Activity Recognition in the Home Using Simple and Ubiquitous Sensors

Emmanuel Munguia Tapia   Stephen S. Intille   Kent Larson

Massachusetts Institute of Technology
1 Cambridge Center, 4FL
Cambridge, MA 02142 USA
{emunguia,intille,kll}@mit.edu

Abstract. In this work, a system for recognizing activities in the home setting using a set of small and simple state-change sensors is introduced. The sensors are designed to be "tape on and forget" devices that can be quickly and ubiquitously installed in home environments. The proposed sensing system presents an alternative to sensors that are sometimes perceived as invasive, such as cameras and microphones. Unlike prior work, the system has been deployed in multiple residential environments with non-researcher occupants. Preliminary results on a small dataset show that it is possible to recognize activities of interest to medical professionals such as toileting, bathing, and grooming with detection accuracies ranging from 25% to 89% depending on the evaluation criteria used [1].

1 Introduction

In this paper, a system for recognizing activities in the home setting using a set of small, easy-to-install, and low-cost state-change sensors is introduced. We show early results that suggest that our sensing technology, which users may perceive as less invasive than cameras and microphones, can be used to detect activities in real homes. The results we present are preliminary but show promise. They are unusual because the ubiquitous computing system and results we describe have been tested in *multiple real homes* with subjects who are not affiliated with the investigators' research group or university.

Our vision is one where a large number of simple, low-cost "tape on and forget" sensors are easily taped on objects throughout an environment and used by a computing system to detect specific activities of the occupant. Computers that can automatically detect the user's behavior could provide new context-aware services in the home. One such service that has motivated this work is proactive care for the aging. Medical professionals believe that one of the best ways to detect emerging medical conditions before they become critical is to look for changes in the activities of daily living (ADLs), instrumental ADLs (IADLs) [17], and enhanced ADLs (EADLs) [24]. These activities include eating, getting

[1] This work was supported, in part, by National Science Foundation ITR grant #0112900 and the Changing Places/House_n Consortium.

in and out of bed, using the toilet, bathing or showering, dressing, using the telephone, shopping, preparing meals, housekeeping, doing laundry, and managing medications. If it is possible to develop computational systems that recognize such activities, researchers may be able to automatically detect changes in patterns of behavior of people at home that indicate declines in health. The system described in this work could potentially be retrofit into existing homes to detect and monitor ADLs.

2 Background

Everyday activities in the home roughly break down into two categories. Some activities require repetitive motion of the human body and are constrained, to a large extent, by the structure of the body. Examples are walking, running, scrubbing, and exercising. These activities may be most easily recognized using sensors that are placed on the body (e.g. [19, 11, 18]). A second class of activities, however, may be more easily recognized not by watching for patterns in how people move but instead by watching for patterns in how people move *things*. For instance, the objects that someone touches or manipulates when performing activities such as grooming, cooking, and socializing may exhibit more consistency than the way the person moves the limbs.

In this work we focus on the latter problem and ask the question, "can activities be recognized in complex home settings using simple sensors that detect changes in state of objects and devices?" Although progress is being made on algorithms that monitor a scene and interpret the sensor signals from complex sensors such as cameras or microphones, the recognition inference problem is often seriously underconstrained. Computer vision sensing, for example, often works in the laboratory but fails in real home settings due to clutter, variable lighting, and highly varied activities that take place in natural environments. Little of the work with video and audio processing in the lab has been extensively tested in the field. Perhaps just as importantly, however, because sensors such as microphones and cameras are so general and most commonly used as recording devices, they can be perceived as invasive and threatening by some people.

For these reasons, we are exploring the recognition potential of deploying very large numbers of extremely simple sensors. Simple sensors can often provide powerful clues about activity. For instance, a switch sensor in the bed can strongly suggest sleeping [1], and pressure mat sensors can be used for tracking the movement and position of people [22, 2]. Although others have written on the potential of sensor networks (e.g. [7, 13, 14]), we are unaware of work where large numbers have been deployed in multiple, non-laboratory home environments and used for ADL pattern recognition.

Previous work where sensors have been placed on objects in the environment have typically been used in laboratories or homes of the researchers themselves and their affiliates. Further, all of these systems have required careful (and usually painstaking) installation and maintenance by research staff and students (e.g. [20, 1, 21]). With few exceptions (e.g. [20]) only a small portion of the homes are sensor-enabled. Prior work, however, has shown the potential of multiple, simple switches for activity detection. In the MARC home, simple sensors in a kitchen (temperature on stove, mat sensors, and cabinet door sensors) have been used to detect meal preparation activities [2]. In that work, mixture models and hierarchical clustering were used to cluster the low-level sensor readings into cooking events using temporal information [2]. However, choosing the number of clusters to use and correlating the clusters of sensor readings to activities may grow more difficult as larger numbers of sensors are added to environments to recognize a more diverse set of activities. RFID tags placed at objects in the environment and combined with unsupervised mining of activity models from the web have also shown promise for activity recognition [23]. Although this approach does not need the subject to label his activities, it could prove difficult to adapt to individual patterns of activities. In this work we explore a supervised learning approach.

Hierarchical hidden semi-Markov models (HHSMMs), a type of dynamic belief network (DBN), have been used to track the daily activities of residents in an assisted living community [15]. The algorithm can distinguish different activities such as “asleep” and “having meals” solely based on noisy information about the location of the residents and when they move. Even though DBNs show some promise, they may not scale to environments that contain hundreds of sensors, particularly if real-time recognition of activity is a goal.

