

# Learning Setting-Generalized Activity Models for Smart Spaces

Diane J. Cook

Washington State University, Pullman, WA USA

cook@eecs.wsu.edu

**Abstract.** The data mining and pervasive computing technologies found in smart homes offer unprecedented opportunities for providing context-aware services, including health monitoring and assistance to individuals experiencing difficulties living independently at home. In order to provide these services, smart environment algorithms need to recognize and track activities that people normally perform as part of their daily routines. However, activity recognition has typically involved gathering and labeling large amounts of data in each setting to learn a model for activities in that setting. We hypothesize that generalized models can be learned for common activities that span multiple environment settings and resident types. We describe our approach to learning these models and demonstrate the approach using eleven CASAS datasets collected in seven environments.

**Keywords.** machine learning, ubiquitous computing, pervasive computing, activity recognition

## Introduction

A convergence of technologies in data mining and pervasive computing as well as the increased accessibility of robust sensors and actuators has caused interest in the development of *smart environments* to emerge. Furthermore, researchers are recognizing that smart environments can assist with valuable functions such as remote health monitoring and intervention. The need for the development of such technologies is underscored by the aging of the population, the cost of formal health care, and the importance that individuals place on remaining independent in their own homes.

To function independently at home, individuals need to be able to complete Activities of Daily Living (ADLs) such as eating, grooming, cooking, drinking, and taking medicine. Automating the recognition of activities is an important step toward monitoring the functional health of a smart home resident. When surveyed about assistive technologies, family caregivers of Alzheimer's patients ranked activity identification and tracking at the top of their list of needs [1].

In response to the recognized need for smart environments to provide context-aware services, researchers have designed a variety of approaches to model and recognize activities. The generally accepted approach is to model and recognize those activities that are frequently used to measure the functional health of an individual. Recognizing resident activities also allows the smart environment to respond in a context-aware way to needs for achieving more comfort, security, and energy efficiency. A typical home may be equipped with hundreds or thousands of sensors. Because the captured data is rich in structure and voluminous, the activity learning problem is a challenging one. Traditionally, each environmental situation has been treated as a separate context in which to perform learning. What can propel research in smart environments forward is the ability to leverage experience of previous situations in new environments or with new residents.

When humans look at video or pictures of residents performing common activities such as eating and sleeping, these activities are recognized immediately, even if the observer has never seen the environment before and never met the residents. We therefore hypothesize that general models of activities can be learned that abstract over specific environments and residents. In this paper, we explore the use of supervised and semi-supervised machine learning algorithms to learn setting-generalized activity models. We evaluate these methods using datasets from the CASAS Smart Home project [2].



Figure 1. Sensor layout for the seven CASAS smart environment testbeds: Bosch1 (first row left), Bosch2 (first row right), Bosch3 (second row left), Cairo (second row right), Kyoto (third row left), Milan (third row right), and Tulum (bottom row).

## 1. Datasets

To test our ideas, we analyze eleven separate sensor event datasets collected from seven physical testbeds, shown in Figure 1. As can be seen in Table 1, the datasets exhibit a great deal of diversity. In addition, because some of the residents were younger adults, some were healthy older adults, some were older adults with dementia, and some were pets, the activities exhibit a great deal of diversity. This makes our goal of learning models to recognize activities across all of these settings even more challenging. Most of these datasets are available on the CASAS web page ([ailab.wsu.edu/casas](http://ailab.wsu.edu/casas)).

Table 1. Characteristics of the eleven datasets used for this study.

