TagFree: Identifying Users without Tags in Smart Home Environments

Wen-Chih Peng, WangTing Huang, YiLing Chen and PeiChing Liao
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan, ROC
E-mail:{huangwt, ylchen, pcliao, wcpeng}@cs.nctu.edu.tw

***Abstract*— Since family members have their unique features when living in a smart home environment, user identifications are able to achieve without any tags. In this paper, we propose TagFree system in which users freely move in a smart home environment and TagFree system is able to intelligently identify family member according to sensed data. Specifically, TagFree system consists of two phases: the training phase and the prediction phase. In the training phase, sensed data are collected and then, given a huge amount of sensed data, the profile of users, including the most common sensed data (i.e., tones, weights and location), are discovered. Once the profile of users is built up, in the prediction phase, we propose two scoring algorithms to generate likelihood scores according to the sensed data given. A simulation is implemented to verify the correctness of our proposed system and extensive experiments are conducted. Experimental results show that our proposed TagFree is able to achieve high accuracy of identifying family member without any tags. Furthermore, from experimental results, we also provided some guidelines to set some important parameters for TagFree.**

***Keywords* : Smart home, pervasive computing, classification, sensor networks.**

I. Introduction

Recent advances in wireless and embedded technologies have already set the stage for the deployment of smart homes. For example, an increasing number of small and inexpensive wireless devices (referred to as sensor nodes [1][4]) are deployed to monitor various environmental measurements such as temperature, humidity and moving behaviors of users. Once collecting such data, a smart home system is able to automatically offer more context-aware services for users. Example services in a smart homes include multimedia services and location-aware resource management [9][10]. A considerable amount of research works elaborate the issues of determining location information of users since location information is an important context. The positioning techniques are roughly classified into two categories. In the first category, users should carry tags or RF-based devices. By exploring tags and RF-based devices, one could easily identify the location information of users. In the second category, sensor networks are deployed to tracking user locations and thus users do not need to equip with any tag or devices [12][5]. In smart homes, the positioning technology employed needs to be weaved into the fabric of our daily life and thus users are able to comfortably live in smart homes. In this paper, we adopt the latter approach in that no any tags or devices are equipped with users. By utilizing positioning techniques in [12][5][13], location information of users could be obtained. In our paper, location information is represented as identifications of sensors. In addition to the location information of users, the identification of users is a vital context as well. However, since no tags or devices are carried by users, identifying users is a challenge issue, which is a very problem we should address in this paper. Though some prior works focus on target classification and tracking problems in wireless sensor networks [6][2], their proposed algorithms mainly deal with the problem of tracking objects in wireless sensor networks deployed in large fields. However, to our best knowledge, our scenario is not the same as the prior works in [6][2] in that the environment is focused on smart home environments. Furthermore, identifying homogeneous objects (i.e., users) is more challenging than heterogeneous object identifications (i.e., identify car or human).

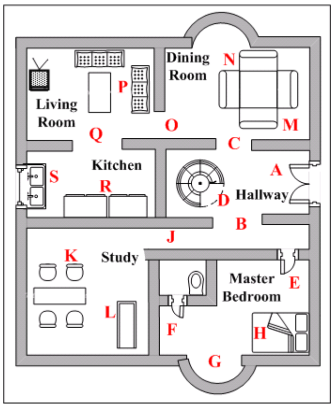


Fig. 1. An example a smart home, where the alphabets denote sensor identifications.

Consider an example profile of users in a smart home, where the floor plan is shown in Figure 1 and profiles of users are given in Table I. Given the profile table in Table I and the sensed values collected from sensors, one should decide the user identification (i.e., the user identification column in Table II). For example in Table II, at time 7:00, from the sensed

User	Weight (kg)	Tone (HZ)	Movement Patterns
U_1	68	103	lh, sh, os, so
U_2	48	338	kh, lh, oh, kho
U_3	53	163	dh, lh, nh, dhn
U_4	53	173	dh, fh, lh, dhf
U_5	43	458	ah, dh, kh, lh

TABLE I
PROFILE OF USERS

Time	Weight (kg)	Tone (HZ)	Location	Identification
07:00	68	103	o	U_1
07:05	47	N/A	k	U_2
07:05	52	164	n	U_3
07:10	54	N/A	l	U_3
07:10	43	457	a	U_5
07:15	47	N/A	k	U_2

TABLE II
A CLASSIFICATION PROBLEM IN A SMART HOME ENVIRONMENT.

values, we could infer that this user is U_1 since some sensed values (i.e., tone and weight) are closely satisfied the profile of U_1 in Table I. However, in reality, some sensed values are not always collected, or cannot be collected accurately. When a user is silent, the sensed value of tones is not collected. Furthermore, some sensed attributes do not always true reflect user profile due to that the change of behaviors of users such as a heavy luggage with users. As can be seen in Table II, at time 7:15, it is hard to identify true user due to that not all sensed readings (i.e., tone) are measured and the attributes of weights cannot provide any unique feature to identify users. Compared to the attributes of weights and tones, the location information is always available. Therefore, one could utilize location information to distinguish users. Location information of users at one time slot is not easy to identify users and note that the moving behavior of users is usually regular and follows some mobility patterns [8][7][3] [11][14]. Therefore, one should collect a considerable amount of location information of users to mine movement patterns of users and these movement patterns could be viewed as one attribute in user profiles.

