A probabilistic model with parsinomious representation for sensor fusion in recognizing activity in pervasive environment

Dung T. Tran† Dinh Q. Phung† Hung H. Bui‡ Sevetha Venkatesh†
School of Computing, Curtin University of Technology
GPO Box U1987 Perth, 6845 Western Australia
†{ trand,phungquo,svetha}@cs.curtin.edu.au, ‡bui@ai.sri.com

Abstract

To tackle the problem of increasing numbers of state transition parameters when the number of sensors increases, we present a probabilistic model together with several parsinomious representations for sensor fusion. These include context specific independence (CSI), mixtures of smaller multinomials and softmax function representations to compactly represent the state transitions of a large number of sensors. The model is evaluated on real-world data acquired through ubiquitous sensors in recognizing daily morning activities. The results show that the combination of CSI and mixtures of smaller multinomials achieves comparable performance with much fewer parameters.

1. Introduction

Activity recognition is essential in building smart home environments such as for aged care monitoring and intelligent homes [2, 7]. The field has attracted much attention from researchers with the advent of ubiquitous sensors which are simple, economical and easy to deploy. The task of recognizing activities is based on the information acquired through the sensors in the environment to infer what the occupant is doing or intends to do. Based on this, decisions can be made, e.g. raising timely alarms or calling for help.

Much recent research into activity recognition using ubiquitous sensors has used probabilistic models. The effort is focused on building models for fusing the sensor data. In [8], Tapia *et al.* introduce a simple model for recognizing daily activities in the home setting using state-change sensors which is built on a naive Bayesian classifier. Wilson *et al.* [10] develop a model for simultaneous tracking and activity recognition using dynamic Bayesian networks to model the sensors and employing particle filters to estimate the belief state. Other probabilistic models [4, 3, 5] restrict the input to trajectory data for activity recognition.

In our previous work [9], we proposed a probabilistic model, termed the Factored State Abstract Hidden Markov Model (FS-AHMM), based on the Abstract Hidden Markov Model (AHMM) [3] which uses a factored representation of state space to reduce the number of model parameters. The AHMM captures the hierarchical structure in policy execution, in which upper levels represent activities at higher levels of abstraction and the lowest level represents the state. An important issue is that when the number of sensors grows, the state transitions at the lowest level grow rapidly as well. The complexity is $O(K^N)$ where K is the size of the state space of a sensor and N is the number of sensors. Thus when N is large, the state transition parameter in the model, normally represented by conditional probability tables (CPT), requires large amounts of training data for estimation. In FS-AHMM [9], we represented state transitions by factorizing the state into sub-states and removing redundant links in the sub-state transitions, considerably decreasing the number of parameters. Although the CPTs are reduced in size, this model still has not solved the fundamental problem when the number of sensors grows.

In this paper, we look at different efficient representations to further reduce the number of model parameters when large numbers of sensors are used. We aim to examine several representations to parsinomiously represent the state transitions, focusing on scalability with a large number of sensors. These include context specific independence (CSI), mixtures of small multinomials and the softmax function which reduces the number of parameters to $O(K^l), O(NK^2), O(KN)$ $(l < N)$ respectively . Next we introduce and apply the model for sensor fusion in activity recognition in smart home environments and finally evaluate the performance of these representations. Our results show that the combination of CSI and mixtures of smaller multinomials is advantageous and achieves a comparable performance whilst achieving a reduction in the number of parameters.

The main contribution of the paper lies in: (1) combining

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

the representation of CSI with mixtures of smaller multinomials for representing large state transitions and thus reducing the number of parameters; (2) providing a probabilistic framework with the above representation for fusing sensor data to recognize activities in pervasive environments.

The layout of the remainder of the paper is as follows. Section 2 introduces the common dynamic Bayesian network (DBN) model for sensor fusion and several parsinomious representations. Section 3 describes the experiments and their results and finally, section 4 summarizes our work.

2. Model, Representation and Parameterization

2.1 DBN representation

2.1.1 Representation

We present a common DBN model for sensor fusion and then examine several different ways to parsinomiously represent it. The DBN representation of the common model with a flat structure is illustrated in Figure 1. The amalgamated state s_t of the model is factorized into N state variables, $s_t = (s_t^1, \ldots, s_t^N)$, where each element represents a sensor. $(o_t^1, \ldots, o_t^N)$ are the observed variables of the sensors and π represents the activity to be modeled. The DBN represents the probabilistic relationship among sensor and the activity nodes. The key issue in the model is to parameterize the state transitions $P(s_t|s_{t-1}, \pi)$. In normal representations using conditional probability tables (CPT), the number of parameters is $O(K^{2N})$, assuming that the domain value of each state variable is K. Thus, the number of parameters is very large especially when K and N are large, and thus parameter learning requires large amounts of training data. We will therefore investigate different ways of compactly representing this probability.