Dataset	Bosch1	Bosch2	Bosch3	Cairo	Kyoto1	Kyoto2
Environment	Bosch1	Bosch2	Bosch3	Cairo	Kyoto	Kyoto
#Residents	1	1	1	2+pet	2	2
#Sensors	32	32	32	27	51	71
#Activity occurrences	5,714	4,320	3,361	600	37	497
Dataset	Kyoto3	Kyoto4	Milan	Tulum1	Tulum2	
Environment	Kyoto	Kyoto	Milan	Tulum	Tulum	
#Residents	3	3	1+pet	2	2	
#Sensors	86	72	32	20	20	
#Activity occurrences	1,342	844	1,513	1,431	166	

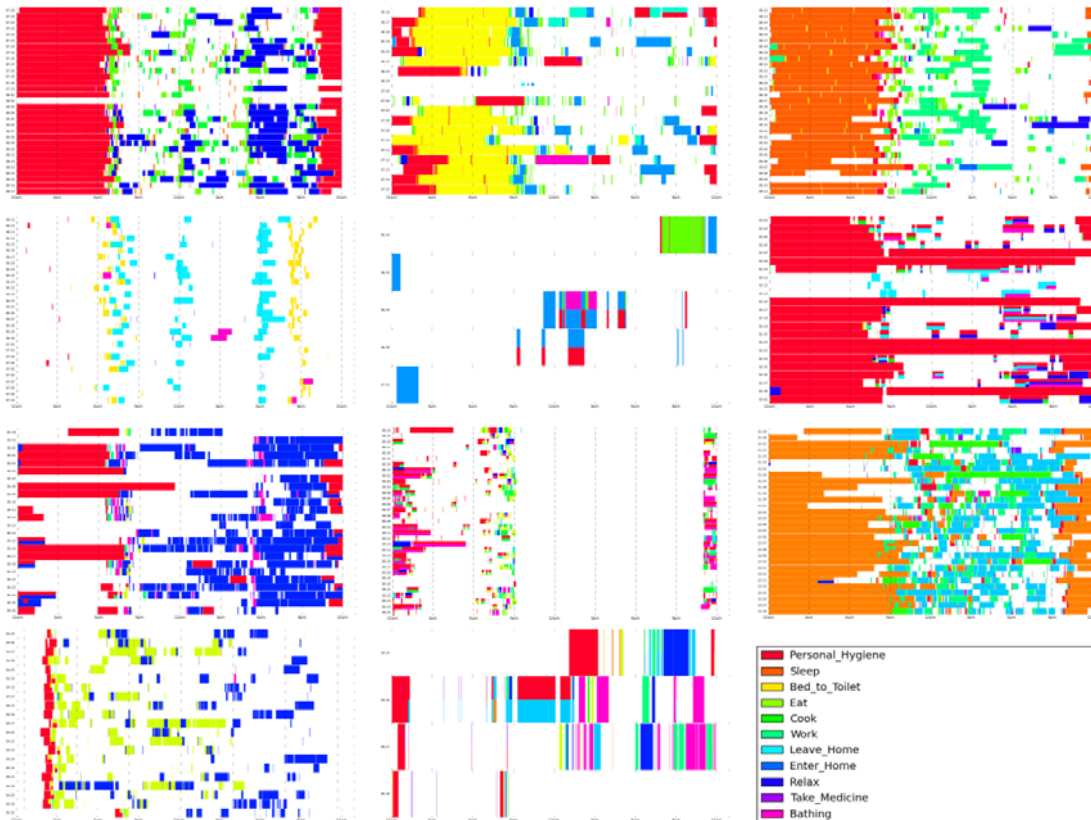


Figure 2. Activity recognition charts for a single month of data from the eleven datasets: Bosch1 (row 1 left), Bosch2 (row 1 middle), Bosch3 (row 1 right), Cairo (row 2 left), Kyoto1 (row 2 middle), Kyoto2 (row 2 right), Kyoto 3 (row 3 left), Kyoto 4 (row 3 middle), Milan (row 3 right), Tulum1 (row 4 left), and Tulum2 (row 4 middle).

Sensor data for each of the environments and datasets is captured using a sensor network that was designed in-house and the data is stored in a SQL database. Our middleware uses a jabber-based publish/subscribe protocol as a lightweight platform and language-independent middleware to push data to client tools with minimal overhead and maximal flexibility. Although each of the datasets was originally monitored for a large number of activities, for this study we are interested in learning abstract models for eleven ADL activities that occur in a majority of the datasets: Personal Hygiene, Sleep, Bed-to-toilet, Eat, Cook, Work, Leave Home, Enter Home, Relax, Take Medicine, and Bathing. These activities are frequently used to measure the functional health of an individual [3]. Figure 2 shows a graph of the occurrences of these activities for a one month period in each of the datasets.