In this paper, given a profile table and a set of sensed values collected from sensor networks, we should judiciously determine user identification. We first model this problem as a classification problem. Due to the dynamic feature of sensed data, we explore a probabilistic model to predict user identifications. Explicitly, for each attribute data collected, we will develop a corresponding probabilistic model to estimate the probabilistic value of classifying users according to sensed data collected. Then, these likelihood probabilistic values are taken into consideration when aggregating into likelihood probabilistic values of predicting user identification for each user in a smart home. The maximal likelihood value is considered and the corresponding user is the identification predicted. It is worthy mentioning that since each attribute data has different discriminated degree and the collected frequency, for each attribute, we assign different weight when it comes to generating aggregated probabilistic values. In order to verify our proposed method, we implement a simulation model and conduct a comprehensive performance study to verify the performance of the proposed TagFree system. Experimental results shows that our proposed TagFree system is able to accurately determine the identification of users in a smart home.

The rest of the paper is organized as follows. In Section 2, some preliminaries are given. In Section 3, we develop TagFree system in which several algorithms are proposed. Experimental results are shown in Section 4. Section 5 concludes with this paper.

II. PRELIMINARIES

To facilitate the presentation of this paper, some preliminaries are given in this section. In Section II.A, we first formulate the problem of identifying users in a smart home environment. Then, mining user moving patterns is presented in Section II.B.

A. Problem of Identifying Users

In this smart home, sensors are deployed to measure sensed values such as weights, tones and locations. Through existing positioning techniques [12][5][13], the location information of users is determined and is represented as sensor identification. Movement paths of users are thus viewed as a series of sensor identifications. Furthermore, user profiles are initially built up and those sensed attributes that are useful to identify user identification will be included in the profile table. As mentioned above, given a profile table and the sensed values collected from sensor networks, we should judiciously determine the identification of users. This problem is intrinsically a classification problem. A traditional classification problem is that given a data record that contains a set of attributes, one would like to assign this data record to one specific class label predefined. Generally speaking, in order to precisely assign class label to each data record, a traditional classification problem would first build up prediction model according to a given training dataset. Interesting readers could refer to the data mining related books. Consequently, in Table II, user identification column is viewed as a class label and the number of class labels is equal to the number of family members. Given a set of sensed values (referred to as a record), one should decide the class label for this record. Similar to traditional classification algorithms, assume that a set of training data in which each data record is assigned to one class label is given. Then, we need to build up one prediction model from the training data. Without loss of generality, Table II is used as a set of training data. From Table II, it can be seen not all attribute values are available at all time and a one time location information (i.e., location information at a specific time slot) cannot provide any guide to distinguish users. To deal with missing attribute values, one could refer to the profile of users so as to fill the data. As mentioned before, since user movements are usually regular, one could utilize existing works to mine user movement patterns. Therefore, one should view location information of users as data streams and the location data stream is able to provide more hints to predict user identification. Hence, in this paper, given a profile of users and a set of attributes, including numerical attributes and location data streams, we should precisely determine user identifications.

B. Mining User Movement Patterns

In a smart home environment, when a user moves, the location of the user is updated in one centralized server. The location of a user is represented as a sensor identification. The successive sensor identification stream seems to be endless, and it changes very frequently in proportion to the sensed rate. Therefore, it costs too much to store the whole sensor identification stream and reprocess it whenever a new sensor identification is sent in. The problem can be solved by using a method which is capable of mining moving patterns in one scan. In fact, mining movement pattern is modeled as a VMM (Variable memory Markov Model) training and we adopt a variation of a suffix tree called *emission tree* to maintain the VMM model [8][7][15].

Each edge of an emission tree represents a moving record (i.e., sensor identification) appearing in the moving path. A tree node of an emission tree is denoted as a concatenation of the edge labels from the node to the root. In other words, a tree node labeled as $r_k ... r_2 r_1$ can be reached from the traversal path from *root* $\rightarrow r_1 \rightarrow r_2 \rightarrow ... \rightarrow r_k$. Each tree node will maintains the occurrence number of its label in the moving path. Furthermore, each tree node also records the conditional probabilities of all consecutive moving records given the node label as the preceding segment. For example, according to the conditional probabilities of consecutive moving records of node A in Figure 2, it can be verified that P(A|A), P(B|A), and P(F|A) are 0.5, 0.38 and 0.12, respectively. Consequently, if the most recently moving record is A, by traversing the emission tree in Figure 2, one can estimate the consecutive movement (i.e., A) in this illustrative example.