2.1.2 Parameterization

The probabilities to be parameterized in the model include: $P(s_t|s_{t-1}, \pi)$ and $P(o_t|s_t)$. The latter is the observation model and can be parameterized by a simple CPT with size K^2 and thus we only focus on parameterizing the former. In complete form, we have:
$\Pr(s_t|s_{t-1}, \pi) = \Pr(s_t^1, \ldots, s_t^N | s_{t-1}^1, \ldots, s_{t-1}^N, \pi)$. A simplified assumption is made that the state variables at the same time slice t are independent on each other. Thus, the state transitions can be factorized as follows:

$$P(s_t^1, \ldots, s_t^N | s_{t-1}^1, \ldots, s_{t-1}^N, \pi) = \prod_{j=1}^{N} P(s_t^j | s_{t-1}^1, \ldots, s_{t-1}^N, \pi)$$

. Therefore, we just need to represent $P(s_t^j | s_{t-1}^1, \ldots, s_{t-1}^N, \pi)$ for each substate j. We will look at several ways to compactly represent this probability which results in a different set of parameters being associated with each type of representation.

2.1.3 Inference and Learning

The learning method used for this model is the EM algorithm. The task of learning in the DBN model is to estimate the parameters of the model θ with the set of training data $D = (D^k, k = 1 \ldots N)$, using the Expectation-Maximization (EM) algorithm. The EM objective is to find a maximum likelihood (ML) estimation $\theta^* = \underset{\theta}{\text{argmax}} Pr(D|\theta)$. The EM algorithm is an iterative procedure to estimate θ^* consisting of an E-step and a M-step, which is guaranteed to converge to a local optimal point. It first computes the expected sufficient statistics (ESS) for the parameters using the smoothing distributions and then re-estimates the parameters by normalizing the ESSs for these parameters.

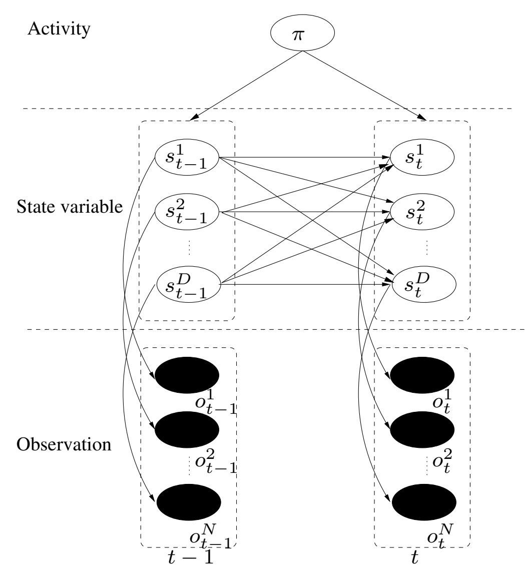

Figure 1. DBN representation of the model with two time slices: $t-1$ and t.

2.2 Compact representations

2.2.1 Context specific independence

In our previous work [9], we have applied the concept of factored state representation (CSI) [1] to compactly represent the state transitions. The basic idea is that the state transition network for each "specific" action π can

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

be simplified by removing some redundant links, instead of being fully connected, which results in reducing the number of parameters. Put simply, for each π we have $P(s_t^j|s_{t-1}^1,\ldots,s_{t-1}^N,\pi) = P(s_t^j|s_{t-1}^{i_1},\ldots,s_{t-1}^{i_l})$ with $l \leq N, 1 \leq i_1 < \ldots < i_l \leq N$.

As a result, the number of parameters in this representation are greatly reduced by a CPT with a size of $O(K^l)$. In the worst case, l is equal to N. To deal with the case of large l, we try to use other representations to parsinomiously represent $P(s_t^j|s_{t-1}^{i_1},\ldots,s_{t-1}^{i_l})$, which has a form of $P(Y|X_1,\ldots,X_l)$. This kind of conditional probabilities can be represented by a mixture of small multinomals and softmax function that will be presented in the next parts.

