CACE: Exploiting Behavioral Interactions for Improved Activity Recognition in Multi-Inhabitant Smart Homes

Mohammad Arif Ul Alam[1], Nirmalya Roy[1], Archan Misra[2], Joseph Taylor[1]
[1]Department of Information Systems, University of Maryland Baltimore County
[2]School of Information Systems, Singapore Management University
alam4@umbc.edu, nroy@umbc.edu, archanm@smu.edu.sg, j.taylor@umbc.edu

***Abstract*—We propose CACE (Constraints And Correlations mining Engine) which investigates the challenges of improving the recognition of complex daily activities in multi-inhabitant smart homes, by better exploiting the spatiotemporal relationships across the activities of different individuals. We first propose and develop a loosely-coupled Hierarchical Dynamic Bayesian Network (HDBN), which both (a) captures the hierarchical inference of complex (macro-activity) contexts from lower-layer micro-activity context (postural and improved oral gestural context), and (b) embeds the various types of behavioral correlations and constraints (at both micro- and macro-activity contexts) across the individuals. While this model is rich in terms of accuracy, it is computationally prohibitive, due to the explosive increase in the number of jointly-defined states. To tackle this challenge, we employ data mining to learn behaviorally-driven context correlations in the form of association rules; we then use such rules to prune the state space dramatically. To evaluate our framework, we build a customized smart home system and collected naturalistic multi-inhabitant smart home activities data. The system performance is illustrated with results from real-time system deployment experiences in a smart home environment reveals a radical ($\max\ 16-fold$) reduction in the computational overhead compared to traditional hybrid classification approaches, as well as an improved activity recognition accuracy of $\max\ 95\%$.**

Keywords*—*multiple inhabitants, multi-modal sensing, scalable activity recognizer, smart communities

I. Introduction

CACE focuses on the problem of determining the macro-level activity context of individuals in a multi-inhabitant smart home environment. Examples of such macro-level (or *complex*) activities include "cooking" or "watching TV", which can be inferred by observing both the sequence of low-level (or *micro*) activities of an individual (e.g., "walking" or ''sitting") and the location context associated with such activities. Past works (e.g., [1], [2], [5]) have looked at the possibility of determining such micro-activity and location context based on either on-body (e.g., smartphone-based) or ambient (e.g., motion detectors, RFID tag) sensing, or by judiciously combining both. Typically, the problem of complex activity detection of each individual has been viewed *in isolation*, without taking into account the *correlations* or *constraints* observed collectively, across inhabitants.

Consequently, the presence of multiple occupants is viewed as a *detriment*, as it confounds the ability to associate certain ambient context directly with a specific individual–e.g., a motion detector can detect that an individual is in the "kitchen", but cannot say whether it was person A or B. The limited work (e.g., [4]) that exploits cross-individual context relations has focused primarily on improving the accuracy of the micro-activity or location activity context, as an intermediate step towards computing the high-level activity of each individual occupant. All these approaches, however, make an implicit assumption that a specific activity is associated with a specific location–e.g., "cooking" with the kitchen, or "watching TV" with the living room (more specifically in the couch area). As a consequence, the accuracy of complex activity recognition stagnates at around 75% in presence of multiple users [9], because, in reality, an individual's performance of a complex activity may naturally straddle multiple locations and include unrelated intermediate activities–e.g., as shown in Figure 1, an inhabitant may start "watching TV" while "cooking" and go back and forth between the kitchen and the living room. In this paper, we extend past work in two distinct ways:

Fig. 1. Hierarchical Structure of ADLs

• *Inclusion of Wearable Context:* With the rapid consumer adoption of wearable devices, we consider how the additional availability of *oral gestural micro-context* (such as "talking" or "laughing") can further improve the accuracy of such complex activity detection. We propose a 9-axis sensor fusion based trajectory generation technique to identify and recognize micro-level oral gestural activity.

• *More Rigorous Modeling of Context Correlations and Constraints:* We believe that the presence of multiple occupants in a home may in fact turn out to be an *advantage*, especially if we can judiciously utilize postulated or observed behavioral correlation and relationships *among* the occupants. For example, individual B's context may be inferred (probabilistically) to be more likely "dining" (instead of "washing dishes") if person A is currently observed to be "dining", and if history suggests that the two occupants usually dine together. Thus, we investigate how such behavioral coupling can be formally defined, both at and across micro- and complex-activity levels, to improve the recognition accuracy.

Unfortunately, as the number of individuals and the number of micro-activity states (per individual) increase, the rigorous modeling of such inter-personal constraints and correlations becomes computationally prohibitive, due to the exponential growth in the number of possible *joint* states.

Research Questions: Our research in this paper consequently tackles the following key research questions:

• What sort of inter-user correlation aware, hierarchical activity recognition models can we use to represent a combination of ambient, mobile and wearable sensor data (from multiple inhabitants)?

• How can we intelligently combine, in such a unified hierarchical model, both spatiotemporal constraints and correlations across multiple users (across both micro-and-complex activity levels) so as to significantly improve the computational tractability of the approach?

• Quantitatively (i.e., given real-world sensor traces from multi-inhabitant environments), how much improvement does our proposed approach offer (in terms of computational overhead and improved accuracy), compared to the existing activity recognition approaches?

To tackle these challenges, CACE delegates a coupled Hierarchical Dynamic Bayesian Network (HDBN) model, which organizes activity recognition in a hierarchy, with low-level micro-activity context states being used to infer higher-level complex activity contexts.

More specifically, we make the following **key contributions**:

• We first propose a generic model using a probabilistic HDBN for a single inhabitant and extend it for multi-inhabitant cases using a coupled HDBN model. This model both (a) incorporates postural and gestural micro-activities, as well as location context, and (b) formally captures the different relationships that exist at the crossroads of micro and macro-level activities for individual and multiple inhabitants.

• Most importantly, we utilize a well-established data mining approach (using past traces of behavioral data) to discover key spatiotemporal constraints in the activity contexts across users, and use these constraints to prune the overall state space of the coupled model (loosely-coupled HDBN) during the model training phase. Such a model effectively trades off a little accuracy (by eliminating very unlikely state sequences) for significant computational gain.

• We evaluate our proposed model using our smart home collected dataset and a realistic dataset that matches best with our scenario [9]. We run our data-driven state space pruning miners and show that (i) a loosely-coupled HDBN model improves the complex activity recognition accuracy to $\max$ 95% (compared to $\approx 75\%$ achieved by prior approaches [9]); (ii) the intelligent fusion of correlations and constraints helps to reduce the computational complexity dramatically (by a factor of 16).

II. Related Works

The idea of combining ambient sensors along with wearable and smart phone sensors to recognize micro and macro ADL in smart home environment has been investigated in the past [1]–[4], [7], [16], [17] but very few of them explored the problem of multiple occupant's activity recognition in smart home environment [4], [5], [17]. However, lack of naturalistic data collection and additional context effect (say oral gestural, correlations and constraints) on activity recognition performance in prior works signify the importance of CACE.