The construction of an emission tree is briefly described as follows. At the beginning, the emission tree has only one root node with the counts of each moving record appearing in the buffer so far. If the count of moving record r_i is larger than the predefined threshold (i.e., minimal support denoted as δ), one tree node labeled as r_i will be created as the child node of the root. Similarly, tree node r_i will maintain the occurrence count of r_i and the probability distribution table is also associated with the node to record the conditional probability of the next moving record with the prefix segment r_i. Assume that the moving records held by the buffer are $r_1 ... r_{l-1}$. When a new moving record, r_l arrives into the buffer, those statistical information (i.e., counts and the conditional probabilities) should be updated accordingly. In order not to distract readers from the main theme of this paper, interested readers are referred to [15] for the detailed procedure of constructing an emission tree.

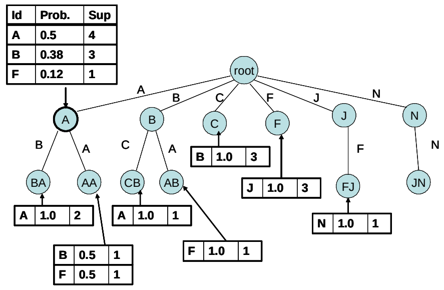


Fig. 2. The resulting emission tree with some selected statistical information.

III. Design of TagFree System

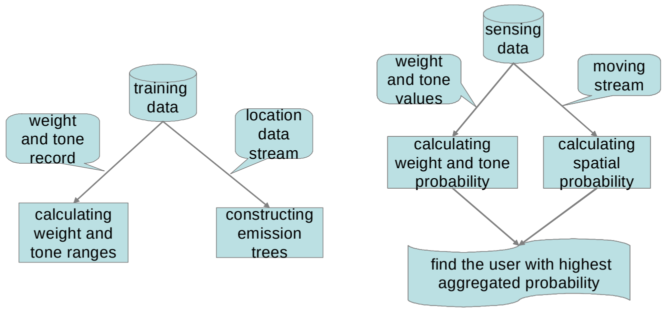
In this section, we first present an overview of TagFree System and then the detailed algorithms are presented.

A. Overview of TagFree System

In TagFree System, profile information should be first built up. Then, given a profile information and the current sensed attributes collected, TagFree system should judiciously identify who the family member is. Such an identification procedure is the same as traditional classification problem. Hence, Figure 3 shows two phases in TagFree system. In the training phase, both of the weight and tone sensed attributes are collected and the average of weight and tone attributes are used as features of individuals. Notice that the location of users is viewed as a series of sensor identification that could be viewed as a data stream). By utilizing mining techniques in [8][7], movement patterns are discovered and each user has his/her own emission tree. Once a profile of users is derived, TagFree system will then be in the prediction phase. In the prediction phase, given a set of attributes (i.e., weight, tone and location) collected, TagFree should classify user identification based on profile information. We model this problem as a Bayesian-like classification problem. For each attribute collected, we will determine the likelihood value that indicates the possibility of inferring one family member. The corresponding likelihood values for a family member are aggregated into one likelihood value (referred to as *aggregate inferring score*). Given a set of aggregate inferring scores, TagFree system will select the maximal aggregate inferring score and thus conclude the corresponding user identification. Since some sensed attributes are numerical and users have their unique features in these numerical attributes, we could employ statistical techniques to decide the likelihood value for these attributes. For location information, given a series of location information (represented as sensor identification), we develop algorithm $SpatilScoring$ to decide the likelihood value. The detailed descriptions are given in the following subsections.

B. Scoring Algorithm for Numerical Attributes

In this section, we will develop a scoring algorithm for numerical sensed attributes. As mentioned before, each user has his/her own unique feature. Note that in the training



(a) training phase

(b) prediction phase

Fig. 3. A framework of TagFree system in a smart home environment.

Time	Weight (kg)	Tone (HZ)	Location (sensor ID)
07:20:00	52	N/A	l
07:20:01	53	N/A	l
07:20:02	52	N/A	h
07:20:03	67	N/A	f
07:20:04	51	N/A	f
07:20:05	52	N/A	h

TABLE III

AN EXAMPLE OF SENSED DATA IN A SMART HOME ENVIRONMENT.

User	w_i	P_i^w
U_1	68	0.2
U_2	48	0.8
U_3	53	0.95
U_4	53	0.95
U_5	43	0.55

TABLE IV

THE RESULT OF DETERMINING $P_{i.}^{W}$.

phase, such attributes are collected. In order to facilitate the presentation of this paper, we consider one numerical sensed attribute (i.e., weight) as an example. Assume that w_i denotes the weight of family member i. Assume that the sensed readings are collected for a period of time window. Therefore, we have the definition of user behavior observed:

Definition 1. User Behavior: The user behavior consists of a series of sensed readings within a sliding window Δt and is expressed as $U(t) = \{S(t - \Delta t + 1), S(t - \Delta t + 2), \ldots, S(t)\}$, where $S(t)$ is the sensed reading at the time t.