2.2.2 Mixture of small multinomals

In sequential data modeling, one needs to represent the probablity $\Pr(X_t|X_{t-l},\ldots,X_{t-1})$, n-order HMMs, where the number of free parameters are proportional to K^{l+1}, assuming that each variable takes on K possible values. The basic idea is to approximate by using a mixture of smaller lower-order Markov models: $\Pr(X_t|X_{t-l},\ldots,X_{t-1}) = \alpha_l(X_{t-l},\ldots,X_{t-1})f(X_t|X_{t-l},\ldots,X_{t-1}) + \ldots + \alpha_1(X_{t-l},\ldots,X_{t-1})f(X_t|X_{t-1})$, where the coefficient α may depend on the history $(X_{t-l},\ldots,X_{t-1})$ and f is a conditional probability. In [SJ99], a mixture of bi-grams is used to model this probability which is called the Mixed-memory Markov Model: $\Pr(X_t|X_{t-l},\ldots,X_{t-1}) = \sum_{i=1}^{l}\mu(i)f(X_t|X_{t-i})$, where $\mu(i)$ does not depend on the history and is treated as a latent switching parent $S_t, P(S_t = i) = \mu(i)$.

Our probability of interest, $\Pr(Y|X_1,\ldots,X_l)$, has the form somewhat like that of the above probability except that all of parents of Y are in the previous time slice t and $X_1,\ldots,X_l$ are not sequentially in time. Motivated by this, our probability can also be represented as a mixture of smaller multinomials by adapting the above representation. The variable Y now is associated with a hidden parent node B as above. We have: $\Pr(Y|X_1,\ldots,X_l) = \sum_{i=1}^{l}\Pr(B = i)\Pr(Y|X_i)$, whereas $\Pr(B = i)$, equivalent to $\mu(i)$ in Mixed-memory HMMs, is the mixing co-efficient which weights the influence of variable X_i on variable Y. Note that here we relax to the assumption that the mixing coefficient B for variable Y does not depend on the variables$X_1,\ldots,X_l$ at the previous time slice. We may consider $\Pr(Y|X_1,\ldots,X_l, B = i)$ as being simplified to $\Pr(Y|X_i)$. The parameters of the model in this representation contain only tables for $P(B = i)$ and $P(Y|X_i)$ which are proportional to $O(lK^2)$.

2.2.3 Softmax function

Another way to model our probability of interest $\Pr(Y|X_1,\ldots,X_l)$ is by using a softmax function. It can be approximated as follows:

$\Pr(Y = k|X_1,\ldots,X_l) \propto \exp(\sum_i \theta_{ki}x_i)$, for each k. This represents the influence of $(X_1,\ldots,X_l)$ on Y by a softmax function with a set of parameters $\theta = (\theta_{ki})$. Therefore, the total number of parameters using this representation is Kl. In [6], the author employs the sigmoid function, a specific case of a softmax function, to represent this kind of probability in the case where only binary sensors are used ($K = 2$). The method for parameter estimation is gradient ascent. In our case, we utilize a softmax function with $K > 2$ and the method used for estimating the parameter is Least Mean Squares.

2.2.4 Combination of CSI and other compact representations

CSI alone above proved to be efficient in modeling these kind of probabilities. As presented in 2.2, depending on a specific action π, we have: $\Pr(s_t^j|s_{t-1}^1,\ldots,s_{t-1}^N,\pi) = Pr(s_t^j|s_{t-1}^{i_1},\ldots,s_{t-1}^{i_l})$, with $l \leq N$. However, in some cases l may be still relatively large ($l \geq N/2$), requiring a huge CPT($O(K^l)$) for parameterization. To get around this, we attempt to represent this probability with a large l ($l \geq N/2$), by using a mixture of smaller multinomials or a softmax function. Note that, $Pr(s_t^j|s_{t-1}^{i_1},\ldots,s_{t-1}^{i_l})$ has the form of $\Pr(Y|X_1,\ldots,X_l)$ that can be represented by a softmax function or a mixture of mutlinomials as discussed above. As a result, we combine the CSI representation with a mixture of smaller multinomials and a softmax function to represent our probability of interest. The number of parameters used in this case is much less than that when using CSI, thus improving the CSI representation in terms of the number of parameters.

3. Experimental results

In this section, we apply the model with the above-mentioned representations in activity recognition in our smart home environment. We will then evaluate the performance of each representation.

3.1 Environment and activity description

3.1.1 Smart home Environment

In our smart home environment, multi-modal sensors are mounted on objects that the occupant is likely to interact with when performing activities. The state of the sensor is changed when a certain activity associated with it is activated. A snapshot of the environment is shown in Figure 2. We use two types of “state-change” sensors. Reed switches are installed on objects such as cupboards, the fridge, the microwave, etc and pressure mats are placed at designated locations such as dining chairs or TV chair.

