IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 25, 2017, accepted November 22, 2017, date of publication December 4, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2779158

An Annotation Technique for In-Home Smart Monitoring Environments

MOHAMMED GH. AL ZAMIL[1], MAJDI RAWASHDEH[2], SAMER SAMARAH[1], M. SHAMIM HOSSAIN[3], (Senior Member IEEE), AWNY ALNUSAIR[4], AND SK MD MIZANUR RAHMAN[5], (Member IEEE)

[1]Computer Information Systems Department, Yarmouk University, Irbid 21163, Jordan
[2]Management Information System Department, Princess Sumaya University for Technology, Amman 11941, Jordan
[3]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[4]Department of Informatics and Computer Science, Indiana University Kokomo, Kokomo, IN 46902, USA
[5]Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: M. Shamim Hossain (mshossain@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, through the research group, under Project RGP-228.

ABSTRACT Advances in multimedia technologies have led to the emergence of smart home applications. In fact, mobile multimedia technologies provide the infrastructure to adopt smart solutions and track inhabitants' activities. In-home activity recognition significantly enhances the performance of healthcare-monitoring and emergency-control applications for elderly and people with special needs. Developing and validating data models for such applications requires training sets that reflect a ground truth in the form of labeled or annotated data. With the accelerated development of Internet-of-Things applications, automated annotation processes have emerged understanding resident behavior in terms of activities. This paper presents a methodology for automatic data annotation by profiling sensing nodes. Our proposed methodology models activities based on spatially recognized actions, with every activity expected to have a direct relationship with a specific set of locations. Furthermore, the proposed technique validates the assignment of labels based on the temporal relations among consecutive actions. We performed experiments to evaluate our proposed methodology on CASAS data sets, which indicated that the proposed methodology achieved better performance, to a statistically significant extent, than the state-of-the-art methodologies presented in the literature.

INDEX TERMS IoT, mobile multimedia, mobile healthcare, data mining, in home activities, wireless sensors.

I. INTRODUCTION

Recent advances in wireless sensor networks and the efficient connectivity enabled by Internet-of-Things (IoT) infrastructure facilitate accessing different forms of multimedia content via mobile devices. Real-time remote access to such content allows the development of advanced applications and services for tracking patients in their homes. Applications of In-Home Activity Recognition (IHAR) are important for developing smart healthcare systems and services, such as monitoring elderly health, detecting communicable disease, and transmitting medically urgent alarms [1].

Recently, the Internet-of-Things (IoT) architecture has been deployed to most in-home technologies, facilitating the interconnection of ubiquitous devices embedded in home appliances in the form of sensors to acquire residents' data [2]. Such architecture simplifies the process of collecting data but makes it difficult to interpret incoming information for the purpose of offering advice or recommendations.

Figure 1 shows a framework for the smart recognition of in-home activities for the purpose of monitoring elderly health. IoT sensors are embedded in home appliances to sense data that, in their basic form, are huge and unexplained, representing only user actions. Since an activity is represented as a set of cohesive actions, such actions must be formulated, modeled, and annotated to reflect a specific activity. An ultimate use for the output of this formulation and annotation is a training model for a machine-learning algorithm to recognize incoming activities. In other words, we need to identify how the recognition engine will detect an activity, since actions arrive rapidly.

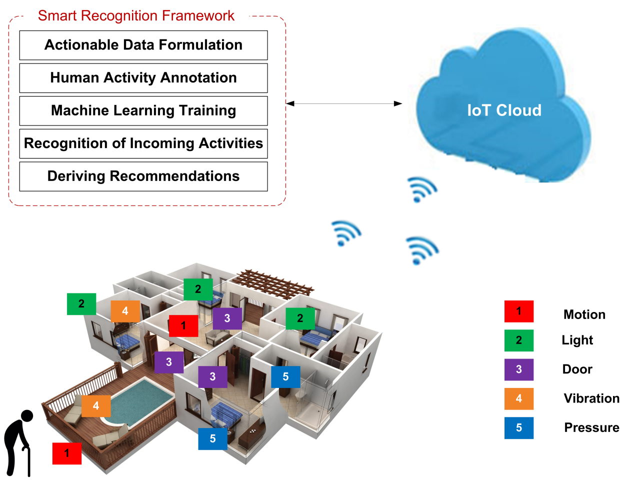


FIGURE 1. IoT smart recognition framework for healthcare monitoring systems.

One major challenge in implementing this smart recognition framework is to assemble residents' actions into data segments, which is a collection of actions that, collectively, represents a single activity (Figure 2). The smart home sensors report actions periodically, while the recognition engine assembles cohesive actions into meaningful segments (i.e., activities). Activity annotation is the process of detecting the appropriate label for a set of actions (segment) that home residents actually perform. Since most at-home activities involve different actions and since every sensing device is responsible for only one kind of action, the temporal and spatial aspects of human action matter for detecting activities—including the order of actions.

Date	Time	Sensor	Value	Label	
02-27	1:49:52	M14	ON	Kitchen_activ_Begin	
02-27	1:49:53	M15	ON		
02-27	1:49:54	M16	ON		
02-27	1:49:55	M17	ON		Segment
02-27	1:49:55	M15	OFF		
02-27	1:49:55	M14	OFF		
02-27	1:49:57	M16	OFF		
02-27	1:49:57	M13	OFF	Kitchen_activ_End	
02-27	1:49:58	AD1-B	0.03	Watch_TV_Begin	
02-27	1:49:59	M17	OFF		
02-27	1:50:01	AD1-B	0.30		
02-27	1:50:10	AD1-B	0.47		Segment
02-27	1:50:25	AD1-B	0.53		
02-27	1:50:38	M17	ON		
02-27	1:50:40	AD1-B	0.46		
02-27	1:50:42	M17	OFF	Watch_TV_End	

FIGURE 2. Data segments.

