Enhancing Activity Recognition in Smart Homes Using Feature Induction

Naveen Nair[1,2,3], Ganesh Ramakrishnan[2,1], and Shonali Krishnaswamy[3,1]

[1] IITB-Monash Research Academy, Old CSE Building, IIT Bombay
[2] Department of Computer Science and Engineering, IIT Bombay
[3] Faculty of Information Technology, Monash University
{naveennair,ganesh}@cse.iitb.ac.in,
Shonali.Krishnaswamy@infotech.monash.edu.au

Abstract. Hidden Markov Models (HMMs) are widely used in activity recognition. Ideally, the current activity should be determined using the vector of all sensor readings; however, this results in an exponentially large space of observations. The current fix to this problem is to assume conditional independence between individual sensors, given an activity, and factorizing the emission distribution in a naive way. In several cases, this leads to accuracy loss. We present an intermediate solution, viz., determining a mapping between each activity and conjunctions over a relevant subset of dependent sensors. The approach discovers features that are conjunctions of sensors and maps them to activities. This does away the assumption of naive factorization while not ruling out the possibility of the vector of all the sensor readings being relevant to activities. We demonstrate through experimental evaluation that our approach prunes potentially irrelevant subsets of sensor readings and results in significant accuracy improvements.

Keywords: Activity recognition, Feature induction, Hidden Markov Model.

1 Introduction

Efficient activity recognition has been a hot topic for researchers since the advent of Artificial Intelligence. An active application area for activity recognition is monitoring elderly activities in homes to ensure their well being. To monitor activities of large number of people living alone in houses and thereby detect unusual patterns of activities, which are indicators of their health condition, automatic activity recognition systems are used [1]. In the following paragraph, we provide a brief overview of a typical activity recognition setting.

A typical minimal intrusive activity recognition setting has on/off sensor devices, which senses user movements or object use, fixed at various locations inside a house. The nodes send binary (sensed) information to a base station/computer. A probabilistic model is trained initially using the information from the sensors and the user annotated information about the activities performed. Later, the

A. Cuzzocrea and U. Dayal (Eds.): DaWaK 2011, LNCS 6862, pp. 406–418, 2011.

model is used for predicting activities performed, based on sensor readings. The Hidden Markov Model (HMM) [5] and the Conditional Random Field (CRF) [6] are two popular probabilistic models used for activity recognition [4]. Kasteren et. al. in [4] reported better class accuracy (average percentage of time a class is predicted correctly) with HMM and better time slice accuracy (fraction of time slices predicted correctly) with CRF. We, in this paper, use HMM as the base line approach and investigate the effectiveness of our approach in HMM. We now give a brief introduction to HMM.

An HMM formulation consists of a hidden variable and an observable/emission variable at each time step. In activity recognition, the activity at each time step t is the label/hidden state y_t. The joint state of the sensors at each time step is the observation x_t. The hidden state y_t at time t is independent of all other variables given the previous hidden state y_{t-1} at time $t-1$ and the observable variable x_t at time t is independent of other variables given y_t. Using these independence assumptions, we can factorize the joint distribution of the sequence of observations (X) and labels (Y) into three factors: the initial state distribution $p(y_1)$, the transition distribution $p(y_t|y_{t-1})$, and the emission distribution $p(x_t|y_t)$ [10][5]. There fore, the joint distribution can be expressed as $p(X,Y) = \prod_{t=1}^{T} p(y_t|y_{t-1})p(x_t|y_t)$, where $p(y_1|y_0)$ is used instead of $p(y_1)$ to simplify notation. Parameters for the distributions are learned by maximizing the joint probability, $p(X,Y)$, of the paired observation and label sequences in the training data. During the inference phase, the parameters are used to determine the sequence of labels that best explains the given sequence of observations. This is efficiently computed using a dynamic programming algorithm called the Viterbi Algorithm [11]. We now discuss the limitations of traditional HMM in the domain of activity recognition.

