70**	**98.12**
		[13]	65.51	61.08	98.09	1.91	0.60	0.60	96.06
		[35]	60.50	61.72	97.80	2.20	0.58	0.60	96.08
		[15]	53.80	23.93	94.43	5.57	0.28	0.26	90.70
		[30]	34.78	18.47	93.94	6.06	0.17	0.17	89.06
8	Milan	ARSH-SV	**87.31**	**83.55**	**99.64**	**0.36**	**0.85**	**0.85**	**99.36**
		[13]	69.50	69.86	98.78	1.22	0.68	0.69	97.84
		[35]	70.66	67.79	98.73	1.27	0.68	0.68	97.75
		[15]	57.95	45.30	97.89	2.11	0.48	0.48	96.29
		[30]	46.83	39.09	96.84	3.16	0.38	0.38	94.51
9	Cairo	ARSH-SV	**94.95**	**94.92**	**99.64**	**0.36**	**0.95**	**0.95**	**99.33**
		[13]	64.98	63.54	96.92	3.08	0.61	0.63	94.36
		[35]	66.85	65.67	97.16	2.84	0.63	0.65	94.79
		[15]	59.63	56.92	96.37	3.63	0.54	0.56	93.36
		[30]	41.53	42.27	95.67	4.33	0.37	0.39	92.18

EH are recognized by the single door sensor making them similar to each other, both LH and EH transfer 18% of their errors among each other. Similarly, *Wash Dishes* (WD) could be a part of *Meal Preparation* (MP) activity and can be carried out during MP, therefore WD is transferring 15% of errors to MP. For the remaining activity classes, ARSH-SV correctly assigns the low confidence to the incorrect instances instead of sharing them with similar activities.

Multi-participants - Datasets

Multiple participants perform individual activities without cooperation, in Kyoto1 and Kyoto2. Similarly to Kasteren dataset, the number of activity instances are very less. Variation in performing the same activity by multiple participants is observed, which results in intra-class and inter-subject variations. However, each activity class is discriminative enough and well separated resulting in high inter-class variation.

In Kyoto1, we achieve the highest precision, recall and specificity $(98.43\%, 98.33\%, 99.58\%)$ for ARSH-SV respectively. Precision and recall of NB is up to 24% and specificity is up to 6% less than ARSH-SV. The results show the higher ability of ARSH-SV in correct labeling of activity instances to the target class and the correct rejection of the activity instances of other classes. Similarly, MCC measure (0.98) remains highest for ARSH-SV compared to PCA-SVM (0.93), ET-KNN (0.94) PNN (0.95) and NB (0.68). It is to be noted that the perfor-

Table 4: *Aruba1* - confusion matrix using ARSH-SV. Rows represent the actual activities and columns represent the predicted activities. Key: Acts - Activities, Tlt - Bed to Toilet, Eat - Eating, EH - Enter Home, HK - House Keeping, LH - Leave Home, MP - Meal Preparation, Rlx - Relax, Res - Resperate, Slp - Sleeping, WD - Wash Dishes, Wk - Work, IA - Irregular Activities

Acts	Tlt	Eat	EH	HK	LH	MP	Rlx	Res	Slp	WD	Wk	IA
Tlt	98.7	0	0	0	0	0	0	0	0	0	0	1.3
Eat	0	90.3	0	0	0	0	1.9	0	0	0	1.2	6.6
EH	0	0	80.3	0	18.8	0	0	0	0	0	0	0.9
HK	0	3	0	69.7	0	0	3	0	0	0	0	24.3
LH	0	0	18.1	0	80	0	0	0	0	0	0	1.9
MP	0	0	0	0	0	81.6	0	0	0	6.7	0	11.7
Rlx	0	0	0	0	0	0	92.4	0	0	0	0	7.6
Res	0	0	0	0	0	0	0	83.3	0	0	0	16.7
Slp	0	0	0	0	0	0	5.5	0	82.3	0	2	10.2
WD	0	0	0	0	0	15.4	0	0	0	78.5	0	6.1
Wk	0	0	0	0	0	0	0	0	0	0	90.6	9.4

Table 5: *Kyoto2* - confusion matrix using ARSH-SV. Rows represent the actual activities and columns represent the predicted activities. Key: Acts - Activities, Cl - Cleaning, Ck - Cooking, Eat - Eating, Ph - Phone Call, WH - Wash Hands and IA - Irregular Activities.

