33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Activity Recognition using Correlated Pattern Mining for People with Dementia

Kelvin Sim, Clifton Phua, Ghim-Eng Yap, Jit Biswas, and Mounir Mokhtari

Abstract— **Due to the rapidly aging population around the world, senile dementia is growing into a prominent problem in many societies. To monitor the elderly dementia patients so as to assist them in carrying out their basic Activities of Daily Living (ADLs) independently, sensors are deployed in their homes. The sensors generate a stream of context information, i.e., snippets of the patient's current happenings, and pattern mining techniques can be applied to recognize the patient's activities based on these micro contexts. Most mining techniques aim to discover frequent patterns that correspond to certain activities. However, frequent patterns can be poor representations of activities. In this paper, instead of using frequent patterns, we propose using correlated patterns to represent activities. Using simulation data collected in a smart home testbed, our experimental results show that using correlated patterns rather than frequent ones improves the recognition performance by 35.5% on average.**

I. Introduction

Dementia is a serious cognitive disorder which affects the sufferer's memory, attention, language, and problem-solving abilities. Recently, several activity recognition systems proposed using *ambient intelligence* to assist dementia patients in carrying out ADLs independently in smart home environments [1], [2], [3], [4], [5], [6]. Ambient intelligence uses pervasive sensors that are sensitive to people's activities to detect the micro contexts or happenings in the homes. Using such information, the system intelligently detects the patient's activities and guides the patient in carrying out ADLs. The purpose is to assist the subject (i.e., the person with dementia) to be more independent at home and to reduce the workload of the caregivers. In such systems, incorporating multiple sensing modalities is important as this enables us to obtain a wide spectrum of micro contexts, which is crucial for capturing a more complete picture of the subject's activities.

II. Problem Motivation

Discovering patterns that accurately represent patient activities is non-trivial due to the complexity of micro contextual data. First, the wide spectrum of micro contexts obtained from the sensor readings means that the data can be extremely high-dimensional. Using traditional machine learning methods (e.g. clustering [7]) to mine high-dimensional data is infeasible due to the effect known as *curse of dimensionality*, i.e., the distinction between data points blurs as more and more micro contexts are considered by these methods. Second, the fact that the micro contextual data streams in every second or millisecond means that the data size can grow extremely large – efficient algorithms are necessary to mine the patterns. Due to the above characteristics of the data, many existing activity recognition systems [3], [4], [5], [6] use frequent pattern mining algorithms because such methods works reasonably well for large data with high-dimensionality.

TABLE I

An example of a transactional dataset of micro contexts

Time	Sensor Readings					Activity
	s_1	s_2	s_3	s_4	s_5	
t_1	1	1	4	1	3	(1) Preparing food
t_2	1	2	4	3	3	(2) Consuming food
t_3	1	2	4	3	3	(2) Consuming food
t_4	1	0	1	4	0	(3) Keeping Utensils
t_5	1	0	4	5	3	(4) Watching TV

In frequent pattern mining, patterns with higher occurrences are considered to be more significant [8] and usually pattern-based activity recognition systems use patterns with high occurrences to detect the activities. However, patterns with high occurrences may not be accurate representations of activities. Table I shows an example of a transactional dataset of micro contexts, where $s_i[j]$ represents sensor s_i with reading value j. $s_1[1]s_3[4]s_5[3]$ (shaded in light grey) is a frequent pattern since it occurs four out of five times in this transactional dataset. Although this pattern has high occurrences, its occurrences are across three activities, namely *preparing food*, *consuming food*, and *watching TV*. Hence, this pattern is not an accurate representation for any particular activity. Instead of using such patterns with high occurrences, using patterns which uniquely represent each activity would be more accurate.