[16] proposed a two-staged method defining multi-inhabitant activities as combined label. SENST* [7] inferred hidden location context by substituting expensive sensors with inexpensive ones in inference cache which supports arbitrary context attributes related to each other by a set of automatic learned general association rules. ACE [1] introduced an association rule mining approach to infer hidden micro contexts by exploiting context correlations in an energy-efficient way. A simple 2-step DBN approach for a single user environment has been proposed considering context classifier's uncertainty into account [2]. [4] added micro context constraints among all users in a multi-inhabitant smart home environment using Coupled Hidden Markov Model (CHMM). Activity recognition algorithms have been proposed to classify either micro or macro-level activities separately [3] or jointly using a two-step classification model [4]. [5] investigated CHMM and Factorial Conditional Random Field (FCRF) for activity daily livings (ADLs) recognition in multi-inhabitant smart home environment relying on multiple body-worn sensors with multi-modal sensing capabilities. The closest work to CACE framework is NCB [17] which exploited the coupled behaviors among socially connected people based on their community behaviors, mining the social links/networks. The central difference between NCB and CACE is that former one proposed context mining to exploit higher level contexts coupling socially connected people. For example, social behaviorally connected people having office at 8.00 AM, take shower at 6.00 AM, take breakfast at 6.30 AM and leave home within 7.00 AM. However, we propose user context correlations and constraints coupling residentially connected people. For example, in two inhabitants apartment, while one is taking shower, other may prepare breakfast or do vice versa. Apparently they take breakfast together and talk each other displaying contexts correlations and constraints. While, neither of the existing frameworks address the multi-inhabitant complex (macro) activity recognition cases in presence of more than one type of micro activities (such as locomotive, postural and gestural) nor handle their performance degradation in case of any missing sensor values, we propose a single unified DBN-based model to infer activities by conglomerating micro and macro-level activities and intra- and inter-user relationships hierarchically. While other failed to handle multiple-occupancy identification, did not consider the correlations to speed-up and loosen the complexity of a hierarchical model sustaining the accuracy, CACE exploits macro contexts hierarchically for each user in a coupled way. However, unlike ACE [1], CACE can generate some initial rule sets through a user-friendly smartphone user-interface.

III. Terminologies and Overall Framework

Our proposed framework exploits different aspects of the relations (both spatial and temporal) between the activities of multiple individuals cohabiting a smart home. Many researchers ([1], [4], [17]) use the terms context correlations and constraints relationships in real world scenario. We formally define these two fundamental types of relationships:

Definition 1. ***C****ontext Correlation: Correlation captures the* deterministic *relationships (must or must not) between two context states (micro-micro, macro-macro and micro-macro) , such that the identification of one context state immediately identifies the possible value for the other context state.*

For example, in a smart home with two individuals, if person A's macro context is "sleeping" in the bedroom, then the other person B's macro context must not be "vacuuming" in the bedroom at the same time (as 'vacuuming' will disrupt person A). Similarly, if person A's micro-context is "sitting" on the couch, then he cannot be 'walking' in the next time instant, as there must be an intervening "standing" postural state. In past work [1], such correlations have been derived (using association rule mining) among various micro-contexts of an *individual* user. In our work, we extend this approach to consider correlations both (i) across different users and (ii) between and across micro and macro-level contexts.

Definition 2. ***C****ontext Constraint: Constraint is a probabilistic measure of the inter-connects and uncertainty levels among different context states, such that the likelihood of occurrence of one context state helps change the a-priori probability of the concurrent occurrence of another context.*

For example, if person A's macro activity is "dining" in the living room, then person B is more likely to be "dining" in the living room. Similarly, if person A is "dining" at present, then, in the next time instant, it is unlikely that person A will be 'jogging" in the living room. Note that our terminology may differ a bit from the lay meaning of the terms: in our framework, correlations are *deterministic*, whereas constraints represent *probabilistic* relationships.

A. Overview of Our Framework

Our framework assumes a multi-inhabitant smart home that includes (a) *ambient, static sensors* that capture context states not directly linked to a specific individual, and (b) *mobile and wearable sensors* that capture individual-level activity context. As a specific exemplar, this paper considers an environment with (i) one binary passive infrared sensor in each room, that indicates whether a particular room is occupied by one or more moving individuals, (ii) one binary high sensitive object sensor on concerned objects, that indicates their possession by one or more inhabitants, and (iii) each inhabitant carrying a smartphone and/or smart-jewellery (in the neck), whose sensors help detect various postural and oral gestural states.

Figure 2 explains the basic steps in our overall context processing pipeline: (1) 'Sensing planar' gathers multi-modal sensor data from smartphones and ambient motion sensors; (2) 'Context planar' performs feature extraction on such ambient, mobile and wearable sensor data; (3) 'State space creation' is responsible for combining micro-level activity tuples (both gestural and locomotive) with ambient context ; (4) 'State space reduction' utilizes data-mined relationships (across both micro and macro-level contexts) in the multi-user state space to reduce the number of unknown context states; (5) 'Loosely-coupled HDBN' model helps construct the probabilistic inter-connections between the macro and micro-level activities of multiple users; and (6) the eventual 'Inference engine' infers the most-likely *sequence* of macro-activities (and their time boundaries), utilizing the well-known Expectation Maximization (EM) algorithm for training and the Viterbi algorithm [13] for runtime inference.

TABLE I. VARIABLE REPRESENTATION

Notation	Description
$\mathcal{N}$	No. of individuals
$context^i_j(t)$	User context state at time t and m-dimensional tuple for i^{th} user where $j = 1$ (micro) and $j = 2$ (macro) level context state $\forall m = 1, 2, \ldots, M$
$\mathcal{K}$	No. of motion sensors, $k = \langle 1, \ldots, k \rangle$ location

Our key innovations in this paper lie in: (a) the use of spatiotemporal rules (both deterministic correlations and statistical constraints) to the state space (step 4) to be explored and (b) the use (step 5) of a coupled HDBN model, incorporating both micro and macro contexts, to improve the accuracy of context estimation.

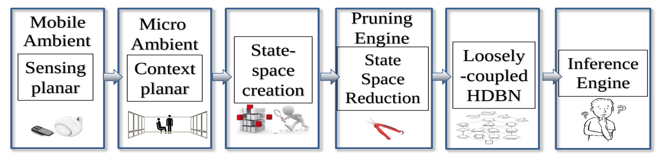


Fig. 2. An overview of our hierarchial activity recognition framework

IV. MULTI-INHABITANT HIERARCHICAL MODEL

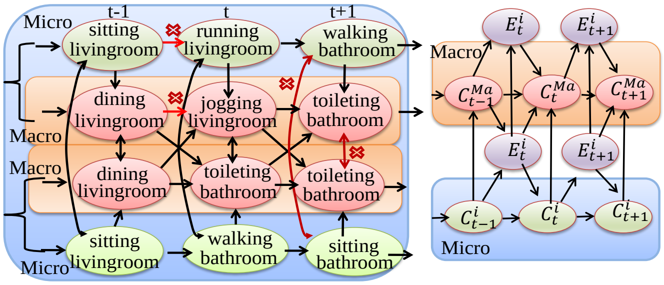
Coupling of a traditional HDBN model across multiple inhabitants helps form a unified model but ramify the complexity of handling the exponential increase in number of state spaces. We propose a correlation miner to prune the state spaces based on Apriori algorithm generated association rules relying on the context relations. We define four dependency augmentations to incorporate our correlation and constraint miner with the traditional coupled model of HDBN.

A. Model Variables Representation

We represent some variables in defining generic model as stated in Table I. An important characteristic of our model is that a subset of the $\mathcal{M}$ elements are 'observable' and can be inferred using solely the sensors embedded within individual's body-worn and personal mobile device such as smartphone or smart tag in neck. For example, the determination of micro-activity can be made using the 3-axis accelerometer (e.g., both postural [6] and gestural [10], [12]), ubiquitously available in modern smartphones or wearable devices. The remaining elements of each tuple are, however 'hidden'; neither the user's location nor macro activities are directly revealed by the smart devices' accelerometer data. While we employed our previously proposed techniques [4] to infer these hidden attributes, our main focus in this work lies on exploiting the constraints and correlation miner to handle the exponential evolution of the state spaces across multiple inhabitants.

B. Generic Model

To define a generic model, we assume two inhabitants are occupying a smart environment, and based on that we propose the following types of correlation and constraint relationships (as shown in Fig 3(a)).