Acts	Cl	Ck	Eat	Ph	WH	IA
Cl	90	0	5	0	0	5
Ck	0	60	0	0	0	40
Eat	0	0	85	0	0	15
Ph	0	0	0	95	0	5
WH	5	0	0	0	90	5

mance of ARSH-SV is not affected by the less amount of training data because of the use of support vectors in SVM, while the inclusion of confidence measure in the proposed approach improves the results by reducing the False Positives. In Kyoto2 dataset, some errors are added in the activity instances such as leaving the stove burner ON, or not using the water for dish cleaning after *Cooking*, leaving the water turned ON after *Wash Hands*, dialing a wrong number, and not bring the medicine container to dining room. ARSH-SV is able to accurately assign the correct labels to 'activity instances with errors'. Low confidence scores are assigned to 14 activity instances. Precision, recall, specificity and F1-score of ARSH-SV are 98.10%, 97.94%, 99.49% and 0.98, respectively. NB is not able to perform better in label assignment and shows nearly 37% less results in precision, recall and 9% less in specificity than ARSH-SV because of the less data available for training. Similarly, MCC values of ARSH-SV, PCA-SVM, ET-KNN, PNN and NB are 0.97, 0.90, 0.87, 0.89 and 0.51, respectively. ARSH-SV exploits multiple sub-clusters within an activity for class variations, which makes it more robust for intra-class and inter-subject variations.

Table 5, shows the confusion matrix of activities in Kyoto2. Most of the activities with errors are identified correctly, while 5% instances of both *Cleaning* and *Wash Hands* activities are recognized as *Eating* and *Cleaning*, respectively. The remaining instances of all the activity classes identified with low confidence are successfully represented as the irregular instances of the corresponding activity classes reducing the false positives.

Table 6: *Tulum1* - confusion matrix using ARSH-SV. Rows represent the actual activities and columns represent the predicted activities. Key: Acts - Activities, CBF - Cook Breakfast, CL - Cook Lunch, EH - Enter Home, GM - Group Meeting, LH - Leave Home, R1BF - R1 Eat Breakfast, R1Snk - R1 Snack, R2BF - R2 Eat Breakfast, WD - Wash Dishes, WT - Watch TV and IA - Irregular Activities.

Acts	CBF	CL	EH	GM	LH	R1BF	R1Snk	R2BF	WD	WT	IA
CBF	100	0	0	0	0	0	0	0	0	0	0
CL	0	98.6	0	0	0	1.4	0	0	0	0	0
EH	0	0	100	0	0	0	0	0	0	0	0
GM	0	0	0	45.5	0	0	0	0	0	0	54.5
LH	0	0	0	0	100	0	0	0	0	0	0
R1BF	0	0	0	0	0	98.5	0	0	0	0	1.5
R1Snk	0.4	0	0	0	0	0.4	98.8	0	0	0	0.4
R2BF	0	0	0	0	0	0	0	100	0	0	0
WD	1.4	0	0	0	0	0	0	0	97.2	0	1.4
WT	0	0	0	0	0	0	0	0	0	99.4	0.6

Multi-residents - Datasets

The activities in Kyoto8, Tulum1 and Tulum2 datasets are performed by two residents each. The residents share kitchen, living/dining and bathroom, while they live in two separate bedrooms. The activities carried out in shared locations are overlapping in features with less discriminative information. The variation within the same activity can also observed because of the activity instances of the same class performed by different residents. Datasets have both inter-class and intra-class variations.

In Kyoto8 dataset, 83 instances are assigned low confidence score from a total of 1290. Precision and recall of ARSH-SV are 94.57% and 97.27% respectively, which are 37.13% and 33.98% higher than PCA-SVM, 33.26% and 37.40% higher than ET-KNN, 49.82% and 56.30% higher than PNN and 78.09% and 78.31% higher than NB. ARSH-SV also has the highest MCC and F1-score of 0.95 each, while PCA-SVM, ET-KNN, PNN and NB remain less accurate in correct recognition of the activity instances. In Tulum1 dataset, a total of 1513 activity instances for ten classes are present. Precision, recall and specificity of ARSH-SV are 95.91%, 96.31% and 99.96% respectively, which is higher than those of PCA-SVM, ET-KNN, PNN and NB. Similarly, MCC and F1-score of ARSH-SV also remained higher at 0.96 each. Tulum2 dataset contains the highest number of activity instances (12637), where some of the activity classes are similar to each other, such as *bathing*, *bed to toilet* and *personal hygiene* are highly interrelated with each other (see Table 1). Precision, recall and specificity obtained by ARSH-SV are 72.07%, 70.74% and 99.01% and remained better than PCA-SVM and ET-KNN. The values for F1-score in case of ARSH-SV, PCA-SVM, ET-KNN, PNN and NB are 0.70, 0.60, 0.60, 0.26 and 0.17, respectively, and show the effectiveness of approach in the correct labeling of the activity instances to the target class in the case of both inter and intra-class variations.