Table 1 shows the sensors used in the experiments. These

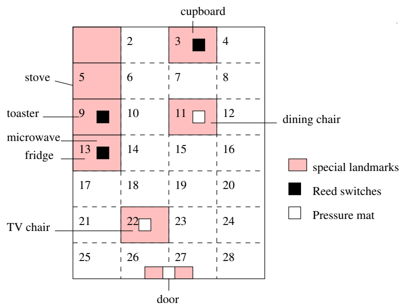

Figure 2. Environment.

sensors will provide information indicating whether the occupant is interacting with specific objects.

Due to the inherent noise of the environment and the sensors themselves, observed values are not always the same as true values. The observation model is learned seperately by comparing the observed value with its respective ground-truth.

Table 1. Sensors used in the experiments.

Reed sensors	Pressure mats
Fridge	Dining chair
Microwave	TV chair
Toaster	Door
Stove	
Cupboard	

3.1.2 Activity description and data

We use the model with the above-mentioned representations to recognize the set of complex daily activities: *having-coffe, having-snack, having-meal*. For each activity, we collected 25 labeled sequences of event data, 15 for training and 10 for testing. A typical data sequence is an array of tupples having the form $(o^1, \ldots, o^N)$ wherein $o^i = 0, 1; N = 8$. For each representation, the model for that activity is then trained with its labeled sequences using an EM algorithm.

3.2 Results and discussion

3.2.1 Evaluation

The experiments are performed with three representations: *CSI, CSI and a mixture of smaller multinomials, CSI and a softmax function* to evaluate their effectiveness. For each representation, we use three models, one each for the three activities to be trained. After training, three sets of parameters for the three models, $\theta_1, \theta_2, \theta_3$, are obtained. For each testing sequence, the likelihood of it given each model is calculated and compared to determine the winning model. It is then compared with ground truth (the label of that testing sequence) to decide whether it is correctly recognized. The recognition performance is evaluated on two criteria: *accuracy rate* and *early detection* which are defined as follows. *Accuracy rate* is the ratio of the number of testing sequences that the system recognizes correctly to the total number of testing sequences, and *early detection* is the ratio of the period of time the testing sequence is recognized correctly to the time length of that testing sequence. In the problem of activity recognition, the higher the accuracy rate and the lower the early detection, the better the recognition performance of the model. For each labeled testing sequence x, the log likelihood of that testing sequence with each set of parameters associated with each model θ_i, $Pr(x|\theta_i), i = 1 \ldots 3$, is computed and the maximum likelihood model is chosen as the winning model.

3.2.2 Results and discussion

Table 2 shows the results of the experiments with the three representations: *CSI, CSI and a mixture of smaller multinomials, and CSI and a softmax function*. As can be seen, the accuracy rate in the case of using the CSI representation only is the highest at 96.7% and the accuracy of using CSI with mixtures and softmax functions is lower at 93.3% and 90.0% respectively. The early detection is 32.13%, 31.85% and 31.88% respectively. This indicates that the softmax function is not appropriate to model the probability of interest although it requires less parameters. However, by using a mixture of smaller multinomials, it still achieves reasonable accuracy compared to the CSI only (96.7% vs 93.3%) while requiring significantly less parameters. This means that there is a trade-off between the accuracy rate and the number of required parameters.

4. Conclusion

We have presented a probabilistic model together with several parsinomious representations for sensor fusion in

Table 2. Performance of the model with different representations.

Representation	Accuracy rate	Early duration
CSI alone	96.7%	32.13%
CSI and Mixtures	93.3%	31.85%
CSI and Softmax	90.0%	31.88%

daily activity recognition in our smart home environment. These encompass CSI, a mixture of smaller multinomials and softmax function representations which compactly represent state transitions with a large number of sensors in use. CSI representation is improved, in terms of the number of parameters, by combining it with a mixture of smaller multinomials and a softmax function. These representations then are evaluated on real-world data acquired through ubiquitous sensors in recognizing daily morning activities. The results demonstrate that the combination of CSI and a mixture of smaller multinomials achieves a reasonable accuracy rate with much fewer parameters.

References

[1] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian networks. In *Uncertainty in Artificial Intelligence*, pages 115–123, 1996.

[2] C. D. Kidd. The aware home: A living laboratory for ubiquitous computing research. In *Cooperative Buildings*, pages 191