Sequence matching approaches have been applied to predict inhabitant’s actions. The SHIP algorithm matches the most recent sequence of events with collected histories of actions to predict inhabitant future actions [4]. This approach, however, does not model ambiguous and noisy information from multiple sensors.

3 Activity Detection Approach

The following design goals motivated the activity recognition algorithms developed in this work.

Supervised learning. Homes and their furnishings have highly variable layouts, and individuals perform activities in many different ways. The same activity (e.g. brushing teeth) may result in a significantly different sensor activation profile based upon the habits, or *routines* of the home occupant and the layout and organization of the particular home. One approach to handling such variability is to use supervised learning with an explicit training phase.

Probabilistic classification. Probabilistic reasoning offers a way to deal with ambiguous and noisy information from multiple sensors.

Model-based vs instance-based learning. Model-based algorithms use the training examples to construct a mathematical model of the target classification function, which avoids the need to save all examples as raw data. This could help alleviate end-user privacy concerns.

Sensor location and type independent. Ideally, the system would operate effectively even when the algorithm is never explicitely told the location (e.g. kitchen) and type (e.g. drawer) of a particular sensor. This would dramatically reduce installation time.

Real-time performance. A system that recognizes activities in the home setting is most useful if it performs in real-time. Training or model construction time is less of a concern.

Online learning. Ideally the system would be capable of adjusting the internal model in real-time as new examples of activities become available. This will allow the algorithm to adapt to changes in the user's routines over time.

In this work, we chose naive Bayesian classifiers[16] to detect activities using the tape-on sensor system . Naive Bayesian classifiers make strong (and often clearly incorrect) assumptions that each class attribute is independent given the class. They also assume that all attributes that influence a classification decision are observable and represented. For these reasons, they are sometimes assumed to perform poorly in real domains. On the contrary, however, experimental testing has demonstrated that naive Bayes networks are surprisingly good classifiers on some problem domains, despite their strict independence assumptions between attributes and the class. In fact, simple naive networks have proven comparable to much more complex algorithms, such as the C4 decision tree algorithm [16, 3, 12, 5]. They also meet the design goals listed above.

One theory on why naive Bayes classifiers work so well is that the low variance of the classifier can offset the effect of the high bias that results from the strong independence assumptions [6]. Although in this preliminary work we are limited to small datasets, over time a tape-on sensor system and the experience sampling data collection method described shortly could be used to collect a large sample of activity from a user's home to train a classification system. Even in the results presented in this work, more data appear to lead to better recognition results. To apply naive Bayes classifiers to the activity recognition problem, however, temporal dependencies may need to be considered. Therefore, one approach would be to encode large numbers of low-order temporal relationships in the networks [8]. In this work, the naive Bayes classifier is extended to incorporate temporal relationships among sensor firings and recognize activities in the home setting. These classifiers are easy to train, fast, and seem to improve in performance with larger training sets.

Two versions of the activity recognition classifier were implemented. The first is a multi-class naive classifier in which the class node represents all the activities to recognize and its child nodes consist of one of two types: *exist* and *before* attributes. In this configuration, all the activities are considered to be mutually-exclusive, which means that the probabilities for all activities sum up to one at any given time. The second version of the activity recognition classifier implemented is multiple binary naive Bayes classifiers, each of them representing an activity to recognize. The main advantage of this binary decomposition is that the representation does not enforce mutual exclusivity. In this way, detection of *listening to music* does not preclude detection of *preparing breakfast*. In

(a) (b)

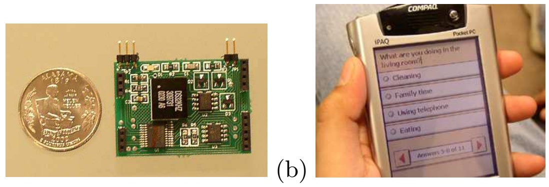


Fig. 1. (a) The state-change sensors that can be installed ubiquitously throughout an environment. Each device consists of a data collection board (shown) and a small sensor. (b) One screenshot from the ESM tool used in this work to collect training data on activities in the home setting.

this work, the prior probabilities for all the activities are assumed to be equal. Moreover, maximum likelihood was used to learn the parameters of the networks.

4 Activity recognition system architecture

The proposed system consists of three major components: (1) *The environmental state-change sensors* used to collect information about use of objects in the environment, (2) *the context-aware experience sampling tool (ESM)* used by the end user to label his or her own activities, and (3) *the pattern recognition and classification algorithms* for recognizing activities after constructing a model based on a training set.

4.1 Environmental State-Change Sensors

Although other low-cost wireless sensing systems have been developed, notably Berkeley Motes [7] and Smart-ITS [14], their power and cost points still pose a challenge for researchers interested in distributing hundreds of units in a single home to collect synchronized data for several weeks or longer. The cost of these devices is relatively high because they are designed as multi-purpose sensors. Therefore, we have designed a new set of tape-on sensors optimized to perform a single task at low cost: measuring change in the state of an object in home [9]. To achieve well-synchronized measurements, the most precise real-time clock hardware was used in each board. Further, the signals from each board were linearly interpolated to match the reference clock better after the end of the study. These highly-specialized boards are 3-5 times less expensive than Smart-ITS and Motes, which dramatically increases the number that can be installed in homes working within a tight research budget. The estimated battery life of the data collection board is one year if the external sensor is activated an average of 10 times per day for 30 seconds.