Table 6 shows the confusion matrix of Tulum1 dataset. Other than *Group Meeting* GM, all the activities are correctly recognized with high confidence. In GM, 45.5% instances of GM are recognized with high confidence, while 54.5% of instances are identified with low confidence. Note that GM is not confused with other activities, however, a less confidence measure compared to other activity classes is due to the complex na-

Table 7: *Cairo* - confusion matrix using ARSH-SV. Rows represent the actual activities and columns represent the predicted activities. Key: Acts - Activities, Tlt - Bed to Toilet, BF - Breakfast, Dnr - Dinner, Ldy - Laundry, LH - Leave Home, Lch - Lunch, NW - Night Wandering, R1Slp - R1 Sleep, R1Wkp - R1 Wakeup, R1WO - R1 Work in Office, R2Slp - R2 Sleep, R2Md - R2 Take Medicine, R2Wkp - R2 Wakeup, and IA - Irregular Activities.

Acts	Tlt	BF	Dnr	Ldy	LH	Lch	NW	R1Slp	R1Wkp	R1WO	R2Slp	R2Md	R2Wkp	IA
Tlt	96.7	0	0	0	0	0	0	0	0	0	0	0	0	3.3
BF	0	89.6	2.1	0	0	0	0	0	0	0	0	0	0	8.3
Dnr	0	4.8	90.5	0	0	0	0	0	0	0	0	0	0	4.7
Ldy	0	0	0	70	10	0	0	0	0	0	0	0	0	20
LH	0	0	0	1.4	97.1	0	0	0	0	0	0	0	0	1.5
Lch	0	2.7	2.7	0	0	91.9	0	0	0	0	0	0	0	2.7
NW	0	0	0	0	0	0	97	0	0	0	0	0	1.5	1.5
R1Slp	0	0	0	0	0	0	0	92	0	0	4	0	0	4
R1Wkp	0	0	0	0	0	0	0	0	96.2	0	1.9	0	0	1.9
R1WO	0	0	0	0	0	0	0	0	0	89.1	2.2	0	0	8.7
R2Slp	5.8	0	0	0	0	0	0	1.9	0	0	80.8	0	3.8	7.7
R2Md	0	0	0	0	0	0	0	0	0	0	0	95.5	0	4.5
R2Wkp	11.5	0	0	0	0	0	0	0	0	0	0	0	76.9	11.6

ture of the activity, since multiple participants are involved and least number of activity instances (only 11) are available for the training that may not cover all the intra-class variations.

Residents with a pet - Datasets

In Milan dataset, one resident living with a pet, and in Cairo, two residents living with a pet, perform the activities. The activity classes in Cairo dataset such as *Breakfast*, *Lunch* and *Dinner*, or in Milan dataset such as *Morning Medication* and *Evening Medication*, have similar features. The presence of pet results in the activation of sensors that do not correspond to any activity and insert noise in the dataset. Additionally, the variation in the activity instances in Cairo dataset can be observed due to more number of users. Less inter-class while more intra-class and inter-subject variations result in more challenging datasets.

In Milan dataset, from the total of 2310 activity instances, precision, recall and specificity of ARSH-SV are 87.31%, 83.55% and 99.64% respectively, and higher than other approaches. The high recall shows that ARSH-SV is least affected by the noise in the data and it is able to correctly identify the activity instances. MCC measure of ARSH-SV is 0.85, which is also better compared to PCA-SVM, ET-KNN, PNN and NB. In Cairo dataset, total activity instances are 600, the precision, recall and specificity of ARSH-SV are 94.95%, 94.92% and 99.64%, respectively. MCC measure of ARSH-SV is also better than PCA-SVM, ET-KNN, PNN, and NB. F1-score of ARSH-SV, PCA-SVM, ET-KNN, PNN and NB is 0.95, 0.63, 0.65, 0.56 and 0.39, respectively. It can be observed that noise due to the presence of pet, overlap in features and variations in activity instances affect the performance of PCA-SVM, ET-KNN, PNN and NB more than ARSH-SV. ARSH-SV shows better results in the case of intra-class and inter-subject variations due to non-related sensor activations.

Table 7 shows the confusion matrix of activities in Cairo dataset. Despite the presence of 2 residents and a pet, it is observed that ARSH-SV has high recognition performance for activities other than *Laundry* (Ldy) and *R2 Wakeup* (R2Wkp).