(a) (b)

Fig. 3. (a) Context correlations and constraints (b) Single Inhabitant HDBN

Proposition 1. *Intra-user spatiotemporal correlations: For a specific user i, micro context state $context^i_{j=1}(t)$* =(running, livingroom) *infers that macro context state is $context^i_{j=2}(t)$* =(jogging, livingroom) *which delineates an intra-user spatiotemporal correlations.*

Proposition 2. *Inter-user spatiotemporal correlations: For a specific user i, micro context state $context^i_{j=1}(t)$* = (sitting, bathroom)*, infers that micro context state $context^i_{j=1}(t)$ for another user must not be* (sitting, bathroom) *as both the users must not be in 'sitting' state in a single bathroom concurrently. Similarly, for user i, if macro context state is $context^i_{j=1}(t)$=* (vacuuming, bedroom)*, then for another user macro context state $context^i_{j=2}(t)$ must not be* (sleeping, bedroom) *as in general people cannot sleep while anyone is vacuuming at the same room.*

Proposition 3. *Intra-user spatiotemporal constraints: For a specific user i, if macro context state is $context^i_{j=1}(t-1)$ =* (dining, livingroom)*, then in the immediate next time stamp, macro context state $context^i_{j=2}(t)$ may not be* (jogging, livingroom) *as it is unusual for someone to start jogging right after the dinner activity.*

Proposition 4. *Inter-user spatiotemporal constraints: For a specific user i, if macro context state is $context^i_{j=2}(t)$ =*(dining, livingroom)*, then, for another user macro context state $context^j_{j=2}(t)$ more likely be* (dining, livingroom) *as it is usual for the inhabitants to have dinner together.*

C. Single Inhabitant Model

We first represent the joint probability distributions of a single inhabitant activity model as follows.

$$p(o^i|context^i_j) = \prod_{t=1}^{N} p(context^i_j(t)| \\ pa(context^i_j(t))) \times p(o^i|context^i_j(t)) \quad (1)$$

where o^i represents observable streams, $pa(context^i_j(t)$ represents antecedents of $context^i_j$ at time t and model parameters are grouped into $\mathcal{D}$ levels i.e., $j = 1, 2, \ldots, \mathcal{D}$.

D. Multi-Inhabitant Model

Considering N users, we generate a N-chain CHDBN model where each chain is associated with a distinct user. Then, we simplify N-chain couplings by considering two users and two levels of hierarchies and represents the posterior of CHDBN for any user as follows.

$$p(c_j^{(n)}|o) = \prod_{(j)}^{2} \prod_{(n)}^{2} \pi_{c_1^{nj}} p_{c_1^{nj}}(o_1^{(nj)}) \\ \prod_{t=2}^{T} p_{c_t^{(nj)}}(o_t^{(nj)}) \prod_{(m)}^{2} p_{c_t^{(mj)}|pa(c_{t-1}^{(mj)})} / p(o) \quad (2)$$

where different users are indexed by the superscript. $p_{c_t^{(nj)}}(o_t^{(nj)})$ represents the emission probability given a state in chain n at level j, $p_{c_t^{(mj)}|pa(c_{t-1}^{(mj)})}$ represents the transition probability of a state in chain m at level j given its parent state, and $\pi_{c_1^{nj}}$ represents the initial state probability.

Further expansion of Eqn. 2, we have initial probabilities ($\pi_{c_1^{11}}$, $\pi_{c_1^{12}}, \pi_{c_1^{21}}$ and $\pi_{c_1^{22}}$), intra-user temporal state transition probabilities ($p_{c_t^{(1j)}|pa(c_{t-1}^{(1j)})}$ and $p_{c_t^{(2j)}|pa(c_{t-1}^{(2j)})}$), intra-user spatial state transition probabilities ($p_{c_t^{(1j)}|pa(c_t^{(1j)})}$ and $p_{c_t^{(2j)}|pa(c_t^{(2j)})}$), inter-user temporal state transition probabilities ($p_{c_t^{(1j)}|pa(c_{t-1}^{(2j)})}$ and $p_{c_t^{(2j)}|pa(c_{t-1}^{(1j)})}$), inter-user spatial state transition probabilities ($p_{c_t^{(1j)}|pa(c_t^{(2j)})}$ and $p_{c_t^{(2j)}|pa(c_t^{(1j)})}$) and emission probabilities of the states ($p_{c_t^{(1)}}(o_t^{(1)})$ and $p_{c_t^{(2)}}(o_t^{(2)})$) at j^{th}-level respectively for user $i = 1$ and 2. We denote the spatiotemporal probabilities for two users as π_{11}, π_{12}, π_{21}, π_{22}, p^1_{11}, p^1_{12},p^1_{21}, p^1_{22}, p^2_{11}, p^2_{12}, p^2_{21}, p^2_{22}, A_{11} and A_{22} respectively, then the posterior can be represented as follows,

$$p(c|o) = \frac{\pi_{11}\pi_{12}\pi_{21}\pi_{22}A_{11}A_{22}}{p(o)} \prod_{t=2}^{T} p^1_{11}p^1_{12}p^1_{21}p^1_{22}p^2_{11}p^2_{12}p^2_{21}p^2_{22}$$

V. Designing Context Miners

We now discuss how our proposed loosely-coupled HDBN model can infer hidden micro and macro context states by exploiting specifically the context correlations within and among multiple inhabitants.

A. Association Rules

We use rule mining techniques to generate rules governing spatiotemporal relationships among micro and macro context attributes. Our rules have the following general form: $\langle c_1, c_2, ..., c_n \Rightarrow \mathcal{R}\rangle$ which implies that $\mathcal{R}$ holds whenever all the tuples $\langle c_1, c_2, ..., c_n\rangle$ hold, $\mathcal{R}$ is true. For example, the rule $\langle$*walking=True; talking=True; kitchen=True ⇒cooking=True*$\rangle$ implies that if a user is in the *location* 'kitchen' with *postural* context state 'walking' and *gestural* context state 'talking', then he is in macro context state 'cooking'. Apriori algorithm is used to identify such rules, with a threshold $\mathcal{T}$ defined, such that we identify the sets of micro-context states which are subsets of at least $\mathcal{T}$ macro-activities. Each association rule has a support and a confidence. For example, if a state space model has 1000 context states, out of which 200 include both context states A and B and 80 of these include context state c, the association rule $A, B \Rightarrow c$ has a support of 8% (= 80/1000) and a confidence of 40% (= 80/200). The algorithm takes two input parameters: *minSup* and *minConf* to generate all the rules with *support* ≥ *minSup* and *confidence* ≥ *minConf*. We assume *minConf = 99%* and *minSup = 4%* which help strike good balance between tolerating occasional inconsistencies and highlighting the viable rules for our hierarchical context state space model (similar to [1]). Towards this rule generation, we consider each context tuple consist of 94 context elements (47 for current time t and 47 for the previous time instant $t-1$). At each instant, we have 11 high-level activities, as well as, for each user, 14 location contexts, 5 gestural contexts and 6 postural contexts, resulting in 47 total different context states.