Previous research has handled this issue by using classical classification techniques [3]–[5] that rely on inducing features from a training set of data into a feature vector for input to different classification algorithms. The disorganized nature of inhabitants' home behavior makes it difficult to overcome the ambiguity problem using these techniques [6]. However, no previous research has considered automatically segmenting data during the process of data acquisition.

Another important challenge in annotating incoming actions is the interleaving among activities, meaning that humans may perform two or more actions concurrently. Since no specific rules indicate the chronological order of actions that represent activities, and since human activities are typically performed in different ways, activity detection often has high ambiguity and degraded performance in terms of accuracy [7].

This paper introduces an annotation technique for in-home activity recognition based on automatic segmentation. The proposed technique defines the annotation process as an optimization problem in which each incoming action is modeled to increase the probability of assigning a given set of actions to a specific activity. Hidden Markov Model (HMM) and Conditional Random Field (CRF) are applied to model the joint probability and features of activities in terms of actions. The proposed feature-generation model handles common challenges, such as actions' spatiotemporal features, ambiguity in detecting activities, and interleaving among home activities.

This research contributes by: (1) modeling activity actions as a set of states and transitions using HMM, (2) modeling a transition feature function that embeds temporal and spatial relations among consecutive actions, and (3) defining the segmentation problem as an optimization problem to minimize the impact of ambiguity on overall accuracy.

Given a set of activities $A = \{a_1, a_2, \ldots, a_M\}$, where each activity a_i is a sequence of atomic actions each of which is detected by a specific sensor. $a_{i,j} = <t_i^+, s_\alpha, v_\beta, t_i^->$ in which a_{ij} denotes the j^{th} action belonging to activity a_i, t_i^+ and t_i^- denote the start and end times of an activity a_i and s_α denotes the sensor that reports the atomic action v_β. Our goal, in this paper, is to define the probability function P as a confidence score to maximize the probability that a given atomic action increase the score of a given segment.

$$\mathrm{Max}[P(a_i, t, r, l) = \prod_{i=1}^{M} P(a_i|t, r, l)] \tag{1}$$

The remaining of this paper is organized as follows. Section II summarizes the related work, and Section III presents our proposed model. Section IV introduces the overall methodology. Section V presents the experimental results, and Section VI concludes.

II. LITERATURE REVIEW

Previous research regarding multivariate time-series data has focused on fully supervised learning approaches in which the training datasets are correctly annotated with labels that point to specific sets of activities. However, such approaches are appropriate for applications, such as intrusion and false alarm detection [8], [9], medical disease recognition [10], and monitoring of human health [11], that allow a classification problem to be defined to detect specific and homogeneous types of activities.

Recently, advances in smart and IoT technologies have led to more general models that can be utilized to automate the process of detecting in-home human activities. Many techniques, frameworks, and algorithms have been proposed to handle different issues in this domain. This paper focuses only on data segmentation, in which an agent must decide the size of the block of actions that represents an activity.

The classification problem is, by definition, a supervised learning task where training datasets are already labeled. Since smart infrastructure perceives the state of residents and their physical environment using sensors, feature enrichment is crucial for developing high-performance classifiers in terms of accuracy [12]. Extracting features automatically is a challenge, too [13], since sensors collect a very small amount of information. For this reason, automatic data segmentation is a challenging problem that requires uncommon techniques to solve.

Manual annotation was once the only way to label datasets for the purpose of training activity models [14], [15]. In this technique, a group of participants is asked to note every activity they perform. In other cases, the experimenters have guided participants toward the exact order in which the activities should be performed, so that the right activity labels are known before the sensor reports its data [16], [17].

In the literature, the data segmentation problem has been resolved using the sliding window approach introduced by Dietterich [18]. The idea behind the sliding window approach is to pick up a fixed number of sensors every time and then move the beginning of the window toward the second entry (and then the third, and so on). Every window represents a sequence of consecutive actions. The size of the sliding window must be chosen in advance, which decreases the accuracy of the approach, even if segments also have fixed size.

Dynamic size windowing is an interesting approach to overcoming the problems with the fixed-size sliding window [19]. This approach relies on making decisions about window size according to certain features. Such an approach works well for datasets that are collected perfectly, with no noisy tuples and where every tuple is annotated with an activity label. However, sensors use wireless communication and activity labels cannot cover every possible action (some actions do not belong to any activity).

Hidden Markov Models (HMM) have been applied to statistically model human behavior for the purpose of activity recognition. Examples of such techniques are in [20]–[22], where the problem is depicted as a set of states and transitions among them. Every state represents a human behavior, while each of them is connected to a specific observation object so that the physical environment is also embedded. However, HMM cannot model interleaved activities.

For this reason, a Conditional Random Field (CRF) is used to model concurrency among activities. CRF allows the statistical model to include feature function [23], [24], helping to create features that recognize both activities and concurrency among them.

Inducing features from training datasets is also a challenging problem, because sensors collect little data. Integrating temporal and spatial features with activities seems to be a solution to this problem [25]–[29]. While these are considered features of binary sensors, other research has focused on multimedia features [30], [31].