Figure 1a shows a sensor device, which actually consists of the sensor itself connected by a thin wire to a 27mm x 38mm x 12mm data collection board. The board fits snugly in a small plastic case of dimensions 37mm x 44mm x 14mm.

Activity	sensor ID	day	activation time	deactivation time	duration (sec)	room (opt)	object type (opt)
Preparing breakfast	PDA	12/1/02	08:23:01		10 min		
	23	12/1/02	08:23:03	08:23:07	4	kitchen	drawer
	18	12/1/02	08:23:09	08:23:17	8	kitchen	cabinet
	89	12/1/02	08:24:49	08:24:59	10	kitchen	fridge door
		⋮	(many readings)				

Table 1. An example of the type of data that was acquired by the state-change sensors and ESM. The activity attributes are acquired using experience sampling during a training period. The sensor activations are collected by the state-change sensors distributed all around the environment. In the table, opt stands for optional attribute.

The boards can use either reed switches, which are activated when brought into contact with a small magnet, or piezoelectric switches, which detect movement of a small plastic strip.

4.2 Context-Aware Experience Sampling

Supervised learning algorithms require training data. In the laboratory, obtaining annotated data is a straightforward process. Researchers can directly observe and label activity in real-time, or later through observation of video sequences. In the home environment, however, direct observation is prohibitively time-consuming and invasive.

One alternative is to use the Experience Sampling Method (ESM) [10, 9]. When using ESM, subjects carry a personal digital assistant (PDA) that is used as timing device to trigger self-reported diary entries. The PDA samples (via a beep) for information. Multiple choice questions can then be answered by the user. Figure 1b shows a screen shot from the ESM tool used in this work. The protocol used to collect subject self report labels of activity in this work using ESM is described in section 5.2.

4.3 Activity Recognition Algorithms

The purpose of the state-change sensors and ESM was to provide the necessary data to create machine learning algorithms that can identify *routines* in activities from sensor activations alone. In order to accomplish this goal, new algorithms that correlate the sensor firings and activity labels and predict activities from new sensor firings are required. Table 1 shows an example of the type of data acquired with the sensors and using the ESM tool.

5 Study and Data Collection

Two studies were run in two homes of people not affiliated with our research group to collect data in order to develop and test the activity recognition algorithms. Both subjects granted informed consent and were compensated with $15.00 dollars per day of participation in the study. The first subject was a professional 30-year-old woman who spent free time at home, and the second was an 80-year-old woman who spent most of her time at home. Both subjects lived alone in one-bedroom apartments. 77 state-change sensors were installed in the first subject's apartment and 84 in the second subject's apartment. The sensors were left unattended, collecting data for 14 days in each apartment. During the study, the subjects used the context-aware ESM to create a detailed record of their activities.

5.1 State-Change Sensors Installation

The state-changes sensors described in section 4.1 were installed on doors, windows, cabinets, drawers, microwave ovens, refrigerators, stoves, sinks, toilets, showers, light switches, lamps, some containers (e.g water, sugar, and cereal), and electric/electronic appliances (e.g DVDs, stereos, washing machines, dish washers, coffee machines) among other locations. The plastic cases of the data collection boards were simply placed on surfaces or adhered to walls using non-damaging adhesive selected according to the material of the application surface. The sensor components (e.g. reed and magnet) and wire were then taped to the surface so that contact was measured. Figure 2 shows how some of the 77 sensors were installed in the home of the first subject. The devices were quickly installed by a small team of researchers: an average of about 3 hours is required for the sensors installation in a small one-bedroom apartment of typical complexity. When sensors were installed, each data collection board (which has a unique ID) was marked on a plan-view of the environment so that when the sensor data was collected, the location (e.g kitchen) and type (e.g cabinet) of each sensor was known.

5.2 Labelling Subject's Activities

Experience sampling. The subjects were given a PDA running the ESM software at the start of the study. As the state-change sensors recorded data about the movement of objects, the subjects used experience sampling to record information about their activities. A high sampling rate was used, where the subject was beeped once every 15 minutes for 14 days (study duration) while at home. At the beep, the subject received the following series of questions. First the user was asked "what are you doing at the beep (now)?". The subject could select the activity that best matched the one that he/she was doing at the time of the beep from a menu showing up to 35 activities. Next, the following question was "For how long have you been doing this activity?" The subject could select from a list of four choices: less than 2 min., less than 5 min, less than 10 min., and more than 10 min. Then, the user was asked, "Were you doing another activity before the beep?". If the user responded positively, the user was presented with a menu of 35 activities once again. For the studies, an adaptation of the activity categories used by Szalai in the multi-national time-use study [25] were used.

Fig. 2. Examples of some of the 77 sensors that were installed in the home of the first subject. The sensors and data collection boards were literally taped to objects and surfaces for the duration of the data collection period.

Several problems were experienced with the ESM annotation method, some of which were learned about via interviews with subjects. Errors were observed where the user selected the wrong activity from the list by mistake. Short duration activities such as toileting were difficult to capture. There were delays between the sensor firings and the labels of the activities specified in the ESM. Fewer labels were collected than anticipated because subjects sometimes did not answer the ESM questions at the beep. Finally, sometimes subjects specified one activity and carried out a different activity.