Table 8: Comparison of existing approaches with ARSH-SV. Recognition rate is in percentage(%) and the range of F1-score is between [0, 1].

Evaluation measure	Cross validation	Dataset	Existing approaches	
Recognition rate	1 day out	Kyoto1	**ARSH-SV**	**93.00**
			ADR-SPLDA(SPL2) [21]	92.49
			ADR-SPLDA(SPL3) [21]	79.44
		Kasteren	**ARSH-SV**	94.24
			C-SVM [43]	93.40
			CRF [43]	**95.60**
			LDA [43]	93.50
		$Kasteren_{10}$	**ARSH-SV**	**88.55**
			AALO [29]	75.86
	three-fold	Cairo	**ARSH-SV**	**92.86**
			CRF [54]	91.00
			HMM [54]	82.00
F1-score	1 day out	Kyoto7	**ARSH-SV**	**0.91**
			AR-SPM(C+CP) [6]	0.73
		Kasteren	**ARSH-SV**	**0.91**
			AR-SPM(C+LP) [6]	0.81
			EFA-AR [45]	0.77
			Murphy rule [45]	0.68
			Dempster-Shafer rule [45]	0.54
			Temporal EDN [48]	0.70
			No time EDN [48]	0.45
		$Kasteren_{10}$	**ARSH-SV**	**0.80**
			HHMM [38]	0.78
			HSMM [38]	0.72
	ten-fold	Kasteren	**ARSH-SV**	**0.90**
			CL-AR [37]	0.87
			J48-DT [37]	0.79

In the case of Ldy, 70% of instances are recognized with high confidence and 20% of instances are recognized with low confidence (irregular). In R2Wkp, 76% of instances are identified correctly with high confidence and 11% with low confidence. The remaining 11% goes into *Bed to Toilet* (Tlt) activity, since Tlt is the next activity in the sequence of performance and shares overlapping features.

4.4. ARSH-SV comparison with existing approaches

Table 8 shows the performance comparison of ARSH-SV with the existing activity recognition approaches [6, 21, 28, 29,

Table 9: Confusion matrices on $Kasteren_{10}$ dataset for activity recognition using leave one day out cross validation (a) AALO [29] (b) ARSH-SV. The rows represent the actual activities and columns represent the predicted activities. Key: Acts - Activities, LH - Leave Home, Tlt - Toileting, Shr - Shower, Slp - Sleep, BF - Breakfast, Dnr - Dinner, Snk - Snack, Drk - Drink, WM - Washing Machine, DW - Dish Washer, IA - Irregular Activities.

(a) AALO [29]

Acts	LH	Tlt	Shr	Slp	BF	Dnr	Snk	Drk	WM	DW	IA
LH	89.9	0	0	0	0	0	0	0	0	0	10.1
Tlt	0	70.3	0	0	0	0	0	0	0	0	29.7
Shr	0	0	77.5	0	0	0	0	0	0	0	22.5
Slp	0	0	0	98.7	0	0	0	0	0	0	1.3
BF	0	0	0	0	64.5	0	4.2	6.3	0	5.8	19.2
Dnr	0	0	0	0	0	69.8	2.4	2.9	0.5	1.3	23.1
Snk	0	0	0	0	0	10	74.3	4.6	0	0	11.1
Drk	0	0	0	0	11.1	9.4	5.2	74.3	0	0	0
WM	0	0	0	0	0	10	0	0	70.4	0	19.6
DW	0	0	0	0	4.6	20.5	0	5.5	0	69.4	0

(b) ARSH-SV

Acts	LH	Tlt	Shr	Slp	BF	Dnr	Snk	Drk	WM	DW	IA
LH	100	0	0	0	0	0	0	0	0	0	0
Tlt	0	98.2	0.9	0	0	0	0	0	0	0	0.9
Shr	0	0	100	0	0	0	0	0	0	0	0
Slp	0	8.3	0	66.7	0	0	0	0	0	8.3	16.7
BF	0	0	0	0	75	0	10	5	10	0	0
Dnr	0	0	0	0	11.1	44.4	33.3	0	0	0	11.1
Snk	0	8.3	0	0	16.7	0	33.3	16.7	16.7	0	8.3
Drk	0	0	0	0	0	0	0	95	0	5	0
WM	0	0	0	0	0	0	0	0	88.9	0	11.1
DW	0	0	14.3	0	0	0	0	0	0	85.7	0