In the case of frequent pattern mining algorithms, the user has to set a minimum support threshold, such that only those patterns that occur above this threshold are considered 'frequent' [8]. However, it is difficult to set the correct minimum support for useful patterns. Using the same transactional dataset shown in Table I and at a minimum support of four, $s_1[1]s_3[4]s_5[3]$ (shaded in light grey) is a frequent pattern as it occurs four times in the transactional dataset. As its occurrences is across three activities, it is not a useful representation of activities. On the other hand, pattern $s_2[2]s_4[3]$ (shaded in dark grey) only occurs twice and would not be considered frequent despite the fact that it is a perfect

K. Sim, C. Phua, G-E. Yap, and J. Biswas are with Institute for Infocomm Research, A*STAR, Singapore {shsim,cwcphua,geyap,biswas}@i2r.a-star.edu.sg
M. Mokhtari is with the CNRS-IPAL Singapore, Institut Telecom France Mounir.Mokhtari@it-sudparis.eu

representation for the *consuming food* activity.

III. Proposed Method

To overcome the weakness of frequent patterns, we propose using *correlated* patterns [9], [10] as representation of activities. We define correlated patterns as patterns that have higher occurrences *only in activities which they are correlated with*. Thus, correlated patterns are more accurate representation of activities than patterns which might be frequent across different activities. For example, pattern $s_2[2], s_4[3]$ (shaded in dark grey) is a closed correlated pattern in Table I. Firstly, the occurrences of this pattern is relatively high, as it occurs 2 out of 5 times in the table. Secondly, $s_2[2], s_4[3]$ only occurs in consuming food activity, and neither $s_2[2]$ nor $s_4[3]$ occur in other activities. Hence, the pattern $s_2[2], s_4[3]$ is a meaningful and useful representation of *consuming food*.

In the training phase, the activity recognition system mines correlated patterns from the training data consisting of micro contexts. Figure 1 shows the overview of the training phase with the correlated pattern mining algorithm MIP. The top table represents the training data (i.e., the micro contexts), the middle box is the correlated pattern mining algorithm, and the bottom table shows examples of correlated patterns and their corresponding activities.

A. Definition of Correlated Patterns

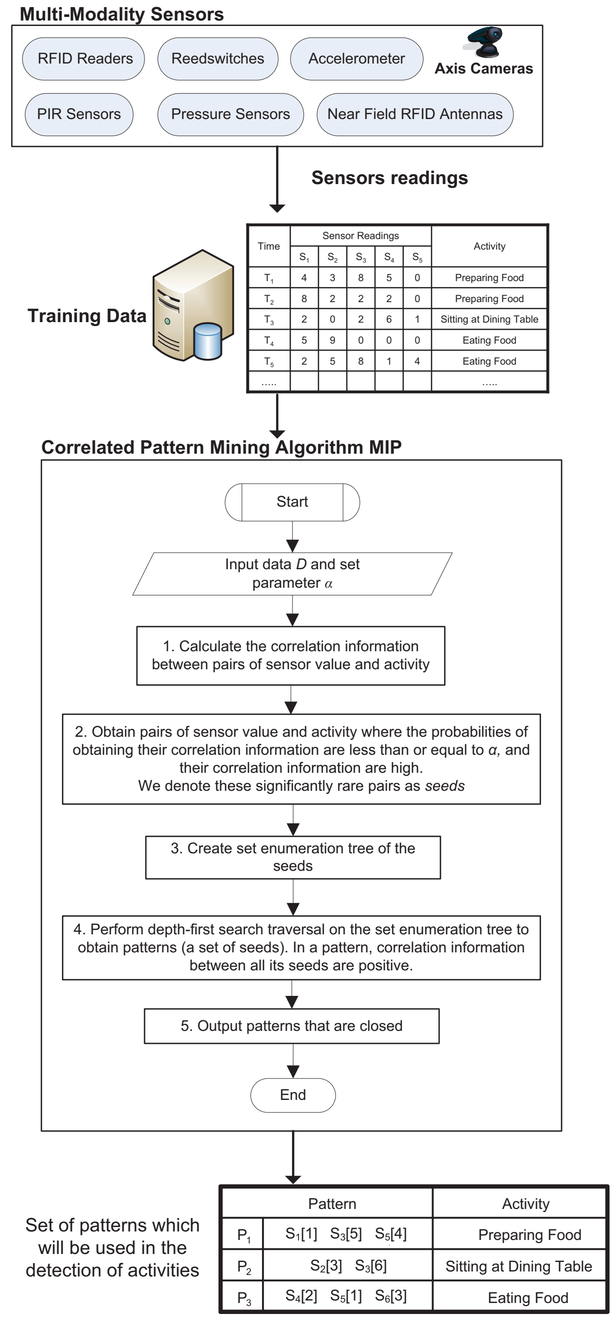
We give some preliminaries before giving the formal definition of correlated patterns. Let the transactional dataset be rows of sensor readings taken at different timestamps, and each row has a label indicating the activity taking place at the particular timestamp. We can represent such a transactional dataset as a **matrix** $\mathcal{D} = \mathcal{T} \times \mathcal{A}$ with timestamps $\mathcal{T}$ and attributes $\mathcal{A}$ as its dimensions. We denote attributes $a_1, \ldots, a_{|\mathcal{A}|-1}$ as the sensor readings and attribute $a_{|\mathcal{A}|}$ as activity label a_{label}. Note that we use integers to represent the activities in activity label a_{label}, e.g., *preparing food*, *consuming food*, *keeping utensils* and *watching TV* activities in Table I are represented as 1,2,3 and 4, respectively. We denote the value of attribute a as $v_a \in \mathbb{Z}$, and we also denote a value of attribute a, at time t, as v_{at}.