Indirect observation of sensors activations. Unfortunately, the number of labels acquired using the ESM method was not sufficient for training the machine learning algorithms. Therefore, we were forced to resort to indirect observation by studying the sensor activations. In this method, the author, with the help of each subject, used self-inference to label the sensor data by visualizing the sensor activations clustered by location, time of activation, and type (description) of each sensor. Photographs of the sensors were also used to help the subject remember her activities during the sensor firings. A few decisions made during the manual annoation step impact the results that follow. First, activities were assumed to occur sequentially. The only activities allowed to occur in parallel with other activities were *Listening to Music* and *Watching TV*. Only the primary activity was labeled if a person was multi-tasking. Finally, only activities for which there exist sensor activations were labeled.

Figure 2 shows the number of labels generated by ESM and by indirect observation of sensor activations. For both subjects, the combined number of labeled activities is far less than desirable for a supervised learning algorithm. In current work, we are improving the subject self-annotation methodology to generate

Measure	Subject 1	Subject 2
Average activities captured per day using ESM	9.5	13
Average activities per day generated by I.O	17.8	15.5
Different activities captured using ESM	22	24
Different activities generated by I.O	21	27
Average ESM Prompts answered to per day	18.7	20.1

Table 2. Average number of labels collected by the ESM and indirect Observation (I.O) per day during the study.

better datasets. However, here we described our work with this admitedly small but still useful pilot dataset.

6 Feature Extraction, Training and Prediction

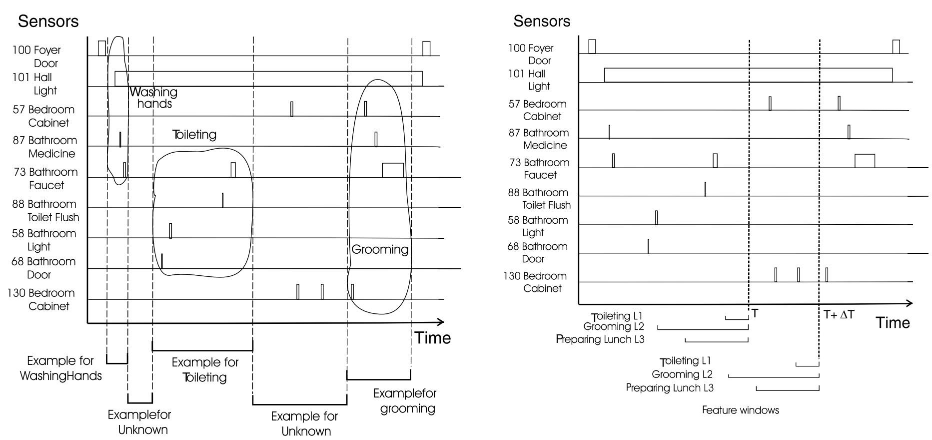
We assumed that temporal information, in addition to which sensors fired, would be necessary to achieve good recognition using the naive Bayesian network approach. Therefore, one idea we explored in this work was to encode large numbers of low-order binary temporal relationships in the naive Bayesian network classifier. Two temporal features have been used. The first is whether activation of a particular sensor *exists* during some time period. The second is whether a particular sensor fires *before* another particular sensor. Table 3 shows the binary features calculated over the sensor data. These features output the evidence entered into the nodes of the naive Bayesian network.

The last two features in the table incorporate high level contextual information about the *type* of object in which the sensor was installed (e.g cabinet) and *location* of the sensor (e.g bathroom). The number of *exist* features that will become nodes in the naive Bayes networks is equal to the number of sensors present in the system (77 and 84 for subject one and two respectively). The number features that become nodes for the *before sensorID*, *before type* and *before location* features is equal to the number of all pairs of sensors, object types, and locations existent in the home environment (77x77=5929, 27x27=729, and 6x6=36 for subject one respectively).

Incorporating activity duration. Different activities have different mean lengths of time. Therefore, in order to incorporate the activity duration, one feature window per activity to recognize was used, and the length of each window corresponded to the activity duration as carried out by the subject. Thus, if M is the number of activities to recognize, there were M different feature windows with lengths $L_i \cdots L_m$. The duration or length L_i for each feature window was the average duration for each activity calculated from all the activity labels generated by ESM and indirect observation. For example, the feature window for toileting for the first subject was estimated to be 7 min, 27 sec. Preparing lunch was estimated to be 37 min, 54 sec.

Feature description	Example
exist(*sensorA*, *start*, *end*)	Sensor A fires within time interval
before(*sensorA*, *sensorB*, *start*, *end*)	Sensor A fires before sensor B within time interval
before(*sensorTypeA*, *sensorTypeB*, *start*, *end*)	Sensor in a drawer fires before a sensor in the fridge within time interval
before(*sensorLocationA*, *sensorLocationB*, *start*, *end*)	Sensor in kitchen fires before sensor in bathroom within time interval

Table 3. Features calculated and evaluated



(a) (b)

Fig. 3. (a) Example of how training examples are created for "washing hands", "toileting", "grooming" and two "unknown" activities. (b) Example of how features are extracted from sensor firings using different feature window lengths for each activity for time t and the next time to analyze $t+\triangle t$ in the prediction step.

Generation of Training Examples. In the training stage, training examples are generated by calculating the features from the start to the end time of each activity label. Figure 3a shows an example of how training examples are generated. Examples for washing hands, toileting, and grooming are generated whenever a label for washing hands, toileting, and grooming is found in the dataset respectively.

Originally, there was no *unknown* activity, but examples of this class were created by generating an example of it whenever no activity labels for other activities were found in the dataset. Figure 3a also shows an example of how two examples for the unknown class were generated.