B. Correlation Mining

We then use these mined rules to eliminate various infeasible state combination from the HDBN, considering relationships between micro and macro-activities. (in contrast, past work [1], [4] has considered such correlations only among micro-activities.) To illustrate our approach, consider the state-trellis for two users, A and B. A is assumed to have 3 possible values for its micro tuple (i.e., postural, gestural, location) and 3 possible values for its macro tuple (i.e., high-level activity, location) at each time instant, whereas B is assumed to have 4 such values for his micro tuple and 4 possible values for its macro tuple. Now, assume that A's postural activity (inferred from the smartphone accelerometer) is "sitting", gestural activity (inferred from the smart tag accelerometer in neck) is "silent" at time $t-1$ while B's postural activity is "walking". Furthermore, we observe that the living room infrastructure sensor was activated at time stamp $t-1$, indicating that the living room was occupied at $t-1$ by both A and B. Based on spatial rule $\langle jogging \vee watchingTV \Rightarrow livingroom\rangle$ generated from rule mining phase suggests that only 'jogging' or 'watching TV' macro activities are associated with A and B at the same time in location livingroom. Subsequently, we apply the rules $\langle WatchingTV \Rightarrow \{sitting, livingroom\}\rangle$ and $\langle Jogging \Rightarrow \{walking, livingroom\} \vee \{running, living room\}\rangle$ where 'sitting' is associated with 'watchingTV' and 'walking' is associated with 'jogging'. So, we can infer A is in "watchingTV" macro state, which helps to reduce the possible state space to 4 distinct combinations: ($A \Rightarrow \{sitting, livingroom\}, \{watchingTV, livingroom\}$, $B \Rightarrow \{walking, livingroom\}, \{jogging, livingroom\}$).

C. Constraint Mining

After appropriate elimination of infeasible states, we have horizontal and vertical correlated state transitions. We apply constraint mining technique to significantly reduce overhead incorporating a new binary-valued variable on vertical state transitions, end of sequence marker E. At a given time each activity level has two particular variables $c_t^{(d)}$ and $E_t^{(d)}$ as shown in Fig 3(b) which control the hierarchical multi-level activity structure. At time t, the variable $c_t^{(d)}$ represents the micro or macro context state and the binary-valued variable $E_t^{(d)}$ represents the continuation or termination point of the d^{th} level activity. To maintain consistency across our proposed multi-level HDBN activity model our constraint miner defines the following two constraints.

Blocking constraint: For user i, context state of the d^{th} level cannot change until the $(d+1)^{th}$ context state level has been terminated which is represented by:

$$E_{t-1}^{(d+1)} = 0 \rightarrow c_t^{(d)} = c_{t-1}^{(d)} \ \ \forall d < D \tag{3}$$

TABLE II. DEFINITION OF SEMANTICS

Notation	Description
c_t^{lMa}, c_t^{lMi}	Macro & micro activity at l^{th}-chain at t
O_t^{lMi}	Observed micro activity at l^{th} chain at time t
n^{Ma}, n^{Mi}	Number of macro and micro states respectively
E^{lMa},E^{lMi}	End of sequence marker (macro and micro)
$\mathbf{S^N}$=$S^{1N}_{1:l_1}, .., S^{LN}_{1:l_{\mid\mathbf{S}\mid}}$	A set of L training sequences for N users.
$\mathbf{S_t^i}$=(c_t^{Ma}, o_t^{Mi})	hidden and observed variables respectively

Termination constraint: The d^{th} context state level may not terminate until the $(d+1)^{th}$ level context state has been terminated which is represented by ($\rightarrow$ represents dependency):

$$E_{t-1}^{(d+1)} = 0 \rightarrow E_t^{(d)} = 0 \ \ \forall d < D \tag{4}$$

We propose to represent these constraints using a DBN model by adding the underlying constraints and correlations dependencies and storing their values in the conditional probability tables (CPT). The multi-level activity dependencies are specified by the following edges as shown in Fig 3(b).

$$c_{t-1}^{(d)} \rightarrow c_t^{(d)}, E_t^{(d+1)} \rightarrow E_t^{(d)}, E_{t-1}^{(d+1)} \rightarrow c_t^{(d)} \tag{5}$$

The blocking and termination constraints are enforced with the following constraints on CPT values $\forall d < D$.

$$\begin{aligned} p(c_t^{(d)} = j | c_{t-1}^{(d)} = i, E_{t-1}^{(d+1)} = 0, pa^*(c_t^{(d)} = .) = \delta_{i,j} \\ \& \ p(E_t^{(d)} = 0 | E_t^{(d+1)} = 0, pa^*(E_t^{(d)}) = .) = 1 \end{aligned} \tag{6}$$

where $\delta_{i,j} = 1$, if $i = j$ and 0 otherwise. (.) represents a probabilistic value and $pa^*(c_t^{(d)})$ denotes rest of macro/ micro states associated with the macro activity state $c_t^{(d)}$ which has not been already appeared in the list of conditioning variables.

VI. DESIGNING MULTI-LEVEL HIERARCHICAL MODEL

We consider smart phone sensors based postural and smart neck tag based oral gestural micro activities along with ambient sensor-based location contexts as observed states to infer the hidden macro state. We build N-chained coupled 'bare bone' HDBN for multiple inhabitants and augment our proposed constraint miner with the base DBN model. We list the semantics of every level of each chained coupled-HDBN model as shown in Table II and illustrated in Fig. 4 (observed states are omitted). Our constraint miner based augmentations and the conditional probability distributions are described as follows.

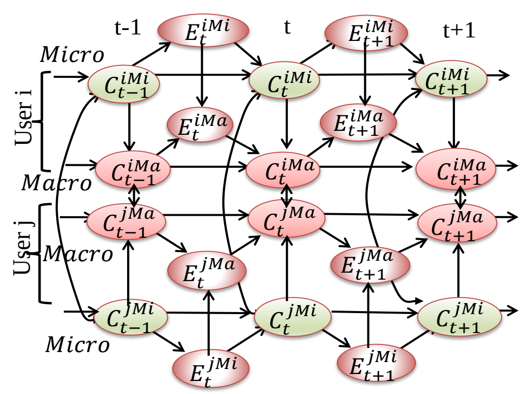


Fig. 4. Multi-Inhabitant loosely-coupled HDBN Model

Augmentation 1. *End of macro-level sequence dependency on macro-level context state and end of sequence marker of micro-level state.*

For example, end of macro state 'exercising' can be happened if end of 'cycling' micro activity has been found. Mathematically, for each macro-level of user l, we add the dependencies $c_t^{lMa} \rightarrow E_t^{lMa}$ and $E_t^{lMi} \rightarrow E_t^{lMa}$, such that the end-of-sequence marker jointly depends on the state of current level and lower level sequence marker i.e., end of macro-level sequence marker E_t^{lMa} jointly depends on macro state (c_t^{lMa}) and micro state (c_t^{lMi}) at time t. This completes the dependencies for the end of the macro-level sequence, E_t^{lMa} giving the CPT for $l = 1$ user as follows:

$$p(E_t^{1Ma} = 1 | X_t^{1Ma} = i, E_t^{1Mi} = 1) = p_{i=end}^{1Ma} \tag{7}$$

The parameter $p_{i=end}^{1Ma}$ represents the probability of ending a macro-level sequence given that we are at macro-level state i and all of the previous micro-level states have ended (end-of-sequence marker of micro-level state $E_t^{1Mi} = 1$). This parameter is used as a transition probability during the training phase.

Augmentation 2. *Each level of state dependency on prior distribution and lower level context state (if any).*

For example, macro activity state 'sleeping' can stay on 'sleeping' for next time stamp if the micro activity 'lying' in the 'bedroom' stays still. For micro activity state 'standing' can be happened if prior micro activity is 'sitting' or 'walking' as from 'lying' to 'standing', there must be an intermediate micro activity state 'sitting'. Mathematically, for each macro-level of user $l = 1$ and $l = 2$, we add the dependencies $c_t^{1Mi} \rightarrow c_t^{1Ma}$ and $c_t^{1Ma} \rightarrow c_t^{2Ma}$ such that the macro-level state jointly depends on the micro-level state and coupled user's macro-level state. This completes the dependencies for the macro-level states, c_t^{lMa} giving the following prior probabilities for user $l = 1$ and user $l = 2$:

$$p(c_t^{1Mi} = j) = \pi_j^{1Mi} \& p(c_t^{2Mi} = j) = \pi_j^{2Mi} \tag{8}$$

$$p(c_t^{1Ma} = i | c_t^{1Mi} = j) = \pi_i^{Ma,j} \tag{9}$$

$$P(c_t^{2Ma} = i | c_t^{2Mi} = j, c_t^{1Ma} = k) = \pi_i^{Ma,j,k} \tag{10}$$

The parameters π_j^{1Mi} and π_j^{2Mi} represent respectively the prior probabilities of micro-level states. The parameters $\pi_j^{Ma,j}$ represent the prior probabilities of macro-level state given that the user is at macro-level state i, micro-level state j. $\pi_i^{Ma,j,k}$ represents the prior probability of macro-level state given that the user is at macro-level state i, micro-level state j and coupled user's macro-level state k. These parameters are initialized with probabilistic values during the training phase.