37, 38, 43, 45, 48, 54]. In addition to the already used three datasets: Kasteren, Cairo and Kyoto2, two more datasets: Kyoto7 (Daily life, Spring 2009) [55] and $Kasteren_{10}$ [11] comprising of 14 and 10 activities, respectively are used (activity descriptions in Fig. 2b and Table 9). We apply the cross validations same as mentioned in the compared approaches. We also use the evaluation measures: recognition rate and F1-score as applied in the state-of-the-art approaches. ARSH-SV shows the recognition rate better or comparable to the existing approaches in [21, 29, 43, 54] using four datasets: Cairo, Kyoto2, Kasteren and $Kasteren_{10}$. Since recognition rate does not consider false positives and false negatives, which are the main focus of ARSH-SV, the complete performance is not measurable through recognition rate alone. Also, since some activities are executed more frequently than others (imbalance), it is possible for a classifier to achieve a high recognition rate by assigning the correct label to the class with the majority of instances. Therefore, F1-score is applied, which incorporates both precision and recall and thus is more reliable. It can be observed that ARSH-SV has the highest F1-score in datasets, Kyoto7 (91%), Kasteren (91%) and $Kasteren_{10}$ (80%) compared to the existing approaches in [6, 37, 38, 45, 48]. The results show that in ARSH-SV, the activities are correctly recognized, while incorrect labels are correctly identified through confidence measure that remain useful in reducing the false positives effectively.

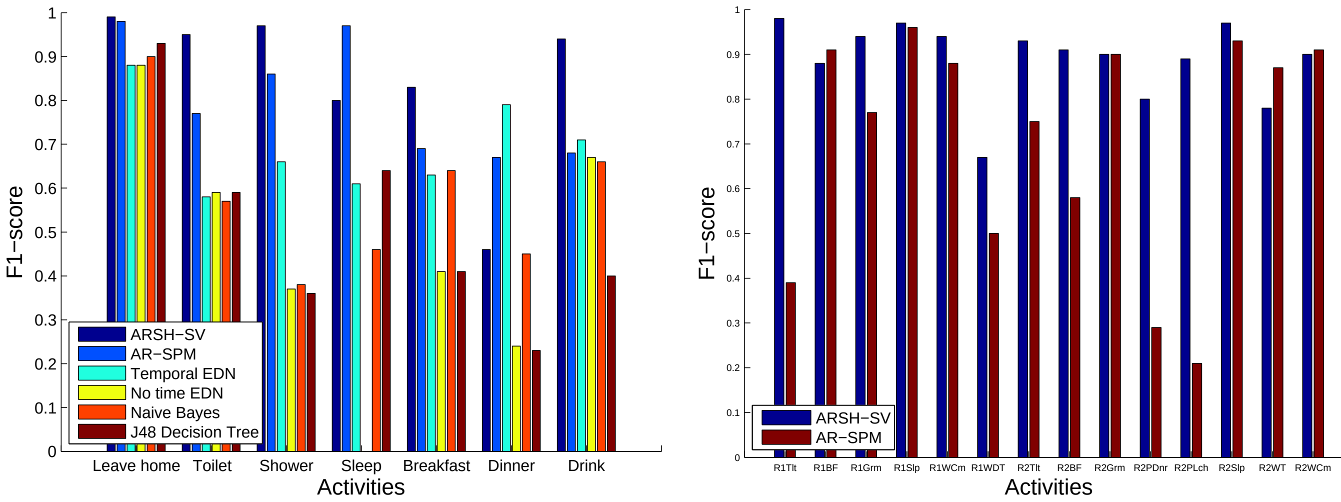
Table 9 shows the comparison of confusion matrices, for AALO [29] and ARSH-SV. AALO is a frequent itemset mining and clustering based activity recognition approach, which assigns labels to the activity instances and separates the instances not associated with any cluster. We can observe that ARSH-SV achieves improved performance than AALO in the activities of *Leave Home, Toilet, Shower, Breakfast, Drink, Washing Machine* and *Dish Washing*. In the case of eating activities (*Snack, Dinner* and *Breakfast*) because of the similarity of activity classes being performed in the same location, 33% of *Dinner* is recognized as *Snack* and 16% of *Snack* is recognized as *Breakfast*. It can be observed from the results that ARSH-SV achieves an overall better recognition performance and correctly assigned the low confidence to the irregular instances, which could otherwise be confused with other activity classes.