Predicting the activity labels. In the prediction step, each feature window (of length L_i) is positioned at the current time to analyze, t. The features are then calculated from time $t - L_i$ to time t. Once the features are calculated, the probability for the current activity is calculated using the naive Bayes classifier.

Figure 4 conceptually shows an example of how the probability for each activity is generated in the prediction step by shifting the feature window for each activity over the sensor activations. Note that the probability is maximum when the feature window aligns with the duration of the activity represented by sensor activations (activity label). This indicates that the classifier may (depending upon noise) report the best match at the moment the activity is ending.

Figure 3b shows an example of how the feature windows for each activity are positioned in the current time to analyze t and in the next time to analyze $t{+}\triangle t$ over simulated sensor data. The $\triangle$t increment in time used in the experiments was three minutes, which was half of the duration of the quickest activity. In a real-time application, however, the $\triangle$t can be chosen to be as small as required, for example 5 seconds. While predicting an activity label for new observed sensor firings, the activity with the maximum likelihood at any given time is considered to be the classification result.

7 Algorithm Recognition Performance

Once the ESM and indirect observation labels were available, they were used to train and test the machine learning algorithms. All activities containing less than six examples were eliminated before training.[2]

Unlike other machine learning and pattern recognition problems, there is no "right" answer when recognizing activities. The boundaries when activities begin and end are fuzzy since they can occur sequentially, in parallel, alternating, and even overlapping. Finally, there is significant variation in the way observers would label the same activities.

Three methods were used to evaluate and measure the accuracy of the activity recognition algorithms. Which method is most informative depends upon the type of application that would be using the activity recognition data. The methods consider different features of the system that could be important for different applications, for example: (1) is the activity detected at all? and (2) for how long is the activity detected? Figure 5 shows examples of each of the three evaluation measures.

Percentage of time that activity is detected. This measures the percentage of time that the activity is correctly detected for the duration of the labelled activity.

Activity detected in best interval. This measures whether the activity was detected "around" the end of the real activity or with some delay ϕ. As discussed in section 6, the end of the activity is "the best detection interval". Thus, the right most edge of each activity (E) is analyzed within an interval of $\pm\phi$. It is important to remember that a detection delay is introduced

[2] The threshold of six was chosen arbitrarily. Given the complexity of the activities and the large amount of variation possible due to day of the week, time, and other factors, to expect an algorithm to learn patterns with less than six examples did not seem reasonable.

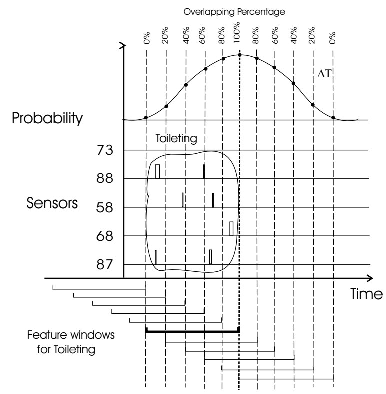


Fig. 4. Example of how the probability for the "toileting" activity is generated in the prediction step by shifting the feature window for "toileting" over the sensor activations with increments of $\triangle t$ (3 minutes for this study). Note that the probability is maximum when the feature window aligns with the duration of the activity represented by the sensor activations.

by the use of the feature windows that capture features back in time in our algorithm. In this work, the interval ϕ was chosen to be 7.5 minutes. Different applications would require different detection delays, thus different values of ϕ could be used.

Activity detected at least once. This measures if an activity was detected at least once for the duration of the activity label (no delay allowed).

Leave-one-out cross-validation was used in each evaluation method in order to calculate the confusion matrix and measure the classification accuracy. Cross-validation permits some classification testing even on small datasets. The activity with the maximum likelihood at a given time was used when determining classification accuracy using each of the three evaluation metrics.

Experiments to determine the discrimination power of the attributes were performed by running the multi-class and multiple binary naive Bayes classifiers with some of the possible combinations of attributes shown in Table 3. Tables 4 and 5 show the accuracies per class for the combination of attributes that performed the best for the multiclass naive Bayes classifier for subject one and two respectively.

8 Discussion

Accuracies vs number of examples. As expected, the activities with higher accuracies were generally those with more examples. For subject one, they were "toileting", "grooming","bathing", and "doing laundry". For subject two, they were "preparing lunch", "listening to music", "toileting" and "preparing breakfast".