In a traditional HMM setting for activity recognition, the observation value is a vector of all the sensor readings in the deployment. Thus there are 2^N possible values for the observation variable in a binary sensor deployment with N sensors, which is computationally expensive for learning and inference in real world settings. Typical approaches in activity recognition tend to assume independence between individual sensors given activity to perform naive factorization of the emission distribution [4]. Since the independence assumption is wrong in most of the real world problems, the method suffers from accuracy loss in many cases. More over, the binary sensor values of all the sensors have to be considered for all the activities in every inference step, which is an overhead in large settings. Now we provide a brief introduction to the proposed solutions to these limitations.

Since strong assumptions of dependence or independence among entire set of sensors given activity have their own limitations in activity recognition domain, we identify the need to find a mapping between activities and their relevant subsets of dependent sensors. Manual imposition of such a mapping is neither novel nor feasible in large settings. An efficient feature induction approach that can automatically capture the mapping between activities and conjunctions of sensors can be used. Inductive Logic Programming (ILP), a branch of machine learning, is a learning paradigm capable of learning such mappings or rules.

Given some background knowledge and a set of facts as examples, ILP systems derive hypothesis (structure) that entails all the positive examples and none of the negative examples. It starts with an initial hypothesis and refines it by searching a lattice of clauses (a partially ordered set of clauses or rules) based on a scoring function. Typical structure learning systems that do a Branch and Bound (B&B) search in the lattice of clauses evaluating scores based on positive and negative examples covered, when used to construct features for HMM in activity recognition, suffer from accuracy loss as shown in our experiments. In this paper, we propose and implement a greedy feature induction approach that adapts the structure of HMM using HMM evaluation on the training set as scoring function. Our experimental results suggest a performance improvement over both the traditional HMM and the B&B learning assisted HMM in terms of accuracy. We also show the statistical significance of our proposed approach against the traditional approaches.

The rest of the paper is organized as follows. We discuss some related works in section 2. Section 3 discusses about using feature induction to assist HMM model construction. Experiments and Results are discussed in section 4. We conclude our work in section 5.

2 Related Work

In this section, we look into some of the related works in both the area of activity recognition and feature induction.

Automatic activity recognition has been an active research area in the current era of pervasive systems. Various approaches have been proposed. Wilson experimented with particle filter and context aware recognition for recognizing ADLs at the MIT Laboratory [1]. Gibson et.al. [2] discussed the idea of clustering sensors for recognizing activities and concluded that trivially imposing clusters differs from reality. A relational transformation based tagging system using ILP concepts is proposed in [13]. The approach starts with an initial tag to all the sequences and then improves by learning a list of transformation rules which can re-tag based on context information. The approach is purely logical and not probabilistic. [3] identifies the minimal set of sensors that can jointly predict all activities in the domain. Binsztok et. al. [14] discussed learning HMM structure (number of states and allowed transitions) from data for clustering sequences. An efficient feature induction method for named entity extraction and noun phrase segmentation tasks using CRFs is presented by McCallum [15]. Landwehr et. al., in [7], construct kernel functions from features induced by an ILP approach. The search for features is directed by a Support Vector Machine performance using the current kernel. [9] aims to classifying relational sequences using relevant patterns discovered from labelled training sequences. Here, the whole sequence is labelled and not the individual components of the sequence. Patterns in each dimension of multi dimensional sequences are discovered and a feature vector is constructed. Then an optimal subset of the features is selected using a stochastic local search guided by a naive Bayes classifier. TildeCRF, an extension to CRF, is introduced in [8] to deal with sequences of logical atoms and to model arbitrary dependencies in the input space. The potential functions are represented as weighted sums of relational regression trees.

Many of the learning approaches discussed above suits for general classification. However, in the case of sequential, skewed, and sparse activity recognition data where temporal dependencies dominate over static dependencies, most of the learning approaches that globally normalize the parameters do not fit well. We find a solution to this problem by identifying relevant conjunctions of sensors for each activity as observation for HMM. We then learn conditional probability values for this emission model and combine it with transition distribution. We propose feature induction assisted HMM model construction which we discuss in the following section.

3 Model Construction for Activity Recognition

In this section, we first give a technical explanation to the problem at hand before we discuss the B&B structure learning assisted HMM model construction and the feature induction assisted HMM model construction for activity recognition.