Augmentation 3. *Micro-level and macro-level context states transition probabilities dependencies on their end-of-sequence marker and context state of coupled user.*

For example, 'dinning' macro context state occurs only when the coupled user's macro context state is 'dinning'. Similarly, for first user's location micro context state 'bathroom' cannot be occurred if coupled user's location context state is 'bathroom'. Mathematically, we can place the edges $c_{t-1}^{1M} \rightarrow c_t^{1M}$, $E_{t-1}^{1M} \rightarrow c_t^{1M}$, $c_t^{2M} \rightarrow c_t^{1M}$ to the model such that transition probabilities are different if we just ended a sequence. Here M is defined as macro or micro-level state. It defines if a sequence has ended then the probability distribution for starting a new sequence should be a prior distribution instead of a normal transition distribution. This gives the following transition model parameters for the macro-level (CPT).

Loosely-coupled HDBN (input: n^{Ma}, n^{Mi}, S; output: Most Likely State Sequence $\mathcal{M}$)
1.Procedure Model_Creation($c_t^{1Ma}, o_t^{1Mi}, c_t^{2Ma}, o_t^{2Mi}$)
//Create CHDBN model coupling two hierarchical DBNs
//E_t^{1Ma}, E_t^{1Mi}, E_t^{2Ma} and E_t^{2Mi} as end-of-sequence
//markers
$\mathcal{M} \leftarrow CreateCHDBNModel(n^{Ma}, n^{Mi}, \mathbf{S^N})$
2.Procedure Set_Conditions()
//Set conditional probabilities for EM algorithm
Initial Probabilities: $\pi_j^{1Mi}, \pi_j^{2Mi}, \pi_i^{Ma,j}, \pi_i^{Ma,j,k}$
Transition Probabilities: $a_{i\rightarrow j}^{1Ma}, \pi_{i\rightarrow j}^{1Ma}, a_{i\rightarrow j}^{1Mi}, \pi_{i\rightarrow j}^{1Ma}$
End of Sequence Probabilities:$p_{i=end}^{1Ma}, p_{j=end}^{1Mi,k}$
3.Run EM algorithm until convergence
$\mathcal{M} \leftarrow LearnParamsEM(\mathcal{M}_0, S)$
return $\mathcal{M}$

Fig. 5. A CHDBN based Multi-level Multi-inhabitant Activity Recognition

$$p(c_t^{1Ma} = j | c_{t-1}^{1Ma} = i, E_{t-1}^{1Ma} = 0, E_{t-1}^{1Mi} = 1, c_t^{2Ma}, c_t^{1Mi}) = a_{i\rightarrow j}^{1Ma} \tag{11}$$

$$p(c_t^{1Ma} = j | c_{t-1}^{1Ma} = i, E_{t-1}^{1Ma} = 1, E_{t-1}^{1Mi} = 1, c_t^{2Ma}, c_t^{1Mi}) = \pi_{i\rightarrow j}^{1Ma} \tag{12}$$

The micro-level CPT is similar, however it does not depend on the macro-level context.

$$p(c_t^{1Mi} = j | c_{t-1}^{1Mi} = i, E_{t-1}^{1Ma} = 0, E_{t-1}^{1Mi} = 1, c_t^{2Mi}) = a_{i\rightarrow j}^{1Mi} \tag{13}$$

$$p(c_t^{1Mi} = j | c_{t-1}^{1Mi} = i, E_{t-1}^{1Ma} = 1, E_{t-1}^{1Mi} = 1, c_t^{2Mi}) = \pi_{i\rightarrow j}^{1Ma} \tag{14}$$

At macro-level context k, $a_{i\rightarrow j}^{1Ma}$ is the transition probability of going from state i to state j and $\pi_{i\rightarrow j}^{1Ma}$ is the prior probability of starting a sequence in state j.

Augmentation 4. *Observations are assumed as multivariate Gaussian distributions and are based on the low-level micro context state (if any).*

Consider O_t^{Mi} represents an observation vector of micro-level context states where observations are continuous valued feature-vector, o_t^{Mi}. We assume that a particular observation o_t^{Mi}, is drawn from a Gaussian distribution, whose parameters are determined by the micro-level state, i.e.,

$$p(O_t^{Mi} = o_t^{Mi} | c_t^{Mi} = k) = \mathcal{N}(o_t^{Mi}; \vec{\mu_k}; \vec{\Gamma_k}) \tag{15}$$

where $\vec{\mu_k}$ is the mean vector and $\vec{\Gamma_k}$ is the covariance matrix parameters to the multivariate Gaussian distribution of observations for micro-level context k. To select and estimate these parameters, we use deterministic annealing clustering [8]. We first take a large set of feature points from our training data, run the deterministic annealing clustering algorithm to find several representative points, represent each cluster center as a low level state from which we can learn transition probabilities, and finally we use the cluster results to estimate our observation likelihoods by fitting a Gaussian to each cluster. Fig 5 shows our loosely-coupled HDBN algorithm.

Fig. 6. (a) PogoPlug Smart Home Customized Devices (from left) PIR sensors, object sensors, Ethernet Tag Manager, PogoPlug Mobile, Simplelink SensorTag used in the neck (top left corner) (b) Base application

TABLE III. ACTIVITIES OF DAILY LIVING

Activity(#)	Description
Macro(11)	1) Exercising, 2) Prepare Clothes, 3) Dining, 4) Watching TV, 5) Prepare Food, 6) Studying, 7) Sleeping, 8) Bathrooming, 9) Cooking, 10) Past Times, 11) Random
Oral gestural(5)	silent, talking, eating, yawning
Postural(5)	walking, standing, sitting, cycling, lying
Sub-location (SR1-SR14)	Area of exercise bike, couch 1, couch 2, dining table, bed, closet 1, reading table, closet 2, bathroom, kitchen, porch, rest of livingroom, corridor, rest of bedroom

VII. EXPERIMENTAL STUDY AND RESULTS

In this section, we first describe our smart home system, activity representations, data collection methodology and experimental setup, then present and discuss the evaluation results obtained from a series of experiments.

A. Smart Home Setup

We develop a real testbed smart home system, PogoPlug, as shown in Fig 7. PogoPlug consists of customized Cloud Engine PogoPlug Mobile [19] base server, 14 wireless sensor tag (WST) sensors (8 object sensors and 6 PIR sensors), 9 iBeacons, one Ethernet tag manager, and one router. PogoPlug base server is placed in a corner of living room with a continuous power supply which is associated with Ethernet tag manager and a router. We also place 6 PIR sensors and 8 object sensors and set the object sensors sensitivity with 55% (best choice tested on trial and error basis) thus slightest vibration on the object associated sensor fires without false alarm. We divide our entire smart home into 14 sub-region to evaluate our spatiotemporal constraints correlation model. Each participant is given a LG Nexus 4 Android phone installed with required apps should be placed in their pocket and a Simplelink Sensor Tag on their appropriate neck position with sticker. We set up three IP cameras in three appropriate places to exclusively collect the ground truth of all concerned activities. Fig 6(a) shows PogoPlug smart home customized devices and the Simplelink SensorTag worn on participant's neck.