III. PROBABILISTIC MODELING OF ACTIVITIES

Consider a dataset D of N tuples that represent M activities in a smart home environment in which $M = |\mathrm{A}|$ and $\mathrm{A} = \{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{M}\}$ is a set of independent activities. Let v_t be an action that happened at time t. Further, we assume that the smart home environment comprises a finite number of sensors, with each sensor associated with exactly one action. Thus, the number of activities and actions in the environment is finite. We define a probabilistic finite state automaton that maps each action to a specific activity (state).

A. HIDDEN MARKOV MODEL (HMM)

The Hidden Markov Model is a generative probabilistic model since it generates hidden states from data observations. Specifically, the goal of HMM is to determine the sequence of actions $\mathrm{V_i} = \{\mathrm{v}_1, \mathrm{v}_2, \ldots \mathrm{v_t}\}$ that strongly correspond to observable outputs from specific sequence of sensors $\mathrm{S_i} = \{\mathrm{s}_1, \mathrm{s}_2, \ldots \mathrm{s_t}\}$. Figure 3 shows an example of HMM states and observation sequence of the activity "Bathing."

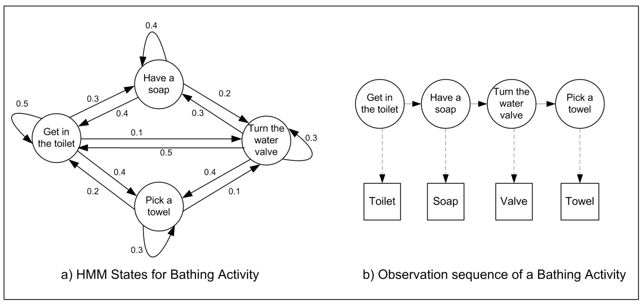


FIGURE 3. Sample HMM bathing activity.

The example in figure 3 explains the modeling of the activity entitled "Bathing." Through scanning the historical data (Training set), a state transition structure is built according to the actions that formulate the activity and the order to executing such actions. Since the order is not consistent, actions may be performed in different orders, the transition probability reflects the possibility of one action to be followed by another one. On the other hand, another structure is formulated; observation sequence. The observation sequence connects activities with the sensors that detect its constituent actions. Specifically, the state transition modeled actions with respect to a specific activity while observation sequence models activities with respect to the sensors that detect its actions.

The first order HMM defines the next state (future one) according to the current state only; not previous history. In other words, at time (t) the action $\mathrm{v_t}$ depends only on $\mathrm{v_{t-1}}$.

$$\mathrm{P}\,(\mathrm{v_t} \,|\, \mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v_{t-1}}) = \mathrm{P}(\mathrm{v_t}|\mathrm{v_{t-1}}) \qquad (2)$$

Observation parameters (S), on the other hand, are only dependent on the current hidden state. At time (t), given an observation parameter s_t, it depends only on hidden state v_t. Such assumption prevents a single action from being shared by two activities simultaneously. The following formula defines the probability of observing s_t while the hidden state v_t is independent from all other actions:

$$P\left(s_t \mid v_t, s_1, s_2, \ldots, s_{t-1}, v_t, \ldots, v_{t-1}\right) = P(s_t|v_t) \quad (3)$$

To map the transition among states in the finite state machine, we must connect an observed output with the most probable hidden state sequence. The transition probability is depicted as $P(v_{t-1}|v_t)$, while observation probability is $P(s_t|v_t)$ that means the probability of s_t observed in hidden state v_t. To maximize the joint probability:

$$P(s, v) = \prod_{t=1}^{T} P(v_t \mid v_{t-1}) P(s_t|v_t) \quad (4)$$

B. CONCURRENT ACTIVITIES

Simple HMMs work well with simple activities and actions that do not interleave in their execution. However, it is normal in smart-home environments for an inhabitant to perform more than one activity at the same time. Therefore, the learning algorithm must be fed with extra information about the nature of incoming activities.

One solution is to apply a conditional random field (CRF) model to define a feature that facilitates detecting such situations, while HMM defines the joint model. This allows the definition of non-independent relationships among observed sequences. In other words, we can embed the historical information that is required to deeply understand the relationships among activities.

$$P(V \mid S) = \frac{1}{Norm(S)} \exp\left(\sum_{i}^{N} \sum_{t=1}^{T} \varphi_i f_i (v_{t-1}, v_t, S, t)\right) \quad (5)$$

Norm(S) : is a normalization factor to make the probability value between 0 and 1.

φ_i is the transition probability (weight)

$f_i (v_{t-1}, v_t, S, t)$ represents the transition feature function, the state feature function, or a combination between them.

In fact, one contribution in this paper is to focus on modeling the function f_i in order to maximize the accuracy of segmenting incoming actions into a cohesive sequence that represents an activity in a specific smart home environment.

C. FEATURE FUNCTIONS

The transition feature function in equation (5) can model generative features from the datasets. Features, in this context, enrich the model with extra information to bias the results toward a specific label (i.e., state). Temporal and spatial features are two common approaches to modeling the function f_i. Therefore, we represent the function as the product of the values from these features.

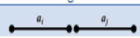
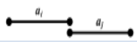
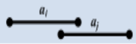
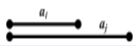
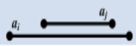
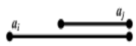
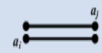
$$f_i (v_{t-1}, v_t, S, t) = \prod \omega_I \varphi_{loc}^t \quad (6)$$

The parameter ω_I models the temporal feature among consecutive actions. It is the ratio of the appearance frequency of two consecutive actions with respect to a specific relation. Equation (7) explains the formal definition of parameter ω_I, where r is the temporal relation between action v_{t-1} and v_t.

$$\omega_I = \frac{Freq([v_{t-1}, r, v_t])}{\text{Number of tuples of } (r, S)} \quad (7)$$

To restrict the value of r, we used the Allen's relationships [30] that depict the temporal relations among two actions (or actions). Table 1 shows 13 temporal relationships that could be, possibly, exist between two actions.