In order to filter out some noise readings, we should first calculate the mean value from the sensed readings within a sliding window. Denote w_m as the mean value and w_{sd} as a standard deviation. Therefore, we only consider the readings within the range $[w_m - \theta w_{sd},\ w_m + \theta w_{sd}]$, where θ is an adjustable variable and is used to verify the reasonable range for normal sensed readings. After filtering out the noise readings, the new mean value, expressed by w'_m, is derived. Assume that the number of family members is n and w_i is the weight of family member i. As such, the likelihood score of inferring family member i from the corresponding sensed attribute (i.e., weight) is formulated as follows:

$$\begin{aligned} P_i^w &= 1 - \left|\frac{w'_m - w_i}{w_{range}}\right|, \text{where} \\ w_{range} &= \max\{w'_m, w_i | i \in [1,n]\} \text{-} \min\{w'_m, w_i | i \in [1,n]\} \end{aligned}$$

To make sure that $0 \leq P_i^w \leq 1$, w_{range} is calculated. With larger value of P_i^w, the more likelihood that the unknown user is family member i from the sensed weight value. Similarly, we could determine the likelihood value for other numerical attributes (e.g., tone).

Consider an example in Table III, where both weight and tone attributes are numerical values and the movement of the user is represented as sensor identifications. Also, we demonstrate how to calculate P_i^w. First, a series of weights (i.e., $\{52, 53, 52, 67, 51, 52\}$) is collected. Then, the mean value w_m and the standard deviation w_{sd} of the sensed values are 54.50 and 5.62, respectively. Assume that θ is set to 2. Clearly, 67 is filtered out and we could derive the new mean w'_m as $\frac{52+53+52+51+52}{5} = 52$. Consider a profile table in Table I, where there are five family members with their corresponding weights (i.e., $w_1 = 68$, $w_2 = 48$, $w_3 = 53$, $w_4 = 53$ and $w_5 = 43$). It can be verified that the $w_{range} = \max\{52, 68, 48, 53, 53, 43\}$-$\min\{52, 68, 48, 53, 53, 43\} = 68 - 48 = 20$. Consequently, $P_1^w = 1 - |\frac{52-68}{20}| = 0.2$. Following the same procedure, we could have likelihood values of other users shown in Table IV.

C. Scoring Algorithm for Location Attribute

Note that family members usually have their own moving behaviors at home. For example, parents are likely to move from their bedroom to the living room. Therefore, one could utilize movement patterns to infer the possible family member. In this section, given a set of emission trees and a location data stream, we develop algorithm *SpatialScoring* to derive the likelihood value of location for each family member (denoted as P_i^L for family member i). From a set of likelihood values of all family members, we select the maximal likelihood value and infer the unknown user is the corresponding family user (e.g., i if P_i^L is maximal among a set of likelihood values in terms of location). In the following paragraph, we demonstrate how to derive the likelihood value of location given one emission tree.

Since the location information is viewed as location data streams, we could match a given location data stream with an emission tree and determine the corresponding likelihood value. The likelihood value is decided according to matching between a location data stream and emission tree nodes. Suppose that each node of an emission tree has one score and the score of a node is set to the value of its tree level. For example, since a root node of an emission tree is at the level 1, the score of a root node is set to 1. The principle of scoring of location data streams is that when the location

Movement	Location stream	Matching score	Spatial score
b	b	1	1
c	bc	1+2=3	4
a	bca	1+2+3=6	10
d	bcad	1+2+3=6	16

TABLE V
AN EXAMPLE OF CALCULATING P_i^L.

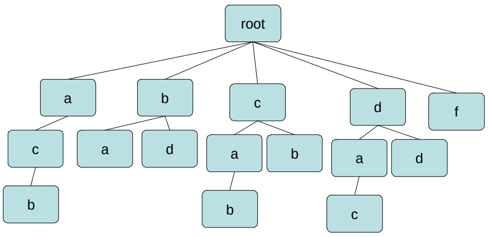


Fig. 4. An example emission tree.

data streams are matched to deep tree nodes, the more likely that the location streams are close to the movement behavior of the corresponding user. Thus, given location streams, we could match location streams with an emission tree and sum the score of tree nodes matched. The sum of the tree nodes matched is referred to as matching score. Furthermore, since location data streams are incoming every time when users move, algorithm $SpatialScoring$ will calculate and accumulate the matching scores. The accumulated matching score is called *spatial score.* Once a new location is collected, this location will be appended as one new location stream. The corresponding matching score will be calculated and the spatial score will be updated. It can be seen that algorithm SaptialScoring is able to dynamically calculate the scores based on the incoming location data.