In an HMM set-up, the probability distribution of observation given label, $p(x_t|y_t)$, is represented as an emission matrix. Here the observation vector is $x_t = (x_t^1, x_t^2, ..., x_t^N)^\top$, where x_t^i represents the value of i^{th} sensor at time t and N is the number of sensors. Considering the entire set of sensors results in 2^N values for observation x_t which is computationally feasible only in small settings. Often independence is assumed among sensors, given activity, to simplify the representation and computation of $p(x_t|y_t)$. Conditional probability, when independence is assumed among sensors, is $p(x_t|y_t) = \prod_{i=1}^{N} p(x_t^i|y_t)$. This approach is prone to accuracy loss in many cases where the independence assumption is wrong. To alleviate both the issues, we identify the need to find a mapping between activities and their relevant conjunctions of dependent sensors. Our work underlines the notion that if a few dependent sensors in conjunction with information regarding the previous activity can jointly decide on whether an activity has happened in the current time, then it is better to consider only these conjunctions of sensor readings. That is, we avoid the non relevant x^is and use conjunctions of relevant x^js to improve the prediction accuracy. This also helps to reduce the effect of noise while doing inference.

We propose learning the HMM emission structure that maximizes probabilistic coverage of the training data. In our problem, since there is no ordering among activities, the model learned should allow all inter state transitions. Therefore, we learn the structure of emission distribution while preserving all the $\binom{n}{2}$ transition probabilities. In the next subsection, the B&B structure learning assisted HMM model construction is discussed.

3.1 B&B Structure Learning Assisted HMM for Activity Recognition

In this subsection, we explore the idea of using the B&B structure learning assisted HMM model construction (B&BHMM) for activity recognition.

As discussed above, we are interested in finding a mapping between activities and relevant conjunctions of sensors. The mapping can be expressed as relationships of the form "Activity if a particular set of sensors fired". These type of rules are called definite clause rules and are represented in the form $A \leftarrow B, C, \ldots$ where $A, B, C, \ldots$ are binary predicates. Traditional structure learning systems are capable of discovering rules of the above form. We now analyze the effectiveness of these systems to construct HMM model for activity recognition.

We hypothesize that traditional structure learning systems that do not do HMM evaluation while refining rules learned in each step of rule induction will have reduced impact on the accuracy of prediction. This is because, in traditional systems, the objective is to logically cover all the positive examples. ILP is one of the traditional structure learning paradigms that learn first order relations among entities. For example, Aleph [12] is an ILP system that in each iteration, selects a positive example, builds the most specific clause based on the example, searches for a more general clause that has the best score, and removes examples made redundant by the current clause. Although the current problem does not require learning complex first order structures, we use Aleph as a benchmark system for our experiments.

The scores used by traditional systems such as Aleph are largely based on the number of positive and negative examples covered by the current model. One of the scoring functions is $pos - neg$, where pos and neg are the number of positive and negative examples covered by the clause respectively. Discovery of each of the clauses leads to the removal of positive examples covered.

Since in real world problems, the support for any emission of an activity and the support for inter state transitions are much fewer than that for same state transitions, in learning both the emission and transition dependencies using traditional systems, rules defining transitions within the same state tend to dominate. Such a model tends to predict fewer inter state transitions, and thus affects the accuracy of inference. Hence we focus only on the induction of emission rules and combine them with the set of $\binom{n}{2}$ interstate transitions while learning the parameters of the model. We first study the applicability of the B&B structure learning systems to learn emission rules and identify the limitations.