An attribute a_i can be considered as a random variable with probability mass function $p(v_{a_i}) = Pr\{a_i = v_{a_i}\} = \frac{occ(v_{a_i})}{|\mathcal{T}|}$, where $occ(v_{a_i})$ is the number of times value v_{a_i} occurs in matrix $\mathcal{D}$. The conditional probability of value v_{a_i} occurring in $\mathcal{D}$, given the occurrence of value v_{a_j} in $\mathcal{D}$, is $p(v_{a_i}|v_{a_j}) = Pr\{a_i = v_{a_i}|a_j = v_{a_j}\} = \frac{occ(v_{a_i}, v_{a_j})}{occ(v_{a_j})}$, where $occ(v_{a_i}, v_{a_j})$ is the number of times the values v_{a_i}, v_{a_j} occur together in $\mathcal{D}$.

Definition 1 (Pattern $P = T \times A$): Let $P = T \times A$ be a sub matrix, where $T \subseteq \mathcal{T}$ and $A \subseteq \mathcal{A}$. P is a pattern if the activity label $a_{label} \in A$ and $\forall a \in A : \forall t \in T : v_a = v_{at}$.

Table I shows an example of matrix $\mathcal{D}$, and $P = \{t_2, t_3\} \times \{a_2, a_4, a_{label}\}$ is a correlated pattern that is present in $\mathcal{D}$.

Definition 2 (Correlation information of a pattern): The correlation information of pattern $P = T \times S$ is measured by



Time	S_1	S_2	S_3	S_4	S_5	Activity
T_1	4	3	8	5	0	Preparing Food
T_2	8	2	2	2	0	Preparing Food
T_3	2	0	2	6	1	Sitting at Dining Table
T_4	5	9	0	0	0	Eating Food
T_5	2	5	8	1	4	Eating Food
.....						.....

	Pattern	Activity
P_1	$S_1[1]$ $S_3[5]$ $S_5[4]$	Preparing Food
P_2	$S_2[3]$ $S_3[6]$	Sitting at Dining Table
P_3	$S_4[2]$ $S_5[1]$ $S_6[3]$	Eating Food

Fig. 1. An overview of the training phase with the mining correlated patterns algorithm (MIP)

$$ci(P) = \sum_{i=1}^{|S|} p(v_{a_1}, \ldots, v_{label}) \log \frac{p(v_{a_i}, v_{label}|v_{a_1}, \ldots, v_{a_{i-1}})}{p(v_{a_i}|v_{a_1}, \ldots, v_{a_{i-1}})p(v_{label})}$$

The correlation information is based on the concept of information theory [11]. Details of the derivations of this measure can be found in [9]. The first part of the equation measures how *frequent* the attributes' values and the activity label occur together, and the second part of the equation measures how *correlated* the attributes' values and the activity label are, i.e. if the occurrence of the attributes' values and the activity label together are by chance. The second part of this equation is the main component that differentiates correlated patterns from frequent patterns; the gist of frequent patterns is in finding attributes' values and the activity label that occurs frequently together, but they do not consider if their occurrences are by chance.