TABLE 1. Allen's relations.

Relation	Abbreviation	Diagram	Reverse	Abbreviation
a_i (before) a_j	a_i (b) a_j		a_i (after) a_j	a_i (b^-) a_j
a_i (meets) a_j	a_i (m) a_j		a_i(met − by)a_j	a_i(m^-)a_j
a_i (overlaps) a_j	a_i (o) a_j		a_i (overlap − by) a_j	a_i (o^-) a_j
a_i(starts)a_j	a_i(s)a_j		a_i(started − by)a_j	a_i(s^-)a_j
a_i(contains)a_j	a_i(c)a_j		a_i(during)a_j	a_i(c^-)a_j
a_i(finished − by)a_j	a_i(f)a_j		a_i(finishes)a_j	a_i(f^-)a_j
a_i(parallel)a_j	a_i(p)a_j		=	=

The second parameter φ_{loc}^t models the spatial aspects of actions, in which the location of the sensor, that detects such actions, maps between them. The parameter φ_{loc}^t is computed using a binary transition function at a specific time. The time variable (t) is important for mobile sensors in which the location is changing over time.

$$\varphi_{loc}^t = \left\{ \begin{array}{l|l} 0 & Loc(v_{t-1}) \neq Loc(v_t) \\ 1 & Loc(v_{t-1}) = Loc(v_t) \end{array} \right\} \quad (8)$$

This parameter plays a significant role in minimizing ambiguity. Consider a situation in which the resident is cooking a meal in the kitchen. During this activity, the resident goes to the living room and then returns to the kitchen. The resident went to the living room and get back again to the kitchen. This interruption will add an action from a motion sensor that is not relevant to the kitchen activity. Comparing the locations of the motion sensor in the living room and all sensors in the kitchen will allow the learning algorithm to recognize this fact.

IV. RESEARCH METHODOLOGY

Data pre-processing, a trivial task in data-mining techniques, involves cleaning up noise from the data, transforming data into an applicable format, normalizing data into a canonical form, and extracting features. Our focus, in this research, is confined to the non-trivial task of segmenting data and extracting features from segments. Figure 4 depicts our methodology to implement, test, and evaluate the proposed framework. The first step after preparing the datasets is to

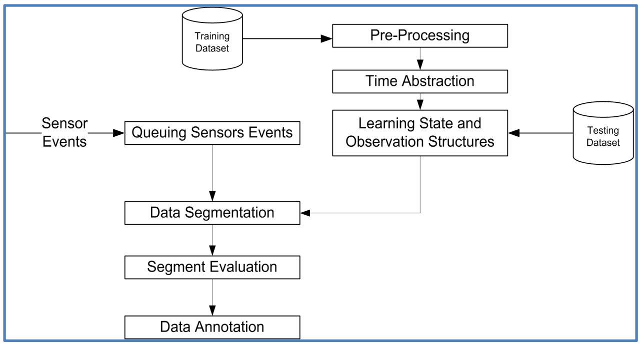


FIGURE 4. Research methodology.

abstract the time intervals in order to overcome time differences and assign more weight to actions that last a long time. After this, the resulting training set is exposed to a learning algorithm to build a state transition structure and observation sequence for each activity. The structures are validated using a testing set of data, which has been prepared for use in segmenting and annotating incoming actions.

Since our target mechanism is to segment actions once they arrive, upon arrival an action is queued and moved to the segmentation module. The segmentation task is implemented as an optimization algorithm, which computes the probability that a given action belongs to every possible activity. Finally, a segment is assigned a label name matching the highest-probability related activity.

A. TEMPORAL ABSTRACTION

During the pre-processing task, time stamps are transformed into intervals that reflect the durations of every action. Time intervals ($\mathrm{I} = \mathrm{t}^- - \mathrm{t}^+$) should be abstracted in a categorical form that can be used to temporally classify segments. Instead of identifying the intervals as a set of starting and ending times or using duration, each interval must be categorized as low, medium, or high duration. To decide the category of each interval, we consider the duration of all intervals in each dataset in order to specify the median duration.

$$\text{Category (I)} = \left\{ \begin{array}{l|c} \text{Low (I)} & 0 \leq \mathrm{d}\,(\mathrm{i}) \leq (\text{Average}/3) \\ \text{Med (I)} & (\text{Average}/3) < d\,(\mathrm{i}) < \mathit{Average} \\ \text{High(I)} & \text{Max} \leq \mathrm{d}\,(\mathrm{i}) \leq \text{Average} \end{array} \right\} \tag{9}$$

Since activities vary in duration, grouping of activities to formulate parameter ω_{I} in equation (7) will depend on the categories of their intervals rather than their exact start and end times. Such an abstraction of interval duration will help understand different instances of the same activity.

B. STATE AND OBSERVATION STRUCTURES

To model the state and observation structure for each activity, we need first to define three parameters: (1) the transition probability among actions, (2) the observation probability of sensors and their actions, and (3) the initial probability vector of each action (usually $1/|\mathrm{V}|$).