B&B systems, when used for learning emission rules, evaluate each refinement of clauses using scoring functions based on positive and negative examples. Since the real world data are vulnerable to noisy information, an exact model is hard to get. As the examples covered by a refinement are removed in each step, rules that are learned in subsequent iterations have less confidence than those learned initially, which leads to a less efficient model. Since the objective of traditional systems is to logically cover all positive examples with clauses which is different from the actual objective of building a probabilistic model (HMM), the approach suffers from accuracy loss significantly. We have experimented with this approach using Aleph combined with a customized implementation of HMM. Each rule returned by aleph is a definite rule, which associates a subset of sensors to an activity. A new attribute (feature) is constructed with each such subset. Therefore, the number of attributes equals the number of rules learned. The learned logical model and the training data are passed to a customized implementation of HMM for constructing the probabilistic model. Later, the probabilistic model is used for inference. Our experiments reveal that, HMM with B&B structure learning for feature construction is less efficient than HMM without structure learning, except in a few cases. Although of-the-shelf branch and bound structure learning system assisted HMM gave comparable time slice accuracies in a few experiments, it gave worse class accuracies in all the experiments. The comparison of time slice and average class accuracies are shown in tables 1 and 2 respectively as well as figures 2 and 3 respectively. We now discuss the feature induction assisted HMM model construction for activity recognition.

3.2 Feature Induction Assisted HMM for Activity Recognition

After analyzing the limitations of branch & bound structure learning using $pos-neg$ to assist HMM model construction, we propose a greedy hill climbing feature induction approach wherein we evaluate, in each refinement step, the current model in an HMM setting. That is, the score which has to be maximized is an HMM evaluation on the training data. We call this approach the Feature Induction assisted HMM model construction (FIHMM). The score can be either time slice accuracy or class accuracy of the current model. Time slice accuracy is the fraction of time slices when classes (activities) are predicted correctly and average class accuracy is the average percentage of time a class is classified correctly as given in the expressions reproduced below from [4].

$$TimesliceAccuracy : \frac{\sum_{n=1}^{N} [inferred(n) = true(n)]}{N} . \quad (1)$$

$$ClassAccuracy : \frac{1}{C} \sum_{c=1}^{C} \left\{ \frac{\sum_{n=1}^{N_c} [inferred_c(n) = true_c(n)]}{N_c} \right\} . \quad (2)$$

where $[a = b]$ is an indicator giving 1 when true and 0 otherwise, N is the total number of time slices, C is the number of classes and N_c is the number of time slices for the class c.

In data that are skewed towards some activities, predicting a frequent activity for all the time slices gives better time slice accuracy but worse average class accuracy. Therefore, if the data set is skewed and some critical activities have less support, we suggest maximizing the average class accuracy. In all other cases, we suggest maximizing time slice accuracy. This is because the average class accuracy computation does not consider the size of a particular class and its maximization leads to a situation where unimportant classes that occur seldom have more impact on the overall efficiency of the model. We pursued both the cases in different experiments, and the results are given in the experiments section. Trying a combination of both time slice and average class accuracies is a future work direction. Leaving the choice of one of these accuracy values to the user, we now discuss the overall learning algorithm for model construction.

1. **procedure** FIHMM_MODEL_CONSTRUCTION
2. $featureSet \leftarrow$ features representing each sensor
3. $currentModel \leftarrow$ model trained with $featureSet$
. $\triangleright$ Here model is synonymous to HMM model
4. **repeat**
5. $previousModel \leftarrow currentModel$
6. **for each** activity i **do**
7. **for each** feature j of activity i **do**
8. $modelDel(i,j) \leftarrow$ model trained with j^{th} feature of i^{th} activity dropped
9. **for each** feature k of activity i **do**
10. $modelAdd(i,j,k) \leftarrow$ model trained with features j and k combined
. to form new feature of activity i
11. **end for**
12. **end for**
13. **end for**
14. $currentModel \leftarrow \arg\max \{\arg\max_{i,j} \; modelDel(i,j).accuracy,$
. $\arg\max_{i,j,k} \; modelAdd(i,j,k).accuracy\}$
15. **until** $currentModel.accuracy \leq previousModel.accuracy$
16. **return** $previousModel$
17. **end procedure**