Definition 3 (Correlated pattern): A pattern $P = T \times A$ is a correlated pattern if its correlated information $ci(P)$ is high.

A pattern P with high $ci(P)$ means that the attributes' values and the activity label occur frequently together and their occurrence are not by chance. We do not explicitly ask the user to set a threshold to determine how high the correlation information should be, as the user will not know what will be the correct threshold to set, which is the same situation of frequent patterns. Instead, we propose using a correlated pattern mining algorithm MIP, which is able to determine how high the correlation information $ci(P)$ should be to be considered as significant. Details of this algorithm is in the next section.

B. Algorithm for Mining Correlated Patterns (MIP)

We present algorithm MIP, which mines correlated patterns with significant correlation information in an efficient way.

Algorithm MIP consists of two main parts:

1) *Generating seeds.* From matrix $\mathcal{D}$, we find pairs of values (v_a, v_{label}) with significant correlation information, and denote these pairs as *seeds* (for growing patterns).
2) *Mining patterns.* The seeds are treated as initial patterns, and these are used to generate the correlated patterns. The correlated patterns are enumerated through a depth-first search using the set enumeration tree of the seeds.

Algorithm MPI is a variant of algorithm MIC [10], which is used to mine correlated clusters. The main difference between them is in their second part, as MIP mines patterns while MIC mines clusters.

The rest of this section describes the two main parts of algorithm MIP in details, with the middle box of Figure 1 as a reference to these details.

1) Generating seeds: Although we could use all possible pairs of values as seeds, the search space would be exponentially large and the computation cost would be overwhelming (the set enumeration tree approach is NP-hard [12]). Hence, this first part is necessary to reduce the exponential search space by filtering away all pairs of values with low correlation information.

For each pair of values (v_a, v_{label}) given $a \in \mathcal{A}$, we calculate its correlation information $ci(v_a, v_{label})$ (Figure 1 Middle box Step 1). For conciseness, we denote $ci(v_a, v_{label})$ as ci. Let the set of positive correlation information of pairs of values be denoted as $CI = \{ci | ci > 0\}$.

The significance of a seed can be determined by calculating the statistical probability of observing its correlation information. Setting the null hypothesis H_0 as "A sample ci is equal to the mean of CI", we let the probability

Algorithm 1 $growSeeds(P, seeds)$

Description:

for all $seed$ in $seeds$ **do**
    if v_{label} of $seed$ and P are the same **and** $ci(extend(P, seed)) > ci(P)$ **then**
        $P \leftarrow extend(P, seed)$;
        $growSeeds(P, seeds/seed)$;
    end if
end for
if P is not extended with any $seed$ **then**
    output P as a pattern;
end if

of observing ci by chance be *p-value*(ci). A very low *p-value*(ci) means that it is very rare to have this pair of values with such high correlation information, so a pair of values is a seed if its ci is statistically significant, i.e., *p-value*(ci) $\leq \alpha$, where α is a preset threshold (Figure 1 Middle box Step 2). In another sense, α determines how many pairs of values should be pruned using statistical principle, and [10] recommends a default setting of $1.0E-4$, which is shown to be insensitive to the results. Formally,

$$seeds = \{seed | seed = (v_a, v_{label}), p\text{-}value(ci) \leq \alpha, a \in \mathcal{A}\}$$

To derive the probability *p-value*(ci), we would first need to model the probability distribution of CI. Given that (1) the values in CI are continuous, (2) they are positive and (3) the probability distribution of CI is unknown and dependent on data $\mathcal{D}$, either the gamma or the Weibull distribution would be suitable. Both distributions offer the flexibility of modeling any continuous and positive probability distribution, as the scale and shape of the distribution can be adjusted by their two parameters [13]. We adopt the gamma distribution for its higher efficiency [14].