The transition probability is defined in equation (2) in its simple form. Since the same action may exist in different time durations, we consider giving higher priority to actions that last longer. Our hypothesis states that actions with large time intervals are expected to identify the context better than those with small time intervals. For this reason, we improved the simple form of embedded interval durations using the following formulae:

$$\mathrm{P}\,(\mathrm{v_t} \mid \mathrm{v_{t-1}}) = \mathrm{P}\,(\mathrm{v_t} \mid \mathrm{I})\,\mathrm{P}(\mathrm{v_{t-1}}|\mathrm{I}) \tag{10}$$

The second parameter to state is the observation probability of sensors pointing to the set of actions. Equation (4) computes such a probability by spanning the training dataset and identifying the value of transition. Since our interval duration hypothesis has been integrated in equation (10), a simple improvement is added to equation (4) to reflect the category of action with respect to its time interval.

$$\mathrm{P}\,(\mathrm{s}, \mathrm{v}, \mathrm{I}) = \prod\nolimits_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{P}\,(\mathrm{v_t} \mid \mathrm{v_{t-1}})\,\mathrm{P}(\mathrm{s_t}|\mathrm{v_{t,I}}) \tag{11}$$

By applying equations (10) and (11) to the training dataset, the state and observation structure can be simply created. Indeed, resolving concurrency requires applying equations (5) and (6). The probability in these equations is as follows

$$\mathrm{P}\,(\mathrm{V} \mid \mathrm{S}, \mathrm{I}) = \frac{1}{\mathrm{Norm(S)}} \times \exp\left(\sum\nolimits_{\mathrm{i}}^{\mathrm{N}} \sum\nolimits_{\mathrm{t=1}}^{\mathrm{T}} \varphi_{\mathrm{i}}\, \mathrm{f_i}\left(\mathrm{v_{t-1.I}}, \mathrm{v_{t,I}}, \mathrm{S}, \mathrm{t}\right)\right) \tag{12}$$

And

$$\mathrm{f_i}\left(\mathrm{v_{t-1,I}}, \mathrm{v_{t,I}}, \mathrm{S}, \mathrm{t}\right) = \prod \omega_{\mathrm{I}} \varphi_{\mathrm{loc}}^{\mathrm{t,I}} \tag{13}$$

C. SEGMENT ANNOTATION

The forwarding algorithm to span the existing structures requires an initialization phase, a recursion phase to emit all states, and a termination phase. This strategy identifies the relevant activities of incoming actions, computes the accumulated probability of incoming actions until the probability value no longer increases, and finally decides to which activity the new segment belongs.

1) INITIALIZATION PHASE

Step #1. $Q \leftarrow push\,(new\ v)$

Step #2. *foreach activity a* $\in A$

a. $W = Compute\ P\,(s, v, I) = \prod_{t=1}^{T} P\,(v_t \mid v_{t-1})\ P(s_t|v_{t,I})$

b. *if* **a** *is a concurrent activity with* A', then Go To Step 1

c. $W = Compute\ P\,(V \mid S, I) = \frac{1}{Norm(S)} \exp\left(\sum_i^N \sum_{t=1}^{T} \varphi_i f_i \left(v_{t-1.I}, v_{t,I}, S, t\right)\right)$

2) SPANNING PHASE

Step #1. $Pop\left(v'\right)$
Step #2. $compute\ W'\ for\ v'\ foreach\ activity$
Step #3. $if\ W' \geq W,\ then\ Go\ To\ Step\ \#1$

3) TERMINATION PHASE

Step #1. $Segment = A'$
Step #2. $Label = Max(P_{i,a})$

V. EXPERIMENT AND RESULTS

In this section, we implemented our proposed technique with three well-known datasets: Tulum, Cairo, and Milan. We cleaned the datasets by converting time stamps into single intervals, removing unlabeled tuples, and transforming intervals into a categorical field according to equation (9).

A. DATASETS

Every dataset comprises instances covering a finite set of activities. Actions are generated using motion, temperature, or detection sensors. Table 2 briefly describes the datasets [32]. Note that every dataset has an attached map that shows the location of sensors, which can be interpreted as the location where a specific action fires.

TABLE 2. Datasets description.

Dataset Name	Size	# Activities	# Sensors
Tulum	1048576	16	20
Cairo	158409	10	25
Milan	433665	15	33

Instances in these datasets represent the daily activities of a single resident and were collected and labeled manually, so that experiments could be supervised using already annotated instances.

B. RECOGNITION ACCURACY

This section reports the results from applying our proposed annotation technique. To measure the performance of the proposed technique, we used accuracy, defined as the ratio between the true positive and negative and all other confusion matrix parameters:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{14}$$

Where TP is the number of instances that have been correctly annotated, TN is the total number of instances that are correctly rejected, FN is the total number of instances that have incorrectly rejected, and FP is the total number of instances that have been incorrectly annotated.

Furthermore, we performed experiments using two different annotation techniques in order to compare our results and measure the resulted enhancements. First we implement an annotation algorithm using TF-IDF similarity model. This model relies on comparing a given testing record with existing training records. The most similar label is chosen accordingly. In addition, we applied KNN (K-nearest Neighbor) algorithm between the training and the testing set in order to assign labels to the testing sets using Euclidian distance function. Indeed the datasets have been processed to cope with these models. Moreover, we presented our methodology into two different experiments; profiling and profiling with temporal relations enrichment. Such partitioning will be useful to measure the impact of using Allen's relations on the annotation process

Table 3 shows the results of comparing the four implementations and reports the accuracy measure in addition to its performance components. Note that, the numbers in Table 3 are the average of performing annotation on all available labels.

TABLE 3. Results in terms of confusion matrix parameters.