Consider an illustrative example in Table V, where a family member has four movements and one example emission tree is shown in Figure 4. In the beginning, the location of the user is at b and then b is buffered. To determine the matching score, one should match location streams collected with the emission tree. From matching scenarios, we could decide that the matching score is 1. Hence, the spatial score is updated to 1. The next movement of this family member is c. Therefore, a new location stream is bc. By traveling the emission tree in Figure 4, one could derive the matching score as (1 (matching c) +2 (matching b)=3). Consequently, spatial score is updated as 1+3=4. Following the same procedure, after four movements of this family member, the spatial score is calculated as 16.

The above procedure is used to determine the spatial score for one emission tree given a location stream. Therefore, we could derive the spatial scores for a set of emission trees given. Since each user will has his/her emission tree, for family member i, we denote the spatial score of emission tree i as SS_i. As such, the likelihood score of inferring family member i from the location attribute is formulated as follows:

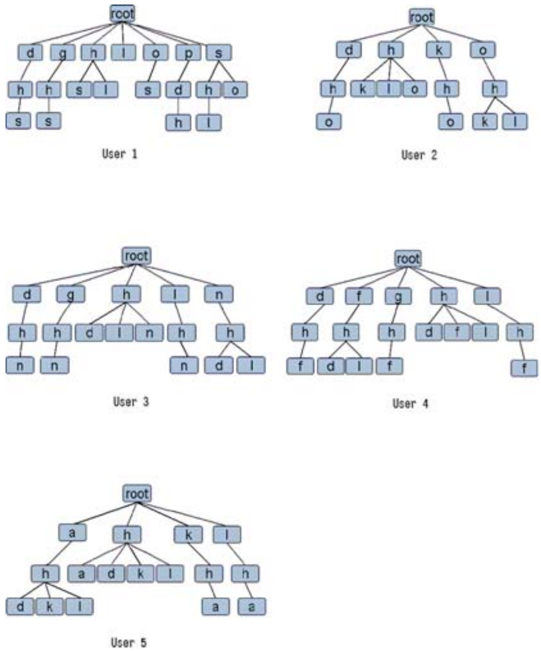


Fig. 5. An example of five emission trees.

Family member U_i	Spatial score	P_i^L
U_1	14	$\frac{14}{14+11+14+33+14} = 0.16$
U_2	11	$\frac{11}{14+11+14+33+14} = 0.13$
U_3	14	$\frac{14}{14+11+14+33+14} = 0.16$
U_4	33	$\frac{33}{14+11+14+33+14} = 0.38$
U_5	14	$\frac{14}{14+11+14+33+14} = 0.16$

TABLE VI
THE FINAL RESULT OF DETERMINING P_i^L.

$$P_i^L = \frac{SS_i}{\sum_{j=1}^{n} SS_j},$$ where n is the number of family members.

Consider our illustrative example in Table III and five emission trees shown in Figure 5. From Table III, there are 6 movements of this family member (i.e., $llhffh$). Thus, we could derive spatial scores for each emission tree and derive their likelihood values shown in Table VI.

D. Determining User Identification

From the above two scoring algorithms, the corresponding likelihood values are determined according to the sensed measurements collected. In this paper, assume that we have two numerical attributes (i.e., tone and weight) and one location attribute. Hence, three likelihood values are available. To facilitate the presentation of this paper, P_i^T and P_i^W are the likelihood values of inferring family member i in terms of tone and weight sensed data. On the other hand, P_i^L is the location likelihood value of inferring family member i. Therefore, we could derive the aggregate likelihood value of inferring family member i as $P_i = r_W * P_i^W + r_T * P_i^T + r_L * P_i^L$, where r_W, r_T

User U_i	r_W	P_i^W	r_T	P_i^T	r_L	P_i^L	P_i
U_1	0.5	0.2	0	N/A	0.5	0.16	0.18
U_2	0.5	0.8	0	N/A	0.5	0.13	0.47
U_3	0.5	0.95	0	N/A	0.5	0.16	0.56
U_4	0.5	0.95	0	N/A	0.5	0.38	0.67*
U_5	0.5	0.55	0	N/A	0.5	0.16	0.36

TABLE VII
THE RESULT OF DETERMING THE AGGREGATED P.

Notation	Definition
N_{train}	The number movement paths in the training phase
N_{pred}	The number movement paths in the prediction phase
δ	A minimum support for mining movement patterns
θ	A threshold value for numerical sensing attributes
V_w	Weight profile for family members
V_t	Tone profile for family members

TABLE VIII
PARAMETERS USED IN THE SIMULATION MODEL.

and r_L are the weight values for these three sensed attributes. These weight values are very dependent to behaviors of family members and one experiment will be conducted to show the impact of these weight values. Consider the example in Table III, where both r_W and r_L are set to 0.5 and r_T is set to zero. Table VII shows the aggregate likelihood value for each user. From Table VII, the maximal likelihood value is selected (i.e., P_4) and hence we could infer that the unknown user is family member U_4.