Let CI be gamma-distributed with parameters shape k and scale θ, i.e. $CI \sim \Gamma(k, \theta)$. The probability density function of the gamma distribution is $f(ci; k, \theta) = \frac{ci^{k-1}}{\Gamma(k)\theta^k} exp(-\frac{ci}{\theta})$, where $ci \in CI$ and $\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt$ is the gamma function.

After obtaining the parameters $\tilde{k}, \tilde{\theta}$ using maximum likelihood estimation [15], we calculate *p-value*(ci) using the cumulative distribution function (cdf) of the gamma distribution

$$p\text{-}value(ci) = \frac{1}{\Gamma(\tilde{k})\tilde{\theta}^{\tilde{k}}} \int_0^{ci} t^{\tilde{k}-1} e^{-t/\tilde{\theta}} dt$$

which is efficiently derived by the Newton-Raphson method [16].

2) Mining Patterns: Algorithm 1 describes our $growSeeds$ algorithm, which uses the generated seeds as building blocks for patterns. The general idea is that each $seed$ is considered as an initial pattern, and we try to 'grow' this initial pattern by extending it with other $seed$ in a depth-first search manner (Figure 1 Middle box Step 3).

The proposed $growSeeds$ algorithm proceeds as follows. Each $seed$ is first initialized to be a pattern $P = T \times A$.

A sub-function $extend$ extends a pattern P with $seed = (v_a, v_{label})$, such that P becomes $T \times \{A \cup a\}$, where $a \notin A$ and $v_a, v_{label} \in T \times \{A \cup a\}$. Pattern P will only be extended if: (1) the activity label v_{label} of P and the $seed$ are the same, and (2) the correlation information of the extended pattern is larger than the correlation information of P (Figure 1 Middle box Step 4). The first condition ensures that we are mining patterns that can accurately represent activities, while the second condition ensures that only useful attributes' values are being added to the pattern P to increase the correlation between the new pattern and the activity. If a pattern P is not extended with any $seed$, we output it as a pattern. After all patterns are mined, we remove redundancy by performing post-processing to retain only closed patterns (Figure 1 Middle box Step 5). A pattern $P = T \times A$ is closed if and only if there does not exist any pattern $P' = T' \times A'$ such that $T \subseteq T'$ and $A \subseteq A'$.

IV. EXPERIMENT

The pattern-based activity recognition system requires a training phase where patterns are mined from a training data of micro contexts. In this experiment, we mined both frequent patterns and our proposed correlated patterns from the training data, and evaluated their accuracy in activity recognition. We first present the experimental setup and then the experimental results.

A. Experimental Setup

The experiment was conducted in our smart home environment [1], [17].We used the Erroneous Plan Recognition (EPR) system [1], [17] as the pattern-based activity recognition system of this experiment.

The EPR system is implemented in Java and the datasets of micro contexts are stored in MySQL databases on a 3GHz Windows Vista PC with 4GB memory. We used the algorithm LCM [18] to mine frequent patterns, and for correlated patterns, we implemented the algorithm MIP in C++.

We used a meal-time scenario as our testbed, where we employed a human actor to simulate the scenario five times so as to get five sets of simulation data. Each scenario took about 6 to 10 minutes, and the sensor readings were sampled every second. Thus, each data contains 360 to 600 rows (representing a second) and 19 columns (representing the micro contexts). Due to space constraints, the full description of our meal-time scenario is available in [1]. Although our given sensors read raw continuous values, in our context, the data we use have been discretized in a separate and prior process. This means that most of our final attributes have binary values.

We performed five-fold cross validation on the dataset. In the training phase, we mined patterns from the training data and in the testing phase, we used the patterns to detect the activities in the testing data.

For frequent patterns, different numbers of patterns were mined at different minimum support thresholds. As it is impossible to test all settings, we used the frequent pattern with the highest support in each activity to represent that activity. For the correlated patterns, we tried the recommended default setting of $1.0E-4$ for the parameter α of the algorithm MIC [10]. However, we could not mine any patterns. Hence, we set a less stringent setting of $\alpha = 0.5$, and since the data is small in size, setting $\alpha = 0.5$ does not lead to an explosion of the search space. The main advantage of setting a less stringent α is the seeds are not aggressively pruned. To have a fair comparison with frequent patterns, we also used the pattern with the highest correlation information in each activity to represent that activity.