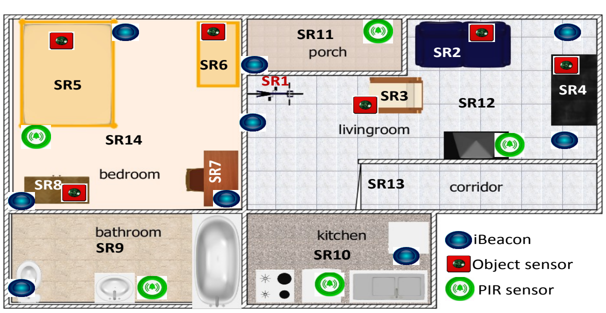


Fig. 7. PogoPlug smart one bedroom apartment testbed with 5 PIR sensors and 9 iBeacons creating 14 sub-regions numbered as SR1–SR14

B. Data Collection Methodology

PogoPlug smart home system is easy to deploy in multiple homes and helps provide and record data with minimal intrusiveness. We describe briefly the several applications with our PogoPlug smart home system. **wireless Sensor Tag application** which is a PHP webservice in Microsoft Azure server that receives Ethernet tag manager (connects all WST devices) provided WST sensor values and broadcasts to our Cloud Engine instantly with proper time-stamp and frequencies. **Simplelink SensorTag application** is an Android app that integrates Simplelink SDK and smartphone for continuous data sensing and controls sampling rate. **iBeacon sensing application** integrates Estimote Android SDK that gives distance measure (in meter) between smartphone and the Beacon. **Base Application** provides a user-friendly Android application interface to generate initial correlation rules. In order to define the semantic correlation rules through our user-friendly application interface, user needs to bring the smartphone at the center of each pre-defined sub-regions touching the concerned object, select/add new sub-region, select correlated low- and high-level activities, optionally set length of regions and click 'Set' button to confirm definition (Fig 6(b)). **Scheduler app** controls the IP camera recordings with pre-scheduled time.

We recruit 10 volunteers in five apartments (5 pairs) equipped with PogoPlug smart home. We choose 10 common activities (usually performed in morning time) to be performed in any order as shown in Table III. The rest of the recorded activities are counted as 'random' activity. Participants are allowed to turn on/off the camera any moment simply unplugging the power supply. Our scheduler app is integrated with wireless IP camera which can continuously record the video while it is turned on. We use trilateration approach on iBeacon sensing app provided distance measure to detect whether the carried smartphone is inside the smart home or not (multiple occupancy detection). We set an alarm at 8.00 AM morning. After 10 minutes of the alarm, our scheduling app automatically turn on the cameras recording. To collect as naturalistic data as possible, we encourage the inhabitants follow their normal activities of daily livings as listed in Table III. If participants do any other activities in the middle, those are counted as 'random' activity. We also consider the interleaved activities as 'random' activity i.e., getting up from bed to entering kitchen, the entire transition period is counted as 'random' activity. We collected data over one month of period for each home with two inhabitants (on average 2 hours of data with video per day) and recruited two graduate students to label all contexts (macro-, gestural-, postural-level activities, room and sub-region level occupancy) using the video camera and users' macro-activity log. The participant themselves validate the macro-activity label given the videos. We do not put any cameras in bathroom to avoid any privacy violations.

C. Datasets

We run the experiments on publicly available dataset CASAS and the data traces collected by us using our PogoPlug smart home system.

CASAS dataset: We first use the CASAS smart home [9] dataset, which consists of multi-resident ADLs obtained from a smart home environment. The dataset consists of a total of 26 different user pairs, from an overall sample size of 40 users. Note that this dataset contains sensor readings from a large number of instrumented sensors and smartphone sensor readings (no gestural activities). To mimic our assumed environment, we consider each motion sensor firing means the sub-location is occupied that is covered by motion sensor range.

CACE dataset: We collect over one month of data from our 5 smart homes for 10 users (each home is inhabited by two users). We maintain a consistent sampling frequency (50 Hz) for smartphone and Simplelink SensorTag for all users. This dataset is involved with 5 gestural, 5 postural, 14 sub-locations, 6 users and 5 smart homes.

D. Micro-level Activity Representation

To recognize micro-level (gestural and postural) activity recognition, we use microelectromechanical (MEMS) sensors, Simplelink Sensor Tag [18]. We calculate the 3D orientation of the device in the form of quaternion and calculate 3-axis acceleration trajectory using 9-axis inertial measurement units (3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer). We first apply high-band pass filter for both of the IMUs (neck mounted Simplelink SensorTag and smartphone integrated) and compute the acceleration trajectories [15] based on quaternion computation. We then represent 9-axis IMUs into rotation matrix form of quaternion. We write a quaternion q as, $q = q_s + q_x\hat{i} + q_y\hat{j} + q_z\hat{k}$ where $\hat{i}$, $\hat{j}$ and $\hat{k}$ are imaginary elements, each of which squares to 1. A quaternion q is a unit quaternion having magnitude $|q|$ given by $\sqrt{p(q_s^2 + q_x^2 + q_y^2 + q_z^2)} = 1$. We consider neck mounted IMU as a reference point while operating smartphone (if available in pocket) based relative trajectory computation. The position of the smartphone w at time t with respect to the neck mounted IMU's frame of reference F, is computed as follows.

$$w = q_t.w_0.q_t^-1 \tag{16}$$

where $w_0 = 0.\hat{i} + 1.\hat{j} + 0.\hat{k}$ represents the position of the smartphone in the local coordinates based on the assumption that the length from neck mounted sensortag to smartphone has unit length.

E. Activity Classification

In this section, we show the results obtained from a sequence of experiments. We also incorporate the activity detection measure using the best interval approach [20] to investigate the time difference between the true end of an activity and the estimate generated by our inference model.

Micro-Level Activity Classification: We use CACE dataset for micro-level activity classification. We calculate the 3-axis absolute acceleration trajectory based on sensor fusion of 9-axis IMUs from neck-positioned Simplelink SensorTag. A total of 32 statistical features (e.g., mean, variance, standard deviation, maximum and minimum, magnitudes, Goertzel coefficients of 1-5 Hz etc.) are computed over each 1.5 seconds long frame (best segment achieved from trial and error) of the absolute acceleration trajectories. We then employ a change-point detection-based classification method towards feature extraction and random forest classification [12] which improved the accuracy significantly on test dataset (95.3% accuracy with a false-positive rate 1.8%) than the current work [10]. For postural activities, we consider neck position as a stationary joint and use this as a reference to calculate relative acceleration trajectory of pocket mounted smartphone. We calculate 3-axis smartphone acceleration trajectory (Eqn 16). We segment the data, extract features with 1.5 seconds framing window and 50% overlap and finally classify with a Random Forest based classifier that provides us an accuracy of $\approx$ 98.6% (with a false-positive rate of 0.6%).