IV. Performance Study

To evaluate the proposed algorithms for TagFree, we implement a simulation to model a smart home environment. In Section IV.A, the simulation model is described. Section IV.B is devoted to experimental results.

A. Simulation Model

In our simulation model, the floor plan of a smart home environment is shown in Figure 1, where sensors are deployed in the smart home and there are four family members. Then, we have developed an object-oriented discrete-event simulation environment to generate family members' movements, associated prediction of likely paths. Specifically, the sensor deployment in the smart home is viewed as a graph, where a vertex denotes a sensor and an edge between vertexes represents that these two sensors are nearby. As mentioned before, there are four family members and each family member has each own movement paths. Note that in order to simulate movement patterns of family members, we use probability model to model user moving behaviors. Explicitly, each family member has his/her own unique moving behavior, meaning that this member will frequently appear or move in some areas. For example, one family user has higher probability in staying in the kitchen since this family member need to cook for other family members. Furthermore, each family user has his/her own bedroom. Consequently, given areas that one family member usually stays, movement paths among these areas are generated, where movement paths are shortest paths. The movement behaviors are similar to prior works in [9][10]. Following the design issues in [9][10], each family member has his/her own movement pattern. The number of consecutive moving paths employed in the training phase is called the training step (denoted by N_{train}). The number of moving paths used in the prediction phase is called the prediction step (denoted by N_{pred}). As to other sensed attributes (i.e., tone and weight), these sensed attributes are generated at each movement of a user. V_w (respectively, V_t) denotes the vector with the corresponding weights values (respectively, tone values) of family members. When a user moves to a new location, a probability indicates whether a user will generate tone value or not. If a user will produce some accosting values, these values are decided according to the corresponding value in V_t with some noise bias. On the other hand, the sensed value for weights is always acquired and the sensed weight is also generated based on the V_w with some noise bias. A measurement, *accuracy*, is represented as the ratio of the number of correct identifications and the total number of identifications. Table VIII summaries the definitions used for some primary simulation parameters.

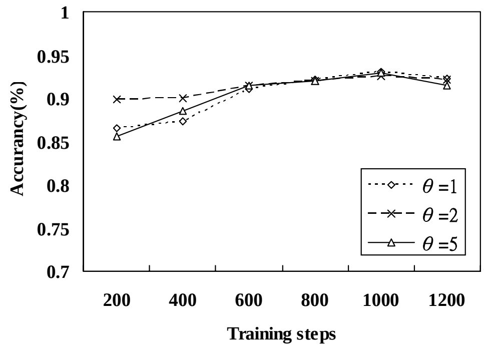


Fig. 6. The accuracy of TagFree with various training steps and θ.

B. Experimental Results

In this section, we conduct some experiments to evaluate the proposed TagFree system. First, we evaluate the impact of training steps for the prediction accuracy. Then, we examine the impact of prediction steps needed in TagFree system. Finally, sensitivity analysis of parameters used in TagFree is conducted.

1) The Impact of Training Steps: As mentioned before, in the training phase, TagFree system will collect data from sensors. Then, these sensed data are used to mine movement patterns and build up the tone and weight profile of family members. In this experiments, we set N_{pred} to 25, δ to 5, V_w to {85, 70, 55, 40}, V_t to {120, 300, 180, 330}, r_W to 0.3, r_T to 0.3 and r_L to 0.4. The experimental results with the number of training steps varied are shown in Figure 6. It can be seen in Figure 6 that with a larger number of training

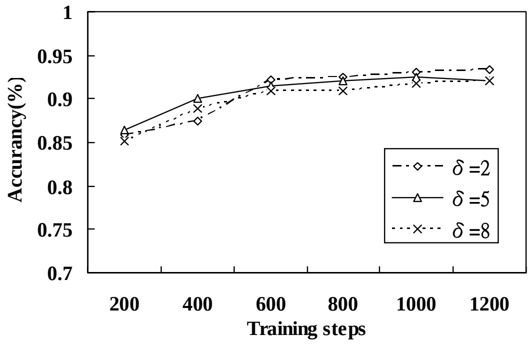


Fig. 7. The accuracy of TagFree with various training steps and δ.

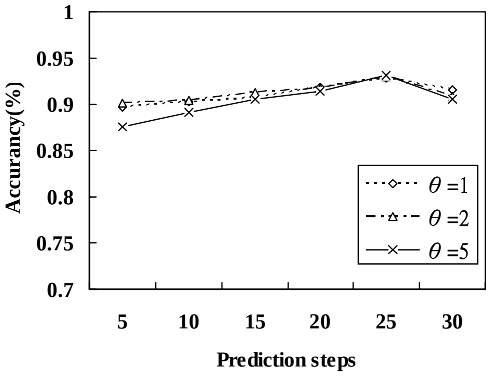


Fig. 8. The accuracy of TagFree with various prediction steps and θ.

steps, the accuracy of TagFree intends to increase. This is due to that with more sensed data available, movement patterns are precisely determined by TagFree. Furthermore, when the number of training steps increases, we could set a smaller value of θ since more data are available to build up the weight and tone profile of user. Next, we investigate the parameter of δ for mining movement patterns of users. Other parameters are set the same as above, except that the value of θ is set to 2. Figure 7 shows the experiential results. It can be verified that with a larger amount of training steps, the accuracy of TagFree tends to increase. Note that for smaller value of δ, an emission tree could quickly generate tree nodes for movement patterns. With a larger number of training steps, the emission tree is able to accurately capture the movement patterns of users, thereby increasing the accuracy of TagFree.