Fig. 1. Feature induction assisted HMM model training for activity recognition

During the training phase, we pursue a greedy hill climbing search in the lattice to find a model. The pseudo code for our approach is given in Fig. 1. Initially, the features for each activity are constructed with each of the individual sensors and an initial model is trained. In every iteration, candidate models are constructed by removing features of each activity one at a time as shown in step 8 of the pseudo code. Step 10 constructs new features by combining the features removed in step 8 with other features of the activity and a new candidate model is trained. The best scoring model among all the candidate models, if better than the previous model, is saved. To evaluate a model, an HMM is constructed from the current emission model and the transition distribution. Each of the conjunctions discovered forms a column in the emission probability matrix and the conditional probabilities are learned for these conjunctions given activity. Further, only those columns that are mapped to an activity have to be considered during inference. Each iteration either deletes or adds a feature to the final model based on the HMM evaluation on training data. Unlike in the traditional approaches, no examples are removed during the iterations. In each iteration, the existing logical model is refined, probabilistic parameters are learned and the model is evaluated on the training data. The process is repeated until convergence. In the next section, we describe our experimental set-up and report our results.

4 Experiments and Results

We have implemented all the approaches in java. All our experiments have been performed on an AMD Athlon 64 bit dual core machine (2.90 GHz) with 2.8 GB RAM and running Ubuntu 8.04.

We have carried out our experiments on the data set made available by Kasteren et. al. [4] of the University of Amsterdam. The dataset consists of binary values reported at each time interval by 14 sensors installed at various locations in a house. There are 8 activities annotated. The data is marked for each one minute time slot and there are 40006 instances. In the dataset, some activities occurred more frequently than others and some activities occurred for a longer duration, and hence the data is not balanced. The data is represented in four binary formats: raw, change point, last observation, and a combination of change point and last observation. Interested readers may refer [4] for more details.

We assume the data is complete in our case. Moreover, the use of discrete data enables us to count the number of occurrences of transitions, observations and states [5]. We have performed our experiments in a leave one day out manner in a 28 fold cross validation set-up. The performance is evaluated by the average time slice accuracy and the average class accuracy. We also evaluate the statistical significance of our claims.

We ran four experiments each on raw, change point, last value and change point plus last value data. The First experiment is the traditional HMM as suggested in [4]. The second experiment, B&BHMM, uses Aleph to learn emission rules in the form of definite clauses for each activity. These rules along with the data are passed to a customized implementation of HMM for probabilistic learning and inference. The third and fourth experiments are the proposed FIHMM which inductively learns HMM emission model using HMM evaluation as the score. The emission model is combined with the $\binom{n}{2}$ inter state transitions and the probabilities are learned to obtain the complete HMM model. The third experiment optimizes average class accuracy while the fourth experiment optimizes time slice accuracy. The results are shown in tables 1 and 2. The comparison of time slice accuracies and class accuracies for all the four approaches in all the data formats is shown in Fig. 2 and Fig. 3 respectively.

From the results, it can be noted that the average time slice accuracies of B&BHMM are not better than traditional HMM in any of the data formats except raw data and the average class accuracies (average of the class accuracies obtained from each fold) of B&BHMM are not better than traditional HMM in any of the data formats. This decline in the accuracies is due to the inappropriate evaluation function used by the branch & bound structure learning systems while doing refinement of learned clauses. In contrast, the proposed feature induction assisted HMM model construction approach that uses HMM evaluation as the scoring function performed better than the other two approaches in all the data formats significantly. The average accuracies of the proposed approach on training data in all the data formats are given in Table 3.

In raw data format, maximizing class accuracy yielded better class accuracy but did not yield better time slice accuracy than the B&B learning assisted

Table 1. Average time slice accuracies in percentage for various data representations using traditional HMM, B&B learning assisted HMM and proposed approach. Proposed approach has been used for maximizing class and time slice accuracies.

Data	**Traditional HMM**	**B&BHMM**	**FIHMM maximizing class accuracy**	**FIHMM maximizing time slice accuracy**
raw	50.49	56.94	54.98	71.59
change point	67.14	44.91	82.93	87.07
last value	86.45	33.69	89.67	93.47
change + last	86.55	64.94	91.15	93.57

Table 2. Average class accuracies in percentage for various data representations using traditional HMM, B&B learning assisted HMM and proposed approach. Proposed approach has been used for maximizing class and time slice accuracies.