## 2. ADL Recognition

We treat a smart environment as an intelligent agent that perceives the state of the resident and the physical surroundings using sensors and acts on the environment using controllers in such a way that the specified performance criteria is optimized [4]. Researchers have designed smart environment algorithms that track the location and activities of residents, that generate reminders, and that react to hazardous situations. Resulting from recent advances, researchers are now beginning to recognize the importance of applying smart environment technology to health assistance and companies are recognizing the potential of this technology for a quickly-growing consumer base.

Activity recognition is not an untapped area of research. Because the need for activity recognition technology is great, researchers have explored a number of approaches to this problem. Researchers have found that different types of sensor information are effective for classifying different types of activities. When trying to recognize body movements (e.g., walking, running), data collected from accelerometers positioned on the body has been effective [5]. Other activities are not as easily distinguishable by body position. In these cases, researchers [6] observe the smart home resident's interaction with objects of interest such as doors, windows, refrigerators, keys, and medicine containers. Other researchers [7] rely upon motion sensors as well as item sensors to recognize ADL activities that are being performed. We note here that while the current study utilizes primarily motion and door sensors, the approach can be applied to a much greater range of sensor types.

The number of machine learning models that have been used for activity recognition varies almost as greatly as the types of sensor data that have been tested. Some of the most commonly-used approaches are naïve Bayes classifiers, decision trees, Markov models, and conditional random fields [5][6][7][8]. In our approach we initially test three models: a naïve Bayes classifier (NBC), a hidden Markov model (HMM), and a conditional random field (CRF) model. These three approaches are considered because they traditionally are robust in the presence of a moderate amount of noise, are designed to handle sequential data, and generate probability distributions over the class labels. These features are all useful for our task. However, among these three choices there is no clear best model to employ – they each employ methods that offer strengths and weaknesses for the task at hand.

The NBC uses the relative frequencies of feature values (the length of the activity, the previous activity, and the sensor event frequencies) as well as the frequency of activity labels found in the sample training data to learn a mapping from activity features,  $D$ , to an activity label,  $a$ , calculated using the formula  $\arg \max_{a \in A} P(a|D) = P(D|a)P(a)/P(D)$ .

The HMM is a statistical approach in which the underlying model is a stochastic Markovian process that is not observable (i.e., hidden) which can be observed through other processes that produce the sequence of observed (emitted) features. In our HMM we let the hidden nodes represent activities. The observable nodes represent combinations of the features described earlier. The probabilistic relationships between hidden nodes and observable nodes and the probabilistic transition between hidden nodes are estimated by the relative frequency with which these relationships occur in the sample data. An example HMM for three of the activities is shown in Figure 3. Given an input sequence of sensor events, our algorithm finds the most likely sequence of hidden states, or activities, which could have generated the observed event sequence. We use the Viterbi algorithm [9] to identify this sequence of hidden states.

Like the HMM, the CRF model makes use of transition likelihoods between states and emission likelihoods between activity states and observable states to output a label for the current data point. The CRF learns a label sequence,  $A$ , which corresponds to the observed sequence of features. Unlike the HMM, weights are applied to each of the transition and emission features. These weights are learned through an expectation maximization process based on the training data [10].