Algorithm	Dataset	TP	TN	FP	FN	Accuracy
TF-IDF	Tulum	52.10%	53.90%	46.10%	47.90%	**53.00%**
	Cairo	58.40%	59.60%	40.40%	41.60%	**59.00%**
	Milan	56.00%	52.00%	48.00%	44.00%	**54.00%**
KNN	Tulum	59.90%	61.00%	39.00%	40.10%	**60.45%**
	Cairo	67.30%	66.80%	33.20%	32.70%	**67.05%**
	Milan	64.70%	59.20%	40.80%	35.30%	**61.95%**
HMM Profiling	Tulum	71.40%	70.90%	29.10%	28.60%	**71.15%**
	Cairo	76.00%	70.50%	29.50%	24.00%	**73.25%**
	Milan	73.80%	69.30%	30.70%	26.20%	**71.55%**
HMM Profiling with Temp	Tulum	72.00%	73.00%	27.00%	28.00%	**72.50%**
	Cairo	81.00%	69.00%	31.00%	19.00%	**75.00%**
	Milan	78.00%	66.00%	34.00%	22.00%	**72.00%**

Table 3 shows that the traditional TF-IDF similarity technique was the worst over other techniques. In fact, TF-IDF is a simple and easy to implement technique that performs well in comparing documents rather than concepts with semantic meaning. It has been applied for annotating free text; while generate a remarkable error (miss annotation) on actions and activities.

While KNN outperformed TF-IDF results, the average accuracy does not exceed 67% on Cairo dataset. Since KNN implement the Euclidian distance function, more information or features are required to narrow similar activities.

TABLE 4. Results in terms of enhancement over other methods.

Algorithm	Dataset	Accuracy	Enhancement
TF-IDF	Tulum	53%	20%
	Cairo	59%	16%
	Milan	54%	18%
KNN	Tulum	60%	12%
	Cairo	67%	8%
	Milan	62%	10%
HMM Profiling	Tulum	71%	1%
	Cairo	73%	2%
	Milan	72%	0%

Table 4 shows the enhancement of applying our profiling annotation technique using Allen's relations enrichment over other techniques. We also included the profiling version before adding the temporal features.

Results in Table 4 shows statically significant enhancements of our profiling technique over TF-IDF and KNN ($p < 0.10$) respectively. While, on the other hand, applying temporal relations does not affect the results significantly.

It is the human behavior, which is usually ad-hoc and responsive, that affects the impact of temporal relations on the performance accuracy. At home, many actions are performed in an unpredicted manner. On the other hand, spatial relations have much impact on the performance as they are, by nature, linked to many in-home human activities.

C. IMPACT OF PROFILING ON CLASSIFICATION ALGORITHMS

In this section, we provide extra experiments that show the impact of our profiling technique on state-of-the-art classification algorithms. We chose algorithms from different categories: Vector Space (Support Vector Machine SVM), Decision Tree (J48), and Neural Network (Naïve Bayes NB).

Our purpose from this experiment is to show that adding profiling features will enhance the performance of the classification task. Table 5 shows the application of these classifiers on the raw datasets before adding the profiling features.

TABLE 5. Results of applying classification algorithms on raw datasets.

Algorithm	Dataset	TP	F-Measure
SVM	Tulum	62%	61%
	Cairo	87%	86%
	Milan	75%	74%
J48	Tulum	66%	65%
	Cairo	89%	89%
	Milan	76%	75%
NB	Tulum	56%	54%
	Cairo	61%	64%
	Milan	62%	65%

TABLE 6. Results of classification algorithms after applying the profiling features.

Algorithm	Dataset	TP	F-Measure	Enhancement
SVM	Tulum	77%	76%	15.9%
	Cairo	91%	91%	4.8%
	Milan	84%	83%	8.9%
J48	Tulum	72%	71%	6.0%
	Cairo	93%	92%	3.3%
	Milan	82%	81%	5.7%
NB	Tulum	68%	67%	12.3%
	Cairo	71%	70%	6.0%
	Milan	68%	67%	2.5%

Table 6 shows the enhancements on F-Measure after applying the profiling features to the raw datasets. The results showed statistically significant enhancements ($p < 0.10$) over state-of-the-art algorithms, which support our hypothesis.

D. SENSITIVITY ANALYSIS

In this section, we present a sensitivity analysis that shows the relationship between accuracy and the number of instances as a training set. The importance of such analysis is that it shows the impact of the training set on a given learning algorithm.

The findings showed a positive, trending relationship between the size of the training set and the resulting accuracy: the bigger the training set, the better the resulting accuracy.

Figure 5 shows a trending analysis of the size of the training datasets against their accuracy. The analysis shows a clear trend and positive relationship between the size of the training sets and the accuracy of the annotation process.

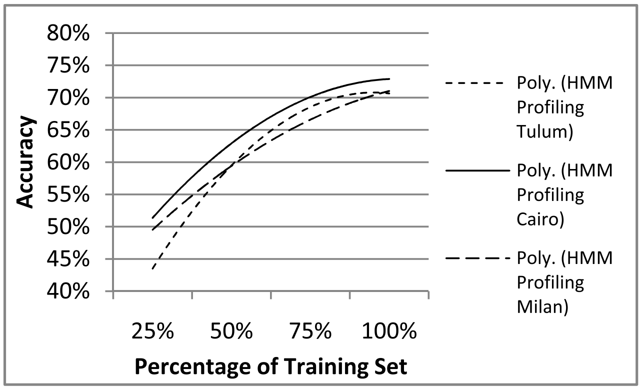


FIGURE 5. Impact of training set size on the accuracy measure.