Data	**Traditional HMM**	**B&BHMM**	**FIHMM maximizing class accuracy**	**FIHMM maximizing time slice accuracy**
raw	44.60	27.81	55.11	55.13
change point	61.68	27.21	75.93	68.03
last value	73.47	15.85	74.90	64.44
change + last	76.41	34.87	79.26	76.78

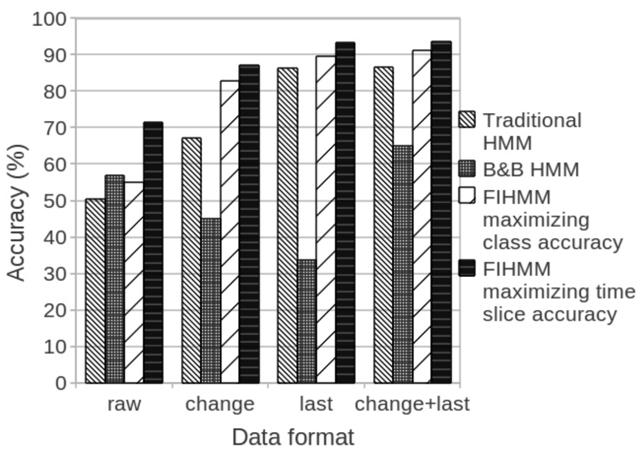

Fig. 2. Comparison of time slice accuracies of traditional HMM, B&BHMM, FIHMM maximizing class accuracy and FIHMM maximizing time slice accuracy on different data representations

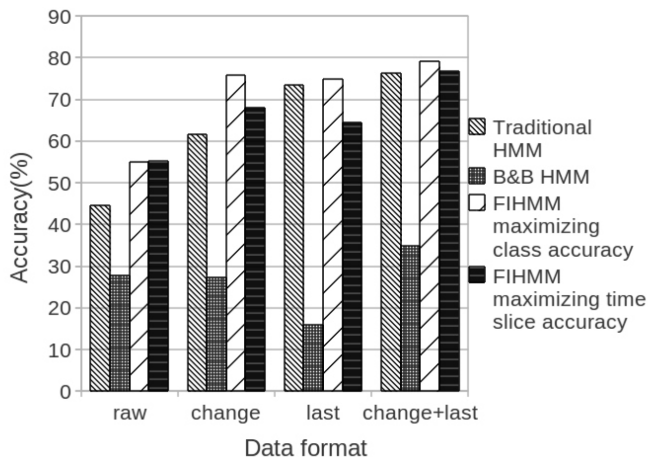

Fig. 3. Comparison of class accuracies of traditional HMM, B&BHMM, FIHMM maximizing class accuracy and FIHMM maximizing time slice accuracy on different data representations

Table 3. Average accuracies of proposed approach on training data in percentage

Data	**FIHMM maximizing class accuracy**		**FIHMM maximizing time slice accuracy**	
	Time slice accuracy	**Class accuracy**	**Time slice accuracy**	**Class accuracy**
raw	64.86	71.06	82.02	61.97
change point	87.59	82.88	93.25	69.13
last value	91.09	83.63	95.76	62.43
change + last	93.14	88.01	96.7	83.79

HMM. This is because the objective function maximized was class accuracy and not time slice accuracy. For similar reasons, the average class accuracy of proposed approach maximizing time slice accuracy in last value data is not better than traditional HMM. Therefore choosing appropriate objective function to maximize has an effect on the accuracies. The average confusion matrices got from the proposed approach using class accuracy optimization for each of the data representations are given in tables 4, 5, 6, 7.