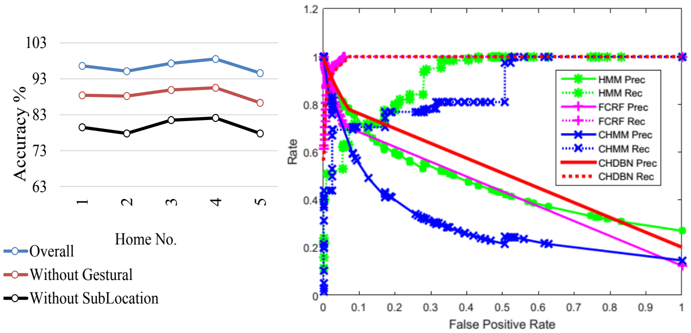
TABLE IV. SOME GENERATED RULES WITH CONFIDENCE

(t=time, U1=user 1, U2=user 2)
$U1(t) : (cycling \vee sitting) \wedge SR1 \Rightarrow U1(t) : exercising; (1)$
$U1(t) : (sitting \vee lying) \wedge SR5 \Rightarrow U1(t) : sleeping; (1)$
$U1(t) : SR9 \Rightarrow U2(t) : \neg SR9; (1)$
$U1(t) : SR4 \wedge U2(t) : SR4 \Rightarrow U1(t) : dining \wedge U2(t) : dining; (1)$

Macro-Level Activity Classification: We implement our loosely-coupled HDBN algorithm in Java using WEKA 3.7.11 ([11]) and evaluate our algorithm using PogoPlug Mobile with a 700 Hz CPU, 128 MB storage and 128 MB memory. We use EM and Viterbi algorithm for the inference of the most likely context state space.

In CACE dataset, at first we generate 58 unified rules on training dataset using intra-inter spatiotemporal correlations. Table IV shows some generated top rules which represent the real world scenario. For example, if individual's postural activity becomes 'cycling' or 'sitting' in the exercise bike area (SR1), in reality, it is more obvious that the person is 'exercising'. We then employ these rules to prune the state space at runtime. We find that our coupled HDBN approach (including micro-level oral gestures, postures and sub-region contexts) results in a macro-level activity recognition accuracy of $\approx$ 95.1% (FP rate 1.5%, precision 97.3%, recall 95.1%, weighted ROC 97.7% and PRC 98.8%). More interestingly, our system achieves on average 99.7% accuracy on shared activities (such as, sleeping, dining, past-times etc.), thus illustrating how *inter-user behavioral correlation can be an asset in activity recognition.* The least performed macro-activities are cooking, preparing food, preparing clothes etc. Fig 8(b) shows the details of macro-level activity accuracy measure including false positive rate, precision, recall and F-measure.

In CASAS dataset (no oral-gestural activity), we use the association rule mining technique to generate optimized rules for intra- and inter-user spatiotemporal correlations; redundant (e.g., transitive) rules were subsequently merge to eventually obtain 47. We employ these rules to prune the state space at runtime. Overall, we find that our coupled HDBN approach results in a macro-level activity recognition accuracy of $\approx$ 94.5% (FP rate 1.4%, precision 96.5%, recall 94.5% and weighted ROC 98.6%). Our system achieves 99.3% accuracy on shared activities (such as, Move Furniture and Play Checker). Fig 9 shows details result of classification on CASAS data.



(a) (b)

Fig. 8. (a) Overall accuracy, without gestural and without location context performance for each home (5 smart home consists with one pair of inhabitants each) (b) Precision & Recall versus False Positive rate, while adjusting the cost function of the different classifier during training.

In contrast, when applied to CACE dataset, in the absence of oral gestural data (manually removed), the accuracy is only 89.7%; when even location (sub-region) context (i.e., the data from the ambient motion sensors) is removed, the recognition accuracy drops to 80.5%. Fig 8(a) complex activity recognition accuracies with or without the gestural level micro activities and location tuple for each apartment. By observing the increased activity recognition accuracy across both different activities and users, we conclude that the gestural micro activity and location are important context states and help improve multi-user complex activity recognition accuracy.

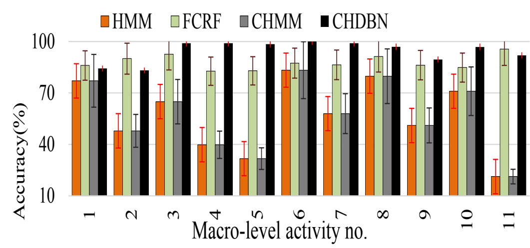
Macro	FP Rate	Preci-sion	Rec-all	F-Meas
1	0.3	97.1	85.3	93.2
2	0.1	100	93.5	96.5
3	0	99	95.1	97.3
4	3.5	53.7	81.4	60.7
5	1.8	55.5	84.5	59.3
6	0	100	93.6	95.5
7	0.7	99.7	87.6	94.5
8	0	99.9	91.3	97.6
9	0	99.9	84.7	94.3
10	0.1	100	81.3	92.3
11	0	100	95.1	98
12	0.1	97.1	90.8	95.5
13	0.1	97.8	91.7	97.9
14	3.4	49.5	77.8	55.8
15	2.1	69.1	79.3	61.5
Overall	1.4	96.5	94.5	93.6

Fig. 9. CASAS Dataset Macro activity classification FP rate, precision, recall, F-measure

F. Comparison with Prior Methods

We next compare our loosely-coupled HDBN approach (including the state-space pruning technique based on mining of correlations and constraints) with three previously-proposed ADL approaches.

(i) the HMM [9] model which has been applied for activity recognition in a multi-resident setting built an individual HMM model for each user; (ii) the FCRF [5] approach dealt with wearable sensor data to exploit the temporal constraints across two users; (iii) the CHMM [4] approach consisted with ambient and postural data to exploit spatiotemporal constraints across multiple users. Note that all of the models (FCRF, CHMM and CHDBN) cause an exponential increase in the number of state spaces for multiple users; however, our intelligent use of mined constraints and correlations help reduce the state space by an order of magnitude (16-fold reduction compared to FCRF). Fig. 10 compares the performance of macro-level activity recognition accuracy of CHDBN, FCRF, CHMM and HMM, and depicts the mean accuracy error of macro activity recognition over a randomly chosen pair of users for all 11 activities. We note that our proposed model outperforms the recognition accuracy (with average accuracies that are 20% higher than the HMM, 8% higher than FCRF and 5% higher than CHMM approaches) for all the 11 complex activities. Note that particularly for four macro activities; namely *2. Prepare Clothes (B)*, *1. Exercising* and *9. Cooking* show high false positive rates with very low precision and F-measure. We observe that macro activity 1 and 2 are mostly happened in the sub-location 'SR1' and 'SR6', 84% and 85% respectively but in case of misclassified instances, they occur mostly in other locations such as 'SR8' or 'SR12'. Similarly, for activity 9 is mostly performed (84% each) in the 'kitchen' location while misclassified instances are mostly occurred in the location 'dining room' or 'living room'.



(a)

Macro	FP Rate	Preci -sion	Rec-all	F-Meas
1	0.1	99.1	97.4	97.8
2	2.2	90.5	89.6	93.5
3	0.05	99.6	98.9	99.1
4	0	100	100	100
5	1.6	92.5	92.5	91.7
6	0	100	100	95.2
7	0	97.5	100	100
8	0.02	100	96.3	96.9
9	1.67	91.5	90.5	93.7
10	0	95.3	100	100
11	0.15	97.1	95.9	95.7
Overall	1.5	97.3	95.1	96.8

(b)

Fig. 10. CACE Dataset (a) Macro activity accuracy comparison (b) FP, Precision, Recall, & F-Measure

G. Comparison with Different Pruning Strategies

We further investigate the benefits of our proposed approach towards pruning of the state space, compared to other plausible approaches. For comparison purposes, we consider CACE dataset with three performance metrics: (1) accuracy: which is defined as (true positive+true negative)/total instances; 2) computational overhead: total time required to build entire model; and (3) start/end duration error: this is defined as the overall error in the estimation of the duration of the activity. An example will clarify our definition. Consider that the true duration of "cooking" is 30 minutes (10:05 AM - 10:35 AM) and that our algorithm predicts 29 minutes (10.10 - to 10.39 AM). Then, the start/end duration error is 9 minutes (|5 minutes delayed start| + |4 minutes hastened end|), in an overall error of e.g., 30% (9/30=0.3).