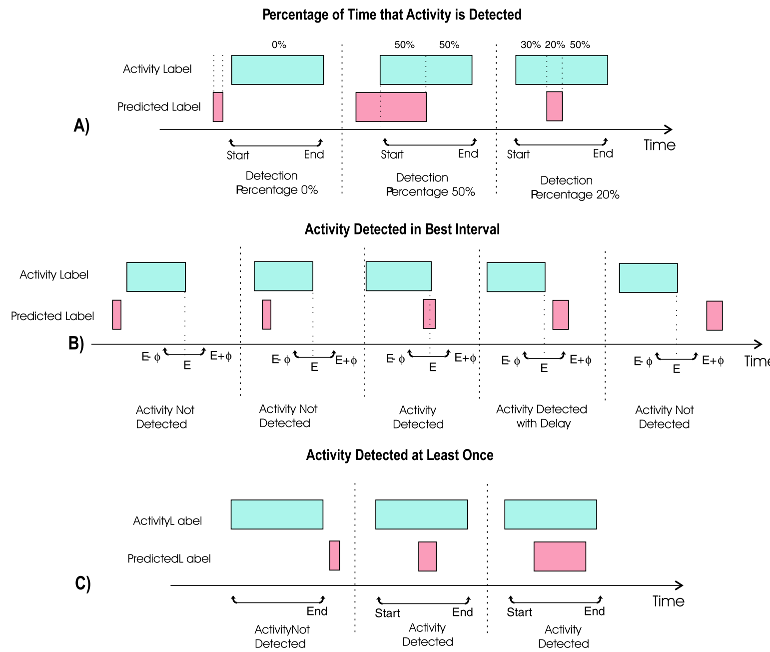


Fig. 5. Example cases of the (a) "percentage of time that activity is detected", (b) "activity detected in best interval" and (c) "activity detected at least once" evaluation methods.

Discriminant attributes. Overall, the *exist* attribute showed the best discriminant power. In this case, the naive Bayesian network is actually acting as a weighted voting mechanism. Adding temporal features such as *before* did not provide the discrimination power expected. We attribute this to the relatively small size of our datasets. When "ground truth" video is available for labelling the subject activities (planned for future installations) and more examples and multi-tasking examples are collected, this feature may show a higher discrimination power.

Optional attributes *Type and Location.* Preliminary results show that adding the attributes using the type of object in which the sensor was installed and location information such as the *before type* and *before location* features to the *exist* attribute did not represent a significant improvement in accuracy. This suggests that the development of activity recognition algorithms that do not use the "type" and "location" information may be possible.

Accuracy vs evaluation method used. The activities show lower accuracies for the "percentage of time" evaluation method. Since activities are being detected from sensor firings, some activities such as watching TV, listening to music and dressing are represented by sensors firing only at the beginning or at the end of the activity. This means that there is no information other

Multiclass Naive Bayes Classifier for Subject One					
Activity	**No. Examples**	**E**	**E+BT**	**Random Guess**	**Evaluation**
Preparing lunch	17	0.25	0.29	0.07	Percentage of Time Activity is Detected
Toileting	85	0.27	0.31	0.07	
Preparing breakfast	14	0.08	0.06	0.07	
Bathing	18	0.25	0.29	0.07	
Dressing	24	0.07	0.03	0.07	
Grooming	37	0.26	0.26	0.07	
Preparing a beverage	15	0.07	0.13	0.07	
Doing laundry	19	0.09	0.07	0.07	
Preparing lunch	17	0.59	0.78	0.30	Activity Detected in Best Interval
Toileting	85	0.71	0.71	0.30	
Preparing breakfast	14	0.45	0.45	0.30	
Bathing	18	0.87	0.79	0.30	
Dressing	24	0.64	0.41	0.30	
Grooming	37	0.89	0.86	0.30	
Preparing a beverage	15	0.36	0.36	0.30	
Doing laundry	19	0.86	0.78	0.30	
Preparing lunch	17	0.50	0.68	0.17	Activity Detected at Least Once
Toileting	85	0.42	0.43	0.03	
Preparing breakfast	14	0.20	0.12	0.07	
Preparing a snack	14	0.08	0.05	0.03	
Bathing	18	0.70	0.75	0.11	
Going out to work	12	0.12	0.00	0.02	
Dressing	24	0.21	0.07	0.02	
Grooming	37	0.68	0.71	0.05	
Preparing a beverage	15	0.22	0.31	0.04	
Doing laundry	19	0.27	0.23	0.05	
Activities with Less than Six Examples					
Work at home(0), Eating(0), Washing hands(1), Sleeping(0), Taking medication(0), Sleeping(0), Talking on telephone(0), Resting(0), Putting away dishes(2), Putting away groceries(2), Putting away laundry(2), Taking out the trash(0), Lawnwork(1), Pet care(0), Home education(0) Going out to school(0), Going out for entertainment(1), Working at computer(0), Going out to exercise(0), Going out for shopping(2), Listening to music(0), and Watching TV(3).					
Activities Not Recognized Better than Random Guess					
Preparing dinner(8), Washing dishes(7), Preparing a snack(14), Going out to work(12), and cleaning(8)					

Table 4. Leave-one-day-out crossvalidation accuracies per class for the multiclass naive Bayes classifier using the best two combination of features for subject one. E stands for the *exist* feature, and BT stands for the *before type* feature.

than the average duration of the activity represented by the feature windows to detect the activities during these "dead intervals" of sensor firings.

The evaluation method with the highest detection accuracies per activity was the "best interval detection". The classes with the highest "best interval detection" accuracies (over 70%) were "bathing", "toileting", "grooming", and "preparing lunch" for the first participant. For the second participant the higher "best interval detection" accuracies (over 50%) were "preparing breakfast", "preparing lunch", "listening to music" and "toileting".