The statistical significance of the performances is analyzed using the Wilcoxon signed rank test [16]. This non parametric test finds the probability of the null hypothesis that a pair of algorithms have no significant difference in their median performance. The test uses signed ranks of absolute values of the differences in performance after discarding the ties. The probabilities for the actual signed ranks are determined by exact computation if there are fewer entries or by normal approximation otherwise. In our experiments, we have evaluated two null

Table 4. Confusion matrix of FIHMM for raw data set

	Idle	Leaving	Toileting	Showering	Sleeping	Breakfast	Dinner	Drink
Idle	32	12	7	21	0	4	12	1
Leaving	11	59	4	11	0	0	15	0
Toileting	4	3	64	6	1	4	3	1
Showering	15	3	17	42	0	1	1	0
Sleeping	12	23	1	3	42	0	2	0
Breakfast	14	0	1	0	0	50	3	3
Dinner	9	4	3	0	0	14	9	1
Drink	6	0	0	0	0	12	8	16

Table 5. Confusion matrix of FIHMM for change data set

	Idle	Leaving	Toileting	Showering	Sleeping	Breakfast	Dinner	Drink
Idle	57	3	4	3	10	4	6	3
Leaving	16	84	0	0	0	0	0	0
Toileting	3	2	73	1	1	5	0	1
Showering	10	0	3	66	0	0	0	0
Sleeping	1	9	1	0	71	0	0	0
Breakfast	13	0	1	0	1	43	6	7
Dinner	13	1	1	1	0	14	8	2
Drink	4	1	5	0	0	6	2	24

Table 6. Confusion matrix of FIHMM for last value data set

	Idle	Leaving	Toileting	Showering	Sleeping	Breakfast	Dinner	Drink
Idle	26	7	7	11	7	8	8	16
Leaving	1	97	1	1	0	0	0	0
Toileting	11	2	64	3	3	2	0	1
Showering	5	0	7	67	0	0	0	0
Sleeping	0	0	1	0	81	0	0	0
Breakfast	15	0	0	4	0	49	1	3
Dinner	2	1	0	0	1	10	21	4
Drink	3	0	5	0	0	8	4	24

Table 7. Confusion matrix of FIHMM for change + last data set

	Idle	Leaving	Toileting	Showering	Sleeping	Breakfast	Dinner	Drink
Idle	47	5	4	10	5	3	8	7
Leaving	2	96	1	0	0	0	0	0
Toileting	4	1	74	2	1	3	1	1
Showering	1	1	5	72	0	0	0	0
Sleeping	1	0	1	0	80	0	0	0
Breakfast	14	0	1	0	0	40	11	6
Dinner	9	0	1	2	0	8	19	1
Drink	8	1	0	0	0	5	7	22

hypothesis: 1. The prediction accuracies of the B&BHMM and the traditional HMM are not different. 2. The prediction accuracy of FIHMM is not different from traditional HMM and the B&BHMM. The probability values returned, in our experiments, by Wilcoxon test for time slice accuracy comparison and class accuracy comparison are shown in Table 8 and 9 respectively. From the tables, there is little evidence in favour of the null hypothesis for overall results (Tables 8(a),9(a)). From other experiments with different data representations (Tables 8(b−e),9(b−e)), there is enough evidence to reject the null hypothesis. There are two exceptions. First is the probability for the B&BHMM against the traditional HMM in the case of time slice accuracy for raw data and the second is the probability of the proposed approach against the traditional HMM in the case of class accuracy for change+last data. These values are slightly higher than the standard significance levels. The second is because of the inherent sparsity in the data representation. Since the two null hypothesis are rejected, the alternate

Table 8. Probabilities of observing the differences in time slice accuracies for different data sets under the null hypothesis that median accuracies of the pair of approaches being compared are equal. Two tailed probability estimates of the null hypothesis being true are shown.

	Traditional HMM	B&B HMM
a) *Overall*		
B&BHMM	< 0.00006	-
FIHMM	< 0.00006	< 0.00006
b) *Raw Data*		
B&BHMM	0.4849	-
FIHMM	0.00012	0.00156
c) *Change Point*		
B&BHMM	0.0009	-
FIHMM	0.00012	< 0.00006
d) *Last Value*		
B&BHMM	< 0.00006	-
FIHMM	< 0.00006	< 0.00006
e) *Change+Last*		
B&BHMM	< 0.00006	-
FIHMM	< 0.00006	< 0.00006

Table 9. Probabilities of observing the differences in average class accuracies for different data sets under the null hypothesis that median accuracies of the pair of approaches being compared are equal. Two tailed probability estimates of the null hypothesis being true are shown.