VI. CONCLUSION

This paper introduces an efficient technique for annotating activities in smart home environments, where performance, ambiguity, and concurrency are frequently required. The contributions of this research were: (1) the modeling of activity actions as a set of states and transitions using HMM, (2) the modeling of a transition feature function that embeds temporal and spatial relations among consecutive actions, and (3) defining the segmentation problem as an optimization problem that minimizes the impact of ambiguity on overall accuracy.

We presented a novel solution that incorporates versions of the Hidden Markov Model and Conditional Random Field model that modified by integrating spatial and temporal relationships among actions to enhance the accurate detection of segment labels. Furthermore, we propose an algorithm to automatically segment incoming actions using state and observation structures. Experimental results showed that our proposed technique is efficient compared to existing, state-of-the-art models.

REFERENCES

[1] M. S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: A voice pathology detection paradigm for smart cities," *Multimedia Syst.*, pp. 1–11, 2017, doi: 10.1007/s00530-017-0561-x.

[2] T. K. L. Hui, R. S. Sherratt, and D. D. Sánchez, "Major requirements for building smart homes in smart cities based on Internet of Things technologies," *Future Generat. Comput. Syst.*, vol. 76, pp. 358–369, Nov. 2017.

[3] M. G. H. Al Zamil and S. Samarah, "The application of semantic-based classification on big data," in *Proc. 5th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2014, pp. 1–5.
[4] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in *Proc. 8th Int. Conf. Intell. Environ. (IE)*, Jun. 2012, pp. 214–221.
[5] M. G. H. Al Zamil, S. M. J. Samarah, M. Rawashdeh, and M. A. Hossain, "An ODT-based abstraction for mining closed sequential temporal patterns in IoT-cloud smart homes," *Cluster Comput.*, vol. 20, no. 2, pp. 1815–1829, 2017.
[6] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Trans. HumanâĂŞMach. Syst.*, vol. 47, no. 3, pp. 368–379, Jun. 2017.
[7] S. Samarah *et al.*, "An efficient activity recognition framework: Toward privacy-sensitive health data sensing," *IEEE Access* vol. 5, pp. 3848–3859, 2017.
[8] M. G. H. Al Zamil and S. Samarah, "Dynamic event classification for intrusion and false alarm detection in vehicular ad hoc networks," *Int. J. Inf. Commun. Technol.*, vol. 8, nos. 2–3, pp. 140–164, 2016.
[9] U. A. B. U. A. Bakar, H. Ghayvat, S. F. Hasanm, and S. C. Mukhopadhyay, "Activity and anomaly detection in smart home: A survey," in *Next Generation Sensors and Systems*. Cham, Switzerland: Springer, 2016, pp. 191–220.
[10] M. Al Zamil and A. B. Can, "Toward effective medical search engines," in *Proc. 5th Int. Symp. Health Inform. Bioinform. (HIBIT)*, Apr. 2010, pp. 21–26.
[11] G. Sprint, D. Cook, R. Fritz, and M. Schmitter-Edgecombe, "Detecting health and behavior change by analyzing smart home sensor data," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2016, pp. 1–3.
[12] E. Chinellato, D. C. Hogg, and A. G. Cohn, "Feature space analysis for human activity recognition in smart environments," in *Proc. 12th Int. Conf. Intell. Environ. (IE)*, Sep. 2016, pp. 194–197.
[13] G. Chetty, M. White, M. Singh, and A. Mishra, "Multimodal activity recognition based on automatic feature discovery," in *Proc. Int. Conf. Comput. Sustain. Global Develop. (INDIACom)* Mar. 2014, pp. 632–637.
[14] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Pervasive Computing* (Lecture Notes in Computer Science), vol. 3001, A. Ferscha and F. Mattern, Eds. Berlin, Germany: Springer-Verlag, Mar. 2004, pp. 158–175.
[15] M. Philipose *et al.*, "Inferring activities from interactions with objects," *IEEE Pervasive Comput.*, vol. 3, no. 4, pp. 50–57, Dec. 2004.
[16] D. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intell. Syst.*, vol. 27, no. 1, pp. 32–38, Jan./Feb. 2012.
[17] T. Duong, D. Phung, H. Bui, and S. Venkatesh, "Efficient duration and hierarchical modeling for human activity recognition," *Artif. Intell.*, vol. 173, nos. 7–8, pp. 830–856, 2009.
[18] T. G. Dietterich, "Machine learning for sequential data: A review," in *Proc. Joint IAPR Int. Workshops SSPR, SPR, Windsor*, ON, Canada, Aug. 2002, pp. 227–246.
[19] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive Mobile Comput.*, vol. 10, pp. 138–154, Feb. 2014.
[20] X. Guan, R. Raich, W.-K. Wong, "Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden Markov model," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 2330–2339.
[21] M. H. Kolekar and D. P. Dash, "Hidden Markov model based human activity recognition using shape and optical flow based features," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2016, pp. 393–397.
[22] Y. Kwon, K. Kang, J. Jin, J. Moon, and J. Park, "Hierarchically linked infinite hidden Markov model based trajectory analysis and semantic region retrieval in a trajectory dataset," *Expert Syst. Appl.*, vol. 78, pp. 386–395, Jul. 2017.
[23] Y. Tong, R. Chen, and J. Gao, "Hidden state conditional random field for abnormal activity recognition in smart homes," *Entropy*, vol. 17, no. 3, pp. 1358–1378, 2015.
[24] A.-A. Liu, W.-Z. Nie, Y.-T. Su, L. Ma, T. Hao, and Z.-X. Yang, "Coupled hidden conditional random fields for RGB-D human action recognition," *Signal Process.*, vol. 112, pp. 74–82, Jul. 2015.
[25] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newslett.*, vol. 12, no. 2, pp. 74–82, 2011.
[26] T. Choudhury *et al.*, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Comput.*, vol. 7, no. 2, pp. 32–41, Jun. 2008.
[27] M. W. Lee, A. M. Khan, and T. S. Kim, "A single tri-axial accelerometer-based real-time personal life log system capable of activity classification and exercise information generation," *Pers. Ubiquitous Comput.*, vol. 15, no. 8, pp. 887–898, 2011.
[28] M. F. Alhamid, M. Rawashdeh, H. Dong, M. A. Hossain, and A. El Saddik, "Exploring latent preferences for context-aware personalized recommendation systems," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 4, pp. 615–623, Aug. 2016.
[29] M. G. A. Zamil, "Verifying smart sensory systems on cloud computing frameworks," *Proc. Comput. Sci.*, vol. 52, no. 1, pp. 1126–1132, May 2015.
[30] J. F. Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM*, vol. 26, no. 11, pp. 832–843, 1983.
[31] M. S. Hossain, G. Muhammad, W. Abdul, B. Song, and B. B. Gupta, "Cloud-assisted secure video transmission and sharing framework for smart cities," *Future Generat. Comput. Syst.*, to be published, doi: 10.1016/j.future.2017.03.029.
[32] (Sep. 15, 2017). *CASAS Dataset*. Accessed: Sep. 15, 2017. [Online]. Available: http://casas.wsu.edu/datasets/