We used precision, recall and F-measure to evaluate the accuracy of the activity recognition by the system using frequent patterns and by the system using correlated patterns.

B. Experimental Results

Figure 2 presents the results of the five-fold cross validation. For recall, the system using frequent patterns and the system using correlated patterns achieve an average of 0.8 and 0.84 respectively, which means that both systems can detect most of the activities. For precision, the system using frequent patterns achieves an average of 0.39 while the system using correlated patterns achieves an average of 0.64, which translates to a 64.1% improvement. This means that the number of wrongly recognized activities by the system using frequent patterns is much higher than the system using correlated patterns. For F-measure, the system using frequent patterns achieves an average of 0.52 while the system using correlated patterns achieves an average of 0.71 – representing a 35.5% improvement.

From these results, we can deduce that the system using frequent patterns attempts to recognize as many activities as possible in the testing data, which results in its large number of correctly and wrongly recognized activities. On the other hand, the system using correlated patterns is able to accurately and correctly recognize the activities. It is possible that the result of the system using frequent patterns can be improved if other minimum support thresholds are used, but it is hard to justify why some thresholds perform better than others. In addition, the heuristical claim of a threshold being the most accurate may only suggest that this threshold is 'overfitted' to the particular dataset.

Therefore, we have shown that a system using correlated patterns achieves higher accuracy than a system using frequent patterns, and there is no hassle of parameter setting, which is normally encountered in using frequent patterns.

V. RELATED WORK

The most widely used activity recognition modeling approach is the Hidden Markov Model (HMM) and its variants, because it has the ability to capture sequence information. It can probabilistically model the complexities and dynamics of the activities of the people in a smart space. With the assumption that different activities map to distinct probability distributions, the activities of each person can be represented as a Markov model. In a dataset of 28 days of annotated sensor data in a smart home, HMM and conditional random fields can achieve a time-slice accuracy of 95.6% and a class

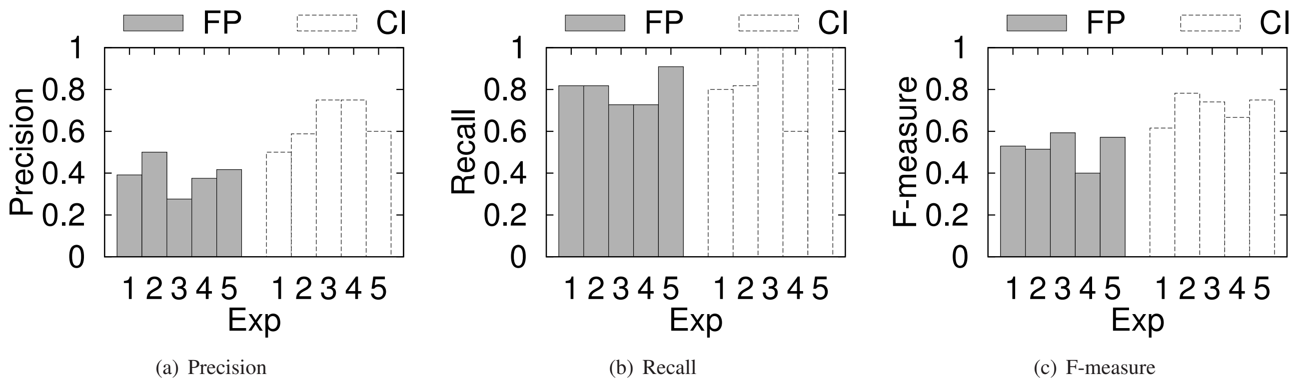


Fig. 2. Activity recognition accuracy using frequent patterns (FP) versus correlated patterns (CI). Exp denotes the experiment label.

accuracy of 79.4% [19]. In a dataset of 200 hours of hospital worker activities in a smart hospital, the proposed HMM can achieve about 92% accuracy for physicians' activities, 94% accuracy for nurses' and interns' activities [20].