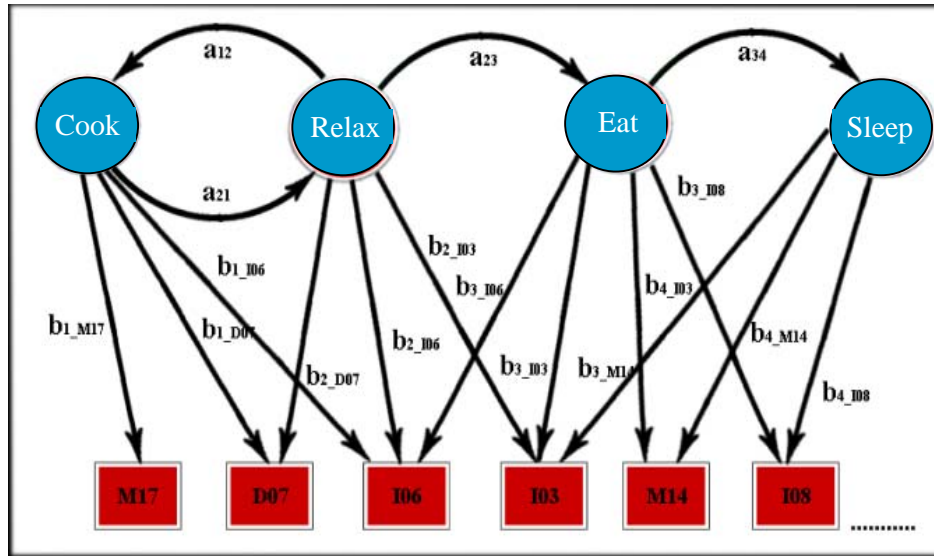


Figure 3. HMM for an activity recognition task with four hidden states (activities) and a set of observable nodes that correspond to possible sensor events. The  $a$  values represent the transition probabilities and the  $b$  values represent emission probabilities.

Table 2 summarizes the recognition accuracy of the three models for each of the eleven datasets, calculated using 3-fold cross validation and averaged over all of the activities. As the table indicates, the accuracy varies dramatically between datasets. The accuracy also varies between individual activities and is affected by the amount of available data, the quality of the labels that were provided for the data, the number of residents in the space that are interacting and performing activities in parallel, and the consistency of the activities themselves.

Table 2. NBC, HMM, and CRF recognition accuracies for each of the eleven datasets.

Dataset	NBC	HMM	CRF
Bosch1	<b>92.91%</b>	92.07%	85.09%
Bosch2	<b>90.74%</b>	89.61%	82.66%
Bosch3	88.81%	<b>90.87%</b>	90.36%
Cairo	<b>82.79%</b>	82.41%	68.07%
Kyoto1	78.38%	78.38%	<b>97.30%</b>
Kyoto2	63.98%	65.79%	<b>66.20%</b>
Kyoto3	77.50%	81.67%	<b>87.33%</b>
Kyoto4	<b>63.27%</b>	60.90%	58.41%
Milan	76.65%	<b>77.44%</b>	61.01%
Tulum1	59.05%	75.12%	<b>79.45%</b>
Tulum2	<b>66.87%</b>	57.83%	39.76%

### 3. Abstracting Activity Models

One approach to learning a setting-generalized activity model is to combine sensor events from all of the environments into one dataset. The first step in generalizing these models is to create a uniform sensor label. Instead of using all of the original sensor IDs, which carry different meanings in each different setting, we map them onto labels corresponding to the room in which the sensor resides: Bathroom, Bedroom, Kitchen, LivingRoom, WorkArea, MedCabinet, and Lounge. We further differentiate the type of sensor: motion, door, or other. The result of applying the three models to this combined dataset is 74.87% for the naïve Bayes classifier, 75.05% for the hidden Markov model, and 72.16% for the conditional random field using 3-fold cross validation over the set of annotated activities.

This result indicates that it is possible to find general patterns for ADL activities across multiple environment and resident settings. However, the experimental approach does not reflect a real-life situation. In a real deployment of this technology a user would set up a new smart home and use activity models learned from other setting to immediately start recognizing activities in their new setting. The appropriate testing for this scenario is a leave one out experiment where activity models are trained on ten of the datasets, then tested on the left-out dataset. Table 3 summarizes the result of this experiment, where the accuracies are averaged over all of the eleven test datasets. As the summary indicates, activity recognition performance fluctuates between datasets and is much lower than when training data is provided specific to the testing environment.

Table 3. Leave one out, ensemble, and semi-supervised results for the eleven datasets.