MOHAMMED GH. AL ZAMIL received the B.Sc. and Master's degrees in computer science from Yarmouk University (YU), Jordan, and the Ph.D. degree in information systems from Middle East Technical University, Ankara, Turkey, in 2010. He is currently an Associate Professor with the Department of Computer Information Systems, YU. His research interests include data mining, wireless sensor networks, model checking, software verification, and software engineering.

MAJDI RAWASHDEH received the Ph.D. degree in computer science from the University of Ottawa, Canada. He is currently an Assistant Professor with the Princess Sumaya University for Technology, Jordan. His research interests include social media, user modeling, recommender systems, smart cities, and big data.

SAMER SAMARAH received the Ph.D. degree in computer science from the University of Ottawa, Canada, in 2008. He is currently an Associate Professor with the Computer Information Systems Department, Yarmouk University, Jordan. His research interests focus on data mining techniques for smart environments, Internet of Things, and big data analysis. He has many publications in this area and a referee for many international journals and conferences.

M. SHAMIM HOSSAIN (SM'09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada. He is currently a Professor with the King Saud University, Riyadh, Saudi Arabia, and an Adjunct Professor of EECS, University of Ottawa. He has authored and co-authored around 160 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include serious games, social media, IoT, cloud and multimedia for healthcare, smart health, and resource provisioning for big data processing on media clouds. He has served as a member of the organizing and technical committees of several international conferences and workshops. He is a member of the ACM and the ACM SIGMM. He has served as the Co-Chair, the General Chair, the Workshop Chair, the Publication Chair, and TPC for over the 12 IEEE and ACM conferences and workshops. He currently serves as the Co-Chair of the IEEE ICME Workshop on Multimedia Services and Tools for Smart-health 2018. He was a recipient of a number of awards, including the Best Conference Paper Award, the 2016 *ACM Transactions on Multimedia Computing, Communications and Applications* Nicolas D. Georganas Best Paper Award, and the Research in Excellence Award from King Saud University. He served as the Guest Editor of the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE (currently JBHI), the *International Journal of Multimedia Tools and Applications* (Springer), *Cluster Computing* (Springer), *Future Generation Computer Systems* (Elsevier), *Computers and Electrical Engineering* (Elsevier), and the *International Journal of Distributed Sensor Networks*. He is on the Editorial Board of the IEEE ACCESS, *Computers and Electrical Engineering* (Elsevier), the *Games for Health Journal and International Journal of Multimedia Tools and Applications* (Springer). He currently serves as a Lead Guest Editor of the *IEEE Communication Magazine*, the IEEE TRANSACTIONS ON CLOUD COMPUTING, the IEEE ACCESS, *Future Generation Computer Systems* (Elsevier), and *Sensors* (MDPI).

AWNY ALNUSAIR received the Ph.D. degree in computer science from the University of Wisconsin–Milwaukee. He was a lecturer with Northwestern University and a Senior Fellow with Robert Morris University. He was also with the Software Development Industry for several years. He is currently an Associate Professor of informatics and computer science at Indiana University-Kokomo. His research interests include software engineering, multimedia information retrieval, programming languages, cloud computing, multimedia systems, and data mining.

SK MD MIZANUR RAHMAN (M'10) received the Ph.D. degree in risk engineering (major in cyber security engineering) from the Laboratory of Cryptography and Information Security, Department of Risk Engineering, University of Tsukuba, Japan, in 2007. He was with the high-tech industry in Ottawa, Canada, where he was involved in cryptography and security engineering for several years. He was also a Post-Doctoral Researcher for several years with the University of Ottawa, the University of Ontario Institute of Technology, and the University of Guelph, Canada. He is currently an Assistant Professor with the Information Systems Department, College of Computer and Information Sciences, King Saud University, Saudi Arabia. He has authored or co-authored over 60 peer-reviewed journal and international conference research papers and book chapters. His primary research interest are cryptography, software security, information security, privacy enhancing technology, and network security. He received the Gold Medal for the distinction marks in his undergraduate and graduate program. He received the IPSJ Digital Courier Funai Young Researcher Encouragement Award from the Information Processing Society Japan for his excellent contribution in IT security research.