Accuracy vs number of sensors. Since activities such as "going out to work" and "doing laundry" are represented by sensor firings from a single sensor (door and washing machine respectively), it was expected that they would show higher detection accuracies than other activities. However, the sensors were also activated during other activities which decreased their discrimi-

Multiclass Naive Bayes Classifier for Subject Two					
Activity	**No. Examples**	**E**	**E+BT**	**Random Guess**	**Evaluation**
Preparing lunch	20	0.22	0.22	0.10	Percentage of Time Activity is Detected
Listening to music	18	0.20	0.09	0.10	
Toileting	40	0.20	0.23	0.10	
Preparing breakfast	18	0.30	0.24	0.10	
Washing dishes	21	0.05	0.11	0.10	
Watching TV	15	0.04	0.16	0.10	
Preparing lunch	20	0.51	0.48	0.40	Activity Detected in Best Interval
Listening to music	18	0.61	0.44	0.40	
Toileting	40	0.52	0.48	0.40	
Preparing breakfast	18	0.68	0.59	0.40	
Washing dishes	21	0.51	0.54	0.40	
Watching TV	15	0.25	0.52	0.40	
Preparing dinner	14	0.38	0.30	0.24	Activity Detected at Least Once
Preparing lunch	20	0.48	0.61	0.26	
Listening to music	18	0.66	0.45	0.38	
Toileting	40	0.46	0.43	0.10	
Preparing breakfast	18	0.75	0.65	0.16	
Washing dishes	21	0.15	0.28	0.09	
Watching TV	15	0.08	0.45	0.30	
Activities with Less than Six Examples					
Work at home(0), Going out to work(0), Eating(0), Bathing(3), Grooming(3), Dressing(5), Washing hands(0), Sleeping(0), Talking on telephone(4), Resting(0), Preparing a beverage(1), Putting away dishes(3), Putting away groceries(1), Cleaning(3), Doing laundry(0), Putting away laundry(1), Taking out the trash(0), Lawnwork(1), Pet care(0), Home education(2), Going out to school(0), Going out for entertainment(1), Working at computer(5), Going out to exercise(0), and Going out for shopping(3).					
Activities Not Recognized Better than Random Guess					
Preparing dinner(14), Taking medication(14), and Preparing a snack(16).					

Table 5. Leave-one-day-out crossvalidation accuracies per class for the multiclass naive Bayes classifier using the best two combination of features for subject two. E stands for the *exist* feature, and BT stands for the *before type* feature.

nant power. The most activated sensors overall for both subjects were the kitchen door, refrigerator, and cabinets. For subject one, the three other most activated sensors were the bathroom sink faucet (165), medicine cabinet (118), and kitchen drawer #84 (74). For subject two the three other most activated sensors were the microwave oven (197), garbage disposal (79), and living room TV (63). The least activated sensors were located in the living room and hallway area.

Accuracy vs sensors installation locations. The state change sensors were not appropriate for installation on some useful objects. For example, sensors were not installed on pans, dishes, chairs, tables and other locations that could improve the recognition accuracy of for preparing dinner. A new version of the tape on sensors in development that uses accelerometry instead of reed switch sensing will permit a 2-3 times increase in the number of sensors that can be installed in a given environment. This may improve recognition, but it also may increase the need for larger training data sets.

Multiclass vs multiple binary classifiers. We have described our implementation of a multiclass naive Bayes classifier. However, we also tested the system with multiple binary classifiers – one per activity. Multiple binary classifiers do not enforce mutual exclusivity, since each classifier is independent of the others. On our dataset, the multiclass and multiple binary classifiers performed with approximately the same accuracy ($\pm 5\%$). However, the multiple binary classifiers may perform better in future studies as more accurate activity labels become available with multi-tasking.

Some of the false positives obtained in this work almost certainly resulted from the fact that a considerable number of short or multi-tasked activities carried out by the subjects were not labelled in our dataset.

Subject one vs Subject two results. Overall, recognition accuracies for the first subject's data were higher than those for the second subject's data. This was mainly for two reasons: (1) the number of sensors that were dislodged, failed and were noisy was higher in the second subject apartment, and (2) the quality of the labels for the first subject was higher because the sensor firings for each activity were easier to label. One possible explanation is that the first subject spent less time at home and the sensor firings were not as complex as those for the second subject.

Improvement over the random guess baseline. The results shown in Tables 4 and 5 represent a significant improvement for some activities over the random guess baseline[3].

Even though the accuracy for some activities is considerably better than chance, it is not as high as expected. The main problems faced were: (1) the quality and number of activity labels, and (2) the small training set of two weeks. It is expected that training sets collected over months, better labels generated by video observation or other methods, and improved versions of the data collection boards and sensors will improve the detection accuracies. Although the results presented here are preliminary, we believe they show promise. Experiments are underway to improve the data collection and annotation process and to acquire substantially more detailed datasets.

9 Subject Experiences

The studies described here have led to some useful qualitative observations about subject reaction to the sensor system and the ESM data collection method. For instance, the participants felt that they became "oblivious" to the presence of the sensors after a few days of the installation. The first subject reported forgetting that the sensors were in her apartment after two days. The second subject was not even sure where some of the sensors were installed. We suspect the acclimation period to more invasive sensors such as cameras would be substantially longer. One of our subjects told us she would not have agreed to the study if

[3] The random guess baseline for the "percentage of time" evaluation method is $1/n$, where n is the number of activities. The random guess is $1-((n-1)/n)^i$ for the "best interval" and "at least once" methods, where $i=(2\phi)/\triangle\mathrm{t}$ for "best interval" and $i=activity\_average\_duration/\triangle\mathrm{t}$ for "detected at least once".

it had involved video observations. The second subject would have agreed but would have restricted cameras from the bathroom.

Subjects had a more difficult time adjusting to the experience sampling device. They started to find the ESM highly disruptive by the second and third day of the study. This was probably because of the high sampling rate (15 minutes), but even with extremely helpful volunteers it has become clear that tolerance for repetitive sampling of the same activities is low. In short, subjects did not mind "teaching the computer" about new activities but did not enjoy having to tell the computer about doing the same activities repetively. The percent of total ESM prompts responded to was 17% and 24% for the first and second subject respectively.