The above two experiments indicate that to increase the accuracy of TagFree, one should collect a sufficient training data by setting a larger number of training steps. Furthermore, smaller values of θ and δ are able to increase the accuracy of TagFree.

2) The Impact of Prediction Steps: After mining movement patterns and building up the profile of family members, in the prediction phase, TagFree will use a series of sensed data collected for identification. Without loss of generality,

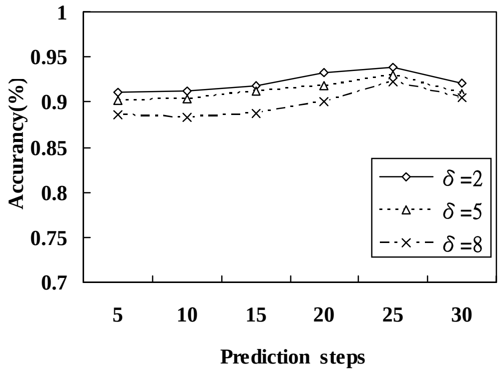


Fig. 9. The accuracy of TagFree with various prediction steps and δ.

we set N_{train} to 1000, δ to 5, V_w to {85,70,55,40}, V_t to {120,300,180,330}, r_W to 0.3, r_T to 0.3 and r_L to 0.4. The experiments of varying the value of θ under various numbers of prediction steps are performed. Figure 8 shows the experimental results. It can be seen Figure 8 that as the number of prediction steps increases, the accuracy of TagFree is also increased. However, when the number of prediction step is too large, the accuracy of TagFree decreases. This is due to that with a larger number of prediction steps, some obsolete location data are not helpful to identify users when scoring the movement path. On the other hand, with a smaller number of prediction steps, the location data does not capture the moving behavior of family users, resulting in the smaller accuracy values. Thus, the number of prediction should judiciously be determined. The setting of θ does not have great impacts for the accuracy of TagFree since the number of prediction steps mainly affect the prediction accuracy of location attributes. Therefore, we conduct experiments with the value of δ varied. The experimental results are shown in Figure 9. Similar to Figure 8, the number of prediction steps should be appropriately determined. The reason is the same as the above. It can be seen in Figure 9, with a smaller δ, more moving behaviors are captured in emission trees, thereby increasing the accuracy of TagFree.

3) Sensitive Analysis of TagFree: In TagFree, the values of r_W, r_T and r_L are important parameters. Clearly, the setting of these parameters depends on the discrimination of each sensed attribute. In this experiments, we model four scenarios to investigate the impact of the discrimination of sensed attributes for the setting of r_W, r_T and r_L. WL (respectively, WS) denotes a significant (respectively, low) discrimination among weights of family members. On the other hand, TL (respectively, TS) represents a significant (respectively, low) discrimination among tones of family members. The profile setting of these four scenarios is shown in Table IX. Under the four scenarios, we set N_{train} to 1000, N_{pred} to 20, θ to 2.

Figure 10 shows the accuracy of TagFree under different scenarios. In Figure 10, it can be seen that the accuracy of

Scenarios	V_w	V_t
WL+TL	{85,70,55,40}	{120,300,130,190}
WL+TS	{85,70,55,40}	{180,270,180,160}
WS+TL	{60,60,60,60}	{120,300,130,190}
WS+TS	{60,60,60,60}	{180,270,180,160}

TABLE IX

SCENARIOS FOR ILLUSTRATING DISCREMINATIONS.

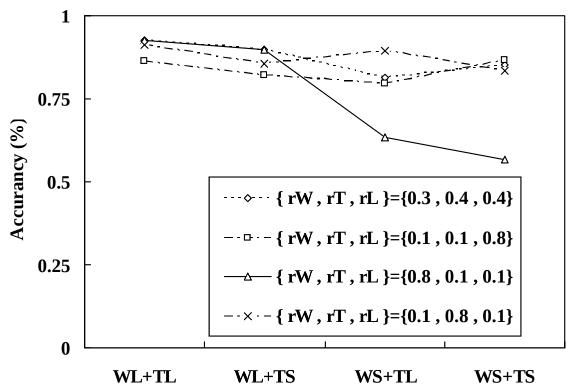


Fig. 10. The accuracy of TagFree under different scenarios.