TABLE V.
DURATION ERROR

Method	duration error
NH	16.9%
NCR	20.6%
NCS	7.72%
C2	8.1%

We consider 4 different approaches, each with varying levels of pruning. **Naive-HMM (NH):** This is the exhaustive strategy considering traditional-HMM with all possible states in the state space [9]. More specifically, this approach ignores the decomposition of a macro activity into micro activities and directly employs macro-activity classification where the classifier is trained with traditional features (5 postural and 6 gestural features) computed over individual frames, and directly labeled with the macro activity label. **Naive-Correlation (NCR):** This strategy is a two-fold approach, where we generate association rules using rule miner and prune the state space, but for each individual user separately (similar to the approach adopted in [1]). **Naive-Constraint (NCS):** This strategy is implemented with our proposed constraint miner (alone) incorporating the 4 augmentations in the CHDBN model. In this strategy, we exclude the correlation miner and employ the constraint miner towards CHDBN model for complex activity recognition. **Correlation-Constraint (C2):**

This approach is hybrid, that combines both the correlation and constraint miner together to perform the state space pruning before employing a loosely-coupled HDBN model for complex activity recognition.

Fig 11(a) and Fig 11(b) plot overall performance in terms of accuracy and computational overhead respectively of above 4 approaches on CACE dataset. We see that macro-activity recognition is quite poor if one ignores the cross-individual correlations: both NH and NCR report very poor performance in terms of classification accuracy (76.2% and 73% respectively) as well as computational overhead (4.95s and 1.5s respectively). In contrast, application of CHDBN concept shows significant improvement in accuracy (98%) costing a lot of computational overhead (15.96 secs. Our unified loosely-coupled HDBN model (C2), which considers correlation and constraint relations jointly among multi-level contexts, shows significant additional reduction (15.96/0.96 secs= 16 fold) in computational overhead. From the cost-accuracy performance graphs (Fig 11), we clearly see that C2 outperforms all other 3 methods. From Table V, we can firmly say that C2 outperforms other approaches in terms of start/end duration error too. Fig. 12 illustrates the incremental performance characteristics of our proposed hierarchical activity recognition model as the knowledge about the state space increases with the increase in data points. We note that i) overhead of building model increases exponentially with the increase in sample size segment (%); and ii) our model attains a modest activity recognition accuracy of 83% with a minimal 30% sample segment. However, initial rules provided by users improve both accuracy and overhead of our CACE framework (Fig. 12).

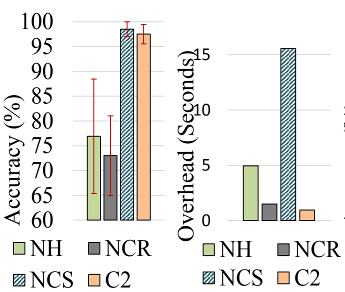


Fig. 11. Complex activity classification performance for NH, NCR, NCS and C2 strategy in terms of (a) accuracy and (b) overhead

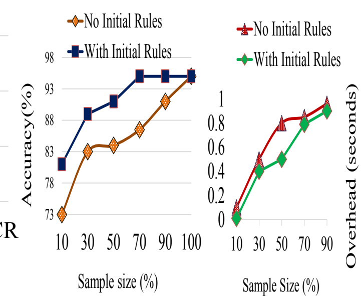


Fig. 12. Incremental complex activity classification performance in terms of accuracy and overhead

VIII. Conclusion

We focus on the problem of using a combination of on-body and ambient sensor data to identify the high-level activities of individuals in a multi-inhabitant smart home. We develop a loosely-coupled HDBN based model for such activity recognition, that explicitly factors in the correlations between the micro and macro activity states of multiple users. "State space explosion" is the key challenge in such a rich interaction model, we show how a rule mining algorithm, operating on multi-inhabitant training data, can provide deterministic correlation and probabilistic constraints that dramatically reduce the state space. Experimental studies with two real world trace-driven, dataset show that (i) the HDBN model offers max. 95% activity recognition accuracy, compared to 75% or lower from alternative approaches that do not exploit such inter-person context relationships; (ii) our state-space pruning method provides a 16-fold reduction in context overhead; and (iii) the use of oral gestural data helps improve the recognition accuracy of common daily activities by $\approx$ 20%. We argue that, CACE model can be used as a smoother of any online complex activity recognition framework as well. Our current experimental results have focused on a smart home facility with only a pair of inhabitants, but we believe that our generic CACE framework can handle 3-4 occupants as well. Though CACE works really well in cross-home and intra-home correlation scenario, adapting our recognition approach to environments with changing relationship patterns (e.g., if the joint dining habits of the inhabitants gets modified) remains an open challenge.

IX. Acknowledgments

This work is supported partially by the NSF Award #1344990, UMB-UMBC Research and Innovation Partnership Grant, and Constellation E2: Energy to Educate Grant.

References

[1] S. Nath, ACE: exploiting correlation for energy-efficient and continuous context sensing. MobiSys 2012.

[2] A. Parate, M. Chiu, D. Ganesan, B. M. Marlin, Leveraging graphical models to improve accuracy and reduce privacy risks of mobile sensing. MobiSys 2013.

[3] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, K. Aberer, Energy-Efficient Continuous Activity Recognition on Mobile Phones, An Activity-Adaptive Approach. ISWC 2012.

[4] N. Roy, A. Misra, D. Cook, Infrastructure-assisted smartphone-based ADL recognition in multi-inhabitant smart environments. PerCom 2013.

[5] L. Wang, T. Gu, X. Tao, H. Chen, J. Lu, Recognizing multi-user activities using wearable sensors in a smarthome. PMC 2011.

[6] J. Kwapisz, G. Weiss, and S. Moore, Activity recognition using cell phone accelerometers. SensorKDD-Workshop 2010.

[7] M. Schirmer, H. opfner, SENST*: approaches for reducing the energy consumption of smartphone-based context recognition. CONTEXT 2011.

[8] J. Muncaster, Y. Ma, Hierarchical Model-Based Activity Recognition With Automatic Low-Level State Discovery. Journal of Multimedia 2007.

[9] G. Singla, D. J. Cook, M. Schmitter-Edgecombe, Recognizing independent and joint activities among multiple residents in smart environments. JAIHC 2010.

[10] T. Rahman, A. T. Adams, Mi Zhang, E. Cherry, Bo. Zhou, H. Peng, T. Choudhury, BodyBeat, a mobile system for sensing non-speech body sounds. MobiSys 2014.

[11] I. Witten, E. Frank, Data Mining: Practicial Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann 1999.

[12] M. Alam, N. Roy, GeSmart: A Gestural Activity Recognition Model for Predicting Behavioral Health. Smartcomp 2014.

[13] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of Royal Statistics Society, Series B, 1977.

[14] M. A. Alam, N. Pathak, N. Roy. Mobeacon: An iBeacon-Assisted Smartphone-based Real Time Activity Recognition Framework, MobiQuitous 2015.

[15] A. Parate, M. Chiu, C. Chadowitz, D. Ganesan, E. Kalogerakis: RisQ: recognizing smoking gestures with inertial sensors on a wristband. MobiSys 2014.

[16] R. Chen, Y. Tong, A Two-stage Method for Solving Multi-resident Activity Recognition in Smart Environments. Entropy 16(4), 2014

[17] N. Lane, L. Pengyu, L. Zhou, F. Zhao, Connecting personal-scale sensing and networked community behavior to infer human activities. UbiComp 2014.

[18] http://www.ti.com/ww/en/wireless_connectivity/sensortag2015/

[19] pogoplug: https://pogoplug.com/devices

[20] E. M. Tapia, S. S. Intille, and K. Larson, Activity Recognition in the Home Using Simple and Ubiquitous Sensors, Pervasive, volume 3001 of Lecture Notes in Computer Science, Springer, 2004.