	Traditional HMM	B&B HMM
a) *Overall*		
B&BHMM	< 0.00006	-
FIHMM	< 0.00006	< 0.00006
b) *Raw Data*		
B&BHMM	0.0003	-
FIHMM	0.00062	0.00022
c) *Change Point*		
B&BHMM	< 0.00006	-
FIHMM	0.00018	< 0.00006
d) *Last Value*		
B&BHMM	< 0.00006	-
FIHMM	0.0345	< 0.00006
e) *Change+Last*		
B&BHMM	< 0.00006	-
FIHMM	0.09186	< 0.00006

hypothesis are considered proved. There fore, we conclude that the efficiency of the proposed approach of feature induction assisted HMM model construction is statistically significant.

The learning part takes an average of three hours. The inference is faster and converges in fraction of a second. Since the learning is done once and inference being done more often, this relatively long learning time is not considered to be affecting the system performance. Moreover, The relatively longer training time can be justified by the significant accuracy gain and fast inference.

5 Conclusion and Future Work

HMM for activity recognition that has exponential observation space is not feasible in operational settings. Assuming independence among sensors given activity simplifies computation, but at the cost of accuracy. We have proposed to use learning methods to find the mappings between activities and relevant subsets of sensors. We reported the results of using the B&B structure learning system to assist HMM learning and inference and discussed the limitations. As a solution, we have proposed and implemented a feature induction assisted HMM model construction system that maximizes the accuracy of HMM inference on training data. Our experiments show good improvement over traditional HMM and B&B learning assisted HMM.

Applying the approach in other models such as CRF, analysis of actual performance benefit in terms of time and energy savings, parallelizing the learning process to run in multiple cores for speed up are some of the future works.

References

1. Wilson, D.H.: Assistive Intelligent Environments for Automatic Health Monitoring. PhD Thesis, Carnegie Mellon University (2005)
2. Gibson, C.H.S., van Kasteren, T.L.M., Krose, B.J.A.: Monitoring Homes with Wireless Sensor Networks. In: Proceedings of the International Med-e-Tel Conference (2008)
3. Wang, S., Pentney, W., Popescu, A.-M., Choudhury, T., Philipose, M.: Common sense based joint training of human activity recognizers. In: 20th International Joint Conference on Artifical Intelligence (2007)
4. van Kasteren, T., Noulas, A., Englebienne, G., Krose, B.: Accurate activity recognition in a home setting. In: 10th International Conference on Ubiquitous Computing (2008)
5. Rabiner, R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)
6. Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: International Conference on Machine Learning (2001)
7. Landwehr, N., Passerini, A., De Raedt, L., Frasconi, P.: KFOIL: Learning Simple Relational Kernels. In: 21st National Conference on Artificial Intelligence (2006)
8. Gutmann, B., Kersting, K.: TildeCRF: Conditional Random Fields for Logical Sequences. In: 15th European Conference on Machine Learning (2006)
9. Di Mauro, N., Basile, T.M.A., Ferilli, S., Esposito, F.: Feature Construction for Relational Sequence Learning. Technical Report, arXiv:1006.5188 (2010)
10. Getoor, L., Taskar, B.: Statistical Relational Learning. MIT Press, Cambridge (2006)
11. Forney, G.D.: The viterbi algorithm. Proceedings of IEEE 61(3), 268–278 (1973)
12. Srinivasan, A.: The Aleph Manual. Technical Report, University of Oxford (2007)
13. Landwehr, N., Gutmann, B., Thon, I., De Raedt, L., Philipose, M.: Relational Transformation-based Tagging for Activity Recognition. Progress on Multi-Relational Data Mining 89(1), 111–129 (2009)
14. Binsztok, H., Artieres, T., Gallinari, P.: A model-based approach to sequence clustering. In: European Conference on Artificial Intelligence (2004)
15. McCallum, A.: Efficiently Inducing Features of Conditional Random Fields. In: Nineteenth Conference on Uncertainty in Artificial Intelligence (2003)
16. Siegel, S.: Nonparametric statistics for the behavioural sciences. McGraw-Hill, New York (1956)