Dataset	Leave One Out	Ensemble	Semi-Supervised
<b>Bosch1</b>	67.75%	<b>89.76%</b>	89.18%
<b>Bosch2</b>	73.52%	<b>85.37%</b>	85.32%
<b>Bosch3</b>	64.74%	<b>88.46%</b>	88.07%
<b>Cairo</b>	46.08%	<b>68.07%</b>	64.82%
<b>Kyoto1</b>	37.84%	<b>59.56%</b>	<b>59.56%</b>
<b>Kyoto2</b>	33.20%	62.78%	<b>62.98%</b>
<b>Kyoto3</b>	36.89%	<b>100.00%</b>	<b>100.00%</b>
<b>Kyoto4</b>	37.32%	<b>100.00%</b>	<b>100.00%</b>
<b>Milan</b>	53.39%	76.39%	<b>78.24%</b>
<b>Tulum1</b>	38.16%	34.73%	<b>32.84%</b>
<b>Tulum2</b>	36.14%	39.16%	<b>40.96%</b>

We note from the earlier experiments that the best-performing recognition model varies from one dataset to another (see Table 2). Factors that influence this outcome are the size of the environment, the number of residents (and thus the amount of activity interleaving that occurs), and the type of activity that is being performed. In order to harness the power of each of these models, our second approach to constructing an abstract activity model is to construct an ensemble of classifiers [11]. The base classifiers for this ensemble are the NBC, HMM, and CRF models and we use a boosted decision tree for the top classifier. The input features to the top classifier are the probability distributions that the three base models output for each of the activity label possibilities and the activity feature values. In addition, dataset-descriptive features are also input including the size of the environment (small, medium, or large) and the number of residents (one, two, or three).

As we can see from Table 3, the ensemble method greatly boosts the accuracy of the classifier for all of the datasets except Tulum1. In fact, for several of the datasets the accuracy is close to the value that is achieved when training data is available from the test environment, and in the Kyoto3 and Kyoto4 datasets the accuracy is actually far better than when only data from the test set is available. These results indicate that abstract activity models can be learned that generalize over multiple environment and resident situations. They also indicate that activity models can in fact be

strengthened by providing data available from other sources outside one particular environment and selection of residents.

In our final approach, we consider a semi-supervised learning method [12]. This approach is inspired by the observation that our ensemble learning approach might actually benefit by having access to data collected in a new setting, even if it is unlabeled data. To make use of the unlabeled data, we first use the ensemble classifier described above to label the new data in the test set. Then we add the newly-labeled data to the training set and reconstruct the classifier with the larger set of labeled data. We iteratively evaluate the new model on the test data and see if accuracy has improved. As is shown in Table 3 the performance does sometimes improve over the original ensemble method. This is not always the case, however. In general, the datasets which are the largest and offer the best original recognition accuracy do not benefit from semi-supervised learning, likely because the erroneous labels generated from the other datasets are reinforced when they applied to the test set and integrated into the model. In contrast, the smaller datasets and those which originally yielded low recognition accuracy (as shown in Table 2) benefitted the most from the semi-supervised approach.

A few additional observations can be made about this study. First, the decision tree that is created from the ensemble and semi-supervised methods utilizes almost all of the base features in the top levels of the tree. This indicates that no single base classifier performs consistently best. It also indicates that the environment description features are useful when deciding which base model to consult when selecting an activity label.

Second, we note that the ensemble methods in general offered a significant ( $p < .005$ ) improvement over the original leave one out method, yielding a 25.56% improvement on average. In contrast, the decrease in accuracy from the original models (shown in Table 2) to the ensemble method is less significant ( $p < 0.600$ ) and represents an average performance decrease of only 3.26%.

#### 4. Conclusions

In order to provide context-aware services such as health monitoring and intervention, smart environment designers need to design robust methods for recognizing ADL activities in these smart spaces. In the past, activity recognition methods have required that training data be collected and labeled in each new environment or even in an existing environment with new residents. In this paper we proposed that general activity models could be learned that abstract over these differences in environment and resident differences.