Finally, even these simple sensors can impact behavior. The first participant reported that that being sensed did cause her to alter some behaviors. For example, she always made sure to wash her hands after using the bathroom. The second subject did not report any changes in behavior due to the sensors.

10 Summary

The work described here is preliminary but demonstrates that ubiquitous, simple sensor devices can be used to recognize activities of daily living from real homes. Unlike prior work in sensor systems for recognizing activities, the system developed in this work was deployed in multiple residential environments with actual occupants. The occupants were not researchers or affiliated with the experimenters. Moreover, the proposed sensing system presents an alternative to sensors that are perceived by most people as invasive such as cameras and microphones. Finally the system can be easily retrofitted in existing home environments with no major modifications or damage.

References

1. Elite Care's Oatfield Estates. http://www.elite-care.com/oatfield-tech.html.
2. T. Barger, M. Alwan, S. Kell, B. Turner, S. Wood, and A. Naidu. Objective remote assessment of activities of daily living: Analysis of meal preparation patterns. Poster presentation, Medical Automation Research Center, University of Virginia Health System, 2002.
3. P. Clark and T. Niblett. The CN2 induction algorithm. *Machine Learning*, 3(4):261–283, 1989.
4. S. K. Das, D. J. Cook, A. Bhattacharya, E. O. Heierman, and T.Y. Lin. The role of prediction algorithms in the MavHome smart home architecture. In IEEE Wireless Communications, editor, *IEEE Press*, volume 9, pages 77–84, 2002.
5. P. Domingos and M. Pazzani. Beyond independence: Conditions for the optimality of a simple bayesian classifier. In L. Saitta, editor, *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 194–202. Morgan Kauffman, 1996.
6. J. Friedman. On bias, variance, 0/1 - loss, and the curse-of-dimensionality. *Data mining and knowledge engineering*, 1:55–77, 1997.

7. S. Hollar. *COTS Dust*. Ph.D. thesis, University of California, Berkeley, 2001.
8. S.S. Intille and A.F. Bobick. Recognizing planned, multi-person action. *Computer Vision and Image Understanding (1077-3142)*, 81(3):414–445, 2001.
9. S.S. Intille, E. Munguia Tapia, J. Rondoni, J. Beaudin, C. Kukla, S. Agarwal, L. Bao, and K. Larson. Tools for studying behavior and technology in natural settings. In A. K. Dey, A. Schmidt, and J. F. McCarthy, editors, *Proceedings of UBICOMP 2003*, volume LNCS 2864. Springer, 2003.
10. S.S. Intille, J. Rondoni, C. Kukla, I. Anacona, and L. Bao. A context-aware experience sampling tool. In *Proceedings of the Conference on Human Factors and Computing Systems: Extended Abstracts*. ACM Press, 2003.
11. J. Himberg J. Mantyjarvi and T. Seppanen. Recognizing human motion with multiple acceleration sensors. *IEEE International Conference on Systems, Man, and Cybernetics.*, pages 747–52, IEEE Press, 2001.
12. G.H. John and P. Langley. Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo, AAAI Press, 1995.
13. J.M. Kahn, R.H. Katz, and K.S.J. Pister. Mobile networking for Smart Dust. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 99)*, pages 271–278. 1999.
14. O. Kasten and M. Langheinrich. First experiences with Bluetooth in the Smart-Its distributed sensor network. In *Workshop on Ubiquitous Computing and Communications, PACT*. 2001.
15. H. Kautz, O. Etziono, D. Fox, and D. Weld. Foundations of assisted cognition systems. Technical report CSE-02-AC-01, University of Washington, Department of Computer Science and Engineering, 2003.
16. P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228. AAAI Press, San Jose, CA, 1992.
17. M.P. Lawton and E.M. Brody. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, 9:179–186, 1969.
18. S.W Lee and K. Mase. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, 1(3):24–32, 2002.
19. M. Makikawa and H. Iizumi. Development of an ambulatory physical activity monitoring device and its application for categorization of actions in daily life. In *MEDINFO*, pages 747–750. 1995.
20. M. Mozer. The Neural Network House: an environment that adapts to its inhabitants. In *Proceedings of the AAAI Spring Symposium on Intelligent Environments*, Technical Report SS-98-02, pages 110–114. AAAI Press, Menlo Park, CA, 1998.
21. B. Ornstein. Care technology: Smart home system for the elderly. In *Proceedings of NIST Pervasive Computing*. 2001.
22. R.J. Orr and G.D. Abowd. The Smart Floor: A mechanism for natural user identification and tracking. In *Proceedings of the 2000 Conference on Human Factors in Computing Systems (CHI 2000)*. ACM Press, 2000.
23. M. Philipose, K.P. Fishkin, D. Fox, H. Kautz, D. Patterson, and M. Perkowitz. Guide: Towards understanding daily life via auto-identification and statistical analysis. In *UbiHealth Workshop*. Ubicomp, 2003.
24. W.A. Rogers, B. Meyer, N. Walker, and A.D. Fisk. Functional limitations to daily living tasks in the aged: a focus groups analysis. *Human Factors*, 40:111–125, 1998.
25. S. Szalai. *The Use of Time. Daily Activities of Urban and Suburban Populations in Twelve Countries*. Mouton, The Hague, 1973. Edited by Alexander Szalai.