Figure 2 shows the activity level comparison of ARSH-SV with AR-SPM [6], Temporal EDN, No time EDN, NB and J48-DT [48] for Kasteren dataset (Fig. 2a), and with AR-SPM [6] for Kyoto7 dataset (Fig. 2b) using F1-score. We apply leave one day out cross validation, a stringent test using single day data for testing and the remaining for training and the process is repeated for all days. In Kasteren dataset, ARSH-SV attains high F1-score in all the seven activities compared to no time EDN, NB and J48-DT and for five activities compared to AR-SPM, while it shows comparatively less F1-score in the *dinner* activity compared to temporal EDN and AR-SPM, due to the similarity with *breakfast* and the availability of less (9) instances for training. In Kyoto7, ARSH-SV achieves better F1-score in most of the activities while comparable to AR-SPM in a few.

ARSH-SV shows its effectiveness in the recognitions in the case of inter-class similarities and intra-class variations. The confidence measure is useful in highlighting the anomalous and incorrectly assigned activity instances and improves the reliability and performance of the activity recognition system.

5. Conclusions

We proposed an activity recognition approach that improves the reliability of the recognized activities by measuring the confidence of the assigned labels. Binary SVM is used for the classification of pre-segmented activity instances, while the underlying distribution of data within an activity class is estimated through sub-clustering. Sub-clusters represent the intra-class variations among the activity instances of the same class, which are exploited to measure the confidence score of the assigned labels. This proves to be useful in reducing the number of false positives in the assignments resulting in more reliability in the recognition performance. The performance of ARSH-SV is evaluated on nine real smart home datasets using a comprehensive evaluation metrics. ARSH-SV correctly measures the confidence score on average up to 94% of the assigned labels and shows a better performance by achieving an average accuracy of 98.85% on nine datasets, which is 6% higher compared to the existing learning approaches for activity recognition.



(a) Kasteren dataset

(b) Kyoto7 dataset

Figure 2: Activity level performance comparison of ARSH-SV with existing methods through F1-score using 'leave one day' out cross validation: (a) AR-SPM (C+LP) [6] Temporal EDN, No time EDN, NB and J48-DT [48] on Kasteren, and (b) AR-SPM (C+CP) [6] on Kyoto7. Key: R1Tlt - R1 Bed to Toilet, R1BF - R1 Breakfast, R1Grm - R1 Groom, R1Slp - R1 Sleep, R1WCm - R1 Work at Computer, R1WDT - R1 Work at Dining room Table, R2Tlt - R2 Bed to Toilet, R2BF - R2 Breakfast, R2Grm - R2 Groom, R2PDnr - R2 Prepare Dinner, R2PLch - R2 Prepare Lunch, R2Slp - R2 Sleep, R2WT - R2 Watch TV, R2WCm - R2 Work at Computer.

References

[1] H. Hachiya, M. Sugiyama, N. Ueda, Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition, Neurocomputing 80 (0) (2012) 93–101.

[2] V. Rialle, F. Duchene, N. Noury, L. Bajolle, J. Demongeot, Health smart home: Information technology for patients at home, Telemedicine Journal and e-Health 8 (4) (2002) 395–409.

[3] B. Taati, J. Snoek, A. Mihailidis, Video analysis for identifying human operation difficulties and faucet usability assessment, Neurocomputing 100 (0) (2013) 163–169.

[4] M. Chana, D. Estevea, C. Escriba, E. Campo, A review of smart homes-present state and future challenges, Computer Methods and Programs in Biomedicine 91 (1) (2008) 55–81.

[5] F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou, Y. Amirat, Joint segmentation of multivariate time series with hidden process regression for human activity recognition, Neurocomputing 120 (0) (2013) 633–644.

[6] U. Avci, A. Passerini, Improving activity recognitionby segmental pattern mining, IEEE Trans. on Knowledge and Data Engineering 26 (4) (2014) 889–902.

[7] G. Okeyo, L. Chen, H. Wang, R. Sterritt, Dynamic sensor data segmentation for real-time knowledge-driven activity recognition, Pervasive and Mobile Computing 10 (0) (2014) 155–172.

[8] T. Gu, S. Chen, X. Tao, J. Lu, An unsupervised approach to activity recognition and segmentation based on object-use fingerprints, Data and Knowledge Engineering 69 (6) (2010) 533–544.

[9] T. Gu, Z. Wu, L. Wang, X. Tao, J. Lu, Mining emerging patterns for recognizing activities of multiple users in pervasive computing, in: Proc. of Int. Conf. on Mobile and Ubiquitous Systems: Networking Services, Toronto, Canada, 2009, pp. 1–10.