TagFree under the scenario of WL+TL is higher accuracy than other scenarios. Since discriminations are enough in terms of weights and tones, the setting of r_W, r_T and r_L does not have impact on the performance of TagFree. In scenario of WL+TS, it can be seen that due to the discrimination in weights, setting larger value for r_W is able to increase the accuracy of TagFree. On the other hand, setting larger value of r_T will improve the accuracy of TagFree under the scenario of WS+TL. Thus, once collecting profiles of family users in the training phase, one should further determine discriminations among these sensed attributes so as to facilitate the setting of r_W, r_T and r_L.

V. CONCLUSIONS

In this paper, we claimed that since family members have their unique features when living in a smart home environment, user identifications are able to achieve without any tags. Hence, we proposed TagFree system in which users freely move in a smart home environment and TagFree is able to intelligently identify family member according to sensed data. Specifically, TagFree system consists of two phases: the training phase and the prediction phase. In the training phase, sensed data are collected and then, given a huge amount of sensed data, the profile of users, including the most common sensed data (i.e., tones, weights and locations), is discovered. Once the profile of users is built up, in the prediction phase, we proposed two scoring algorithms to generate likelihood scores according to the sensed data given. We implemented a simulation model to verify the correctness of our proposed system and extensive experiments are conducted. Experimental results show that our proposed TagFree is able to achieve high accuracy of identifying family member without any tags. Furthermore, from experimental results, we also provided some guidelines to set some important parameters for TagFree. In the future, we will implement TagFree system in a real smart home environment to demonstrate the feasibility of TagFree.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communication Magazine*, 40(8):102–114, 2002.

[2] R. R. Brooks, P. Ramanathan, and A. M. Sayeed. Distributed Target Classification and Tracking in Sensor Networks. *Proceedings of the IEEE*, 91(8):1163–1171, Aug. 2003.

[3] G. Cho and L. F. Marshall. An Efficient Location and Routing Scheme for Mobile Computing Environments. *IEEE Journal on Selected Areas in Communications*, 13(5):868–879, June 1995.

[4] D. Estrin, R. Govindan, J. S. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In *Mobile Computing and Networking*, pages 263–270, 1999.

[5] L. Girod, M. Lukac, V. Trifa, and D. Estrin. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proc. of the 4th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 71–84, New York, NY, USA, 2006. ACM Press.

[6] L. Gu, D. Jia, P. Vicaire, T. Yan, L. Luo, A. Tirumala, Q. Cao, T. He, J. A. Stankovic, T. F. Abdelzaher, and B. H. Krogh. Lightweight detection and classification for wireless sensor networks in realistic environments. In *Prof. of the 3rd International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 205–217, 2005.

[7] W.-C. Peng and M.-S. Chen. Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system. *IEEE Transactions on Knowledge and Data Engineering*, 15(1):70–85, 2003.

[8] W.-C. Peng, Y.-Z. Ko, and W.-C. Lee. On mining moving patterns for object tracking sensor networks. In *Proc. of the 7th International Conference on Mobile Data Management (MDM)*, pages 41–46, 2006.

[9] A. Roy, Das, A. Bhattacharya, K. Basu, D. J. Cook, and S. K. Das. Location aware resource management in smart homes. In *Proc. of the 2004 International Conference on Pervasive Computing and Communications (PerCom)*, pages 481–488, 2003.

[10] N. Roy, A. Roy, and S. K. Das. Context-aware resource management in multi-inhabitant smart homes: A nash h-learning based approach. In *Proc. of 2006 International Conference on Pervasive Computing and Communications (PerCom)*, pages 148–158. IEEE Computer Society, 2006.

[11] N. Shivakumar, J. Jannink, and J. Widom. Per-User Profile Replication in Mobile Environments: Algorithms, Analysis and Simulation Results. *ACM Journal of Mobile Networks and Applications*, (2):129–140, 1997.

[12] N. Shrivastava, Mudumbai, and S. Suri. Target tracking with binary proximity sensors: fundamental limits, minimal descriptions, and algorithms. In *Proc. of the 4th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 251–264, New York, NY, USA, 2006. ACM Press.

[13] R. Stoleru, P. Vicaire, T. He, and J. A. Stankovic. Stardust: a flexible architecture for passive localization in wireless sensor networks. In *Proc. of the 4th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 57–70, New York, NY, USA, 2006. ACM Press.

[14] H.-K. Wu, M.-H. Jin, J.-T. Horng, and C.-Y. Ke. Personal Paging Area Design Based on Mobile's Moving Behaviors. In *Proc. of the 2001 International Conference on Computer Communications (INFOCOM)*, pages 21–30, April 2001.

[15] J. Yang and W. Wang. Agile: A general approach to detect transitions in evolving data streams. In *Proc. of the 2004 IEEE International Confernece on Data Mining (ICDM)*, pages 559–562. IEEE Computer Society, 2004.