Of the approaches that we considered, the fully-supervised and semi-supervised ensemble methods performed best. These approaches made use of a variety of information including the base classifier output probability distributions and features of the environments themselves. The results from the experiments indicate that activity recognition in a new setting can be accomplished by generating a model specific to the setting or by using the setting-generalized model with almost the same accuracy. In some cases, the generalized model actually outperforms the setting-specific model due to the increased availability of training data for a particular type of activity. These results are encouraging and offer users the possibility of employing smart home technologies “out of the box” with very little or no training.

Ultimately, we want to use our algorithm design as a component of a complete system that performs functional assessment of adults in their everyday environments. This type of automated assessment also provides a mechanism for evaluating the effectiveness of alternative health interventions. We believe these activity profiling techniques are valuable for providing automated health monitoring and assistance in an individual’s everyday environments.

Evidence from our experiments suggests that this approach to activity recognition will effectively generalize to settings with new floor plans, new sensor layouts, and new residents. This generalization is very useful for applying activity recognition algorithms in new settings with little or no training data. The type of generalization we have not considered is generalization to new, similar activities. This can be considered in the future.

We also hypothesize that utilizing ensemble classifiers with abstracted features can be applied to other types of learning problems as well. Future work can investigate applying this type of abstracted learning problem to classification of sequences in transaction data or in gene data.

## 5. Acknowledgements

The author would like to thank Larry Holder and Brian Thomas for their contributions to this work. This material is based upon work supported by the National Science Foundation under Grant Number 0852172, by the Life Sciences Discovery Fund, and by NIBIB Grant Number R01EB009675.

## References

- [1] V. Rialle, C. Ollivet, C. Guigui, and C. Herve. What do family caregivers of Alzheimer's disease patients desire in smart home technologies? *Methods of Information in Medicine*, 47:63-69, 2008.
- [2] 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*, 39(5):949-959, 2009.
- [3] V. Wadley, O. Okonkwo, M. Crowe, and L. A. Ross-Meadows. Mild cognitive impairment and everyday function: Evidence of reduced speed in performing instrumental activities of daily living, *American Journal of Geriatric Psychiatry*, 16:416-424, 2007.
- [4] D. Cook and S. Das, eds. *Smart Environments: Technology, Protocols, and Applications*, Wiley, 2004.
- [5] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions, *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, pages 99-102, 2004.
- [6] E. Munguia-Tapia, S.S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors, *Proceedings of Pervasive*, pages 158-175, 2004.
- [7] D. Cook and M. Schmitter-Edgecombe. Assessing the quality of activities in a smart environment, *Methods of Information in Medicine*, 48(5):480-485, 2009.
- [8] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational Markov networks, *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 773-778, 2005.
- [9] A. Viterbi. Error bounds for convolutional codes and an asymptotically-optimum decoding algorithm, *IEEE Transactions on Information Theory*, 13(2):260-269, 1967.
- [10] A. McCallum. Efficiently inducing features of conditional random fields. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 403-410, 2003.
- [11] C.-X. Zhang and J.-S. Zhang. A local boosting algorithm for solving classification problems. *Computational Statistics and Data Analysis*, 52(4):1928-1941, 2008.
- [12] Q. Yang, S. Pan, and V. Zheng. Estimating location using wi-fi. *IEEE Intelligent Systems*, 23(1):8-13, 2008.



## ***Biography***

Dr. Diane J. Cook is a Huie-Rogers Chair Professor in the School of Electrical Engineering and Computer Science at Washington State University. Dr. Cook received a B.S. degree in Math/Computer Science from Wheaton College in 1985, a M.S. degree in Computer Science from the University of Illinois in 1987, and a Ph.D. degree in Computer Science from the University of Illinois in 1990. Her research interests include artificial intelligence, machine learning, graph-based relational data mining, smart environments, and robotics. Dr. Cook is an IEEE Fellow.



## ***Contact Information***

Diane Cook  
School of Electrical Engineering and Computer Science  
Washington State University  
Box 642752  
Pullman, WA 99164-2752  
(509) 335-4985  
cook@eecs.wsu.edu