[10] P. Rashidi, D. J. Cook, L. B. Holder, M. Schmitter-Edgecombe, Discovering activities to recognize and track in a smart environment, IEEE Trans. on Knowledge and Data Engineering 23 (4) (2011) 527–539.

[11] T. V. Kasteren, A. Noulas, G. Englebienne, B. Krose, Accurate activity recognition in a home setting, in: Proc. of Int. conf. on Ubiquitous computing, Seoul, Korea, 2008, pp. 1–9.

[12] E. M. Tapia, S. S. Intille, K. Larson, Activity recognition in the home using simple and ubiquitous sensors, Pervasive Computing 3001 (2004) 158–175.

[13] A. Fleury, M. Vacher, N. Noury, Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results, IEEE Trans. on Information Technology in Biomedicine 14 (2) (2010) 274–283.

[14] J.-Y. Yang, J.-S. Wang, Y.-P. Chen, Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers, Pattern Recognition Letters 29 (16) (2008) 2213 – 2220.

[15] L. G. Fahad, A. Ali, M. Rajarajan, Long term analysis of daily activities in smart home, in: Proc. of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 2013, pp. 419–424.

[16] L. Bao, S. S. Intille, Activity recognition from user-annotated acceleration data, in: Proc. of Int. conf. on Pervasive computing, Vienna, Austria, 2004, pp. 1–17.

[17] X. Hong, C. Nugent, M. Mulvenna, S. McClean, B. Scotney, Evidential fusion of sensor data for activity recognition in smart homes, Pervasive and Mobile Computing 5 (3) (2009) 236–252.

[18] D. J. Cook, M. Schmitter-Edgecombe, Assessing the quality of activities in a smart environment, Methods of Information in Medicine 48 (5) (2009) 480–485.

[19] E. Kim, S. Helal, D. Cook, Human activity recognition and pattern discovery, IEEE Pervasive Computing 9 (1) (2010) 48–53.

[20] G. Iannello, G. Percannella, C. Sansone, P. Soda, On the use of classification reliability for improving performance of the one-per-class decomposition method, Data and Knowledge Engineering 68 (12) (2009) 1398–1410.

[21] B. Chikhaoui, S. Wang, H. Pigot, Activity discovery and recognition by combining sequential patterns and latent dirichlet allocation, Pervasive and Mobile Computing 8 (6) (2012) 845–862.

[22] L. Atallah, G.-Z. Yang, The use of pervasive sensing for behaviour profiling - a survey., Pervasive and Mobile Computing 5 (5) (2009) 447–464.

[23] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, I. Korhonen, Activity classification using realistic data from wearable sensors, IEEE Trans. on Information Technology in Biomedicine 10 (1) (2006) 119–128.

[24] M. Ermes, J. Parkka, J. Mantyjarvi, I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, IEEE Trans. on Information Technology in Biomedicine 12 (1) (2011) 20–26.

[25] Z. Wang, M. Jiang, Y. Hu, H. Li, An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors, IEEE Trans. on Information Technology in Biomedicine 16 (4) (2012) 691–699.

[26] S. Wang, J. Yang, N. Chen, X. Chen, Q. Zhang, Human activity recognition with user-free accelerometers in the sensor networks, in: Proc. of IEEE Int. Conf. on Neural Networks and Brain, Vol. 2, Beijing, China, 2005, pp. 1212 –1217.
[27] A. Bulling, J. A. Ward, H. Gellersen, G. Troster, Eye movement analysis for activity recognition using electrooculography, IEEE Trans. on Pattern Analysis and Machine Intelligence 33 (4) (2011) 741–753.
[28] S. Mckeever, J. Ye, L. Coyle, C. Bleakley, S. Dobson, Activity recognitio6n using temporal evidence theory, Journal of Ambient Intelligence and Smart Environments 2 (3) (2010) 253–269.
[29] E. Hoque, J. Stankovic, Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities, in: Proc. of IEEE Int. Conf. on Pervasive Computing Technologies for Healthcare, 2012, pp. 139–146.
[30] D. J. Cook, Learning setting-generalized activity models for smart spaces, IEEE Intelligent Systems 27 (1) (2012) 32–38.
[31] C. Chen, B. Das, D. J. Cook, A data mining framework for activity recognition in smart environments, in: Proc. of IEEE Int. Conf. on Intelligent Environments, Kuala Lumpur, Malaysia, 2010, pp. 80–83.
[32] T. V. Duong, H. H. Bui, D. Q. Phung, S. Venkatesh, Activity recognition and abnormality detection with the switching hidden semi-markov model, in: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, Vol. 1, 2005, pp. 838–845.
[33] D. F. Specht, Probabilistic neural networks, Neural Networks 3 (1) (1990) 109–118.
[34] L. G. Fahad, S. F. Tahir, M. Rajarajan, Activity recognition in smart homes using clustering based classification, in: Proc. of IEEE Int. Conf. on Pattern Recognition, 2014.
[35] T. Denoeux, A k-nearest neighbor classification rule based on dempster-shafer theory., IEEE Trans. on Systems, Man, and Cybernetics 25 (5) (1995) 804–813.
[36] L. M. Zouhal, T. Denoeux, An evidence-theoretic k-nn rule with parameter optimization., IEEE Trans. on Systems, Man, and Cybernetics, Part C 28 (2) (1998) 263–271.
[37] J. Ye, S. Dobson, Exploring semantics in activity recognition using context lattices, Journal of Ambient Intelligence and Smart Environments 2 (4) (2010) 389–407.
[38] G. E. Tim van Kasteren, B. Krose, Hierarchical activity recognition using automatically clustered actions, in: Proc. of Int. Conf. on Ambient Intelligence, 2011, pp. 82–91.
[39] T. Gu, L. Wang, Z. Wu, X. Tao, J. Lu, A pattern mining approach to sensor-based human activity recognition, IEEE Trans. on Knowledge and Data Engineering 23 (9) (2011) 1359–1372.
[40] T. Gu, L. Wang, H. Chen, X. Tao, J. Lu, Recognizing multiuser activities using wireless body sensor networks, IEEE Trans. on Mobile Computing 10 (11) (2011) 1618–1631.
[41] P.-C. Huang, S.-S. Lee, Y.-H. Kuo, K.-R. Lee, A flexible sequence alignment approach on pattern mining and matching for human activity recognition, Expert Systems with Applications 37 (1) (2010) 298–306.
[42] S. Luhr, G. West, S. Venkatesh, Recognition of emergent human behaviour in a smart home: A data mining approach, Pervasive and Mobile Computing 3 (2) (2007) 95–116.
[43] M. Abidine, B. Fergani, Evaluating c-svm, crf and lda classification for daily activity recognition, in: Proc. of IEEE Int. Conf. on Multimedia Computing and Systems, 2012, pp. 272–277.
[44] M. Zhang, A. A. Sawchuk, A feature selection-based framework for human activity recognition using wearable multimodal sensors, in: Proc. of Int. Conf. on Body Area Networks, Beijing, China, 2011, pp. 92–98.
[45] F. Sebbak, A. Chibani, Y. Amirat, A. Mokhtari, F. Benhammadi, An evidential fusion approach for activity recognition in ambient intelligence environments, Robotics and Autonomous Systems 61 (11) (2013) 1235–1245.
[46] J. Liao, Y. Bi, C. D. Nugent, Using the dempster-shafer theory of evidence with a revised lattice structure for activity recognition, IEEE Trans. on Information Technology in Biomedicine 15 (1) (2011) 74–82.
[47] A. Tolstikov, X. Hong, J. Biswas, C. Nugent, L. Chen, G. Parente, Comparison of fusion methods based on dst and dbn in human activity recognition, Journal of Control Theory and Applications 9 (1) (2011) 18–27.
[48] S. McKeever, Recognising Situations Using Extended Dempster-Shafer Theory, Ph. D. Thesis, University College Dublin, 2011.
[49] S. Lloyd, Least squares quantization in pcm, IEEE Trans. on Information Theory 28 (2) (1982) 129–137.
[50] V. N. Vapnik, Statistical learning theory, John Wiley and Sons, New York, USA, 1998.
[51] A. Idris, A. Khan, Y. S. Lee, Intelligent churn prediction in telecom: employing mrmr feature selection and rotboost based ensemble classification., Applied Intelligence 39 (3) (2013) 659–672.
[52] M. Tahir, A. Khan, A. Majid, Protein subcellular localization of fluorescence imagery using spatial and transform domain features, Bioinformatics 28 (1) (2012) 91–97.
[53] A. Zoubir, B.Boashash, The bootstrap and its application in signal processing, IEEE Signal Processing Magazine 15 (1) (1998) 56–76.
[54] E. Nazerfard, B. Das, L. B. Holder, D. J. Cook, Conditional random fields for activity recognition in smart environments, in: Proc. of ACM Int. Symp. on Health Informatics, 2010, pp. 282–286.
[55] CASAS, Smart Home Projects, Washington State University (USA), Washington State University, 2013.
URL `http://ailab.wsu.edu/casas/datasets/ (last accessed: Jun, 2013)`