However, HMM does not scale well with respect to the number of activities and micro-contexts, as the increasing number will lead to an explosion of hidden states, which will decrease the efficiency of the model. Activity data is often high-dimensional, and both HMM and frequent/correlated pattern mining approaches rely on different assumptions to reduce the complexity of reasoning about high-dimensional data. Therefore, our future work will provide accuracy and run-time comparisons with HMM and its variants.

VI. Conclusion

Assisting dementia patient in performing Activities of Daily Living (ADLs) independently is an important task and several activity recognition systems have been proposed to perform this task. Due to the high dimensionality and large data size, many systems use frequent patterns to detect activities. However, activity recognition using frequent patterns may not be accurate. In this paper, we have proposed using correlated patterns in activity recognition systems, which we have shown in our experiments that the accuracy of the new system is on average 35.5% higher than the system using frequent patterns. A part of our future work will be conducting experiments on real data and on a wider variety of activities.

References

[1] C. Phua, V. Foo, J. Biswas, A. Tolstikov, A. Aung, J. Maniyeri, W. Huang, M.-H. That, D. Xu, and A. Chu, "2-layer erroneous-plan recognition for dementia patients in smart homes," in *Healthcom*, 2009, pp. 21–28.

[2] O. Amft, M. Kusserow, and G. Tröster, "Probabilistic parsing of dietary activity events," in *BSN*, vol. 13, 2007, pp. 242–247.

[3] P. Rashidi and D. Cook, "Keeping the resident in the loop: Adapting the smart home to the user," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 39, no. 5, pp. 949–959, 2009.

[4] D. Lymberopoulos, A. Bamis, and A. Savvides, "Extracting spatiotemporal human activity patterns in assisted living using a home sensor network," in *PETRA*, 2008, pp. 1–8.

[5] Q. Yang, K. Wu, and Y. Jiang, "Learning action models from plan examples using weighted MAX-SAT," *Artificial Intelligence*, vol. 171, no. 2-3, pp. 107–143, 2007.

[6] A. Aztiria, J. Augusto, A. Izaguirre, and D. Cook, "Learning accurate temporal relations from user actions in intelligent environments," in *UCAmI*, 2009, vol. 51, pp. 274–283.

[7] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[8] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining association rules between sets of items in large databases," in *SIGMOD*, 1993, pp. 207–216.

[9] K. Sim, V. Gopalkrishnan, H. N. Chua, and S.-K. Ng, "MACs: Multi-attribute co-clusters with high correlation information," in *ECML/PKDD'09*, 2009, pp. 398–413.

[10] K. Sim, A. Aung, and G. Vivekanand, "Discovering correlated subspace clusters in 3D continuous-valued data," in *ICDM*, 2010, pp. 471–480.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[12] Y. Cheng and G. M. Church, "Biclustering of expression data," in *ISMB*, 2000, pp. 93–103.

[13] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*, 2nd ed. Wiley-Interscience, 1994.

[14] L. J. Bain and M. Englehardt, *Statistical Analysis of Reliability and Life-testing Models*, 2nd ed. CRC-Press, 1991.

[15] T. P. Minka, "Estimating a gamma distribution," 2002, available online at http://research.microsoft.com/en-us/um/people/minka/papers/minka-gamma.pdf.

[16] R. L. Burden and J. D. Faires, *Numerical analysis*, 9th ed. Cengage Learning, 2010.

[17] K. Sim, G. Yap, C. Phua, J. Biswas, P. Wai, A. Aung, A. Tolstikov, W. Huang, and P. Yap, "Improving the accuracy of erroneous-plan recognition system for Activities of Daily Living," in *Healthcom*, 2010, pp. 28–35.

[18] T. Uno, M. Kiyomi, and H. Arimura, "LCM ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets," in *FIMI-ICDM*, 2004.

[19] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *UbiComp*, 2008, pp. 1–9.

[20] D. Sánchez, M. Tentori, and J. Favela, "Activity recognition for the smart hospital," *IEEE Intelligent Systems*, vol. 23, pp. 50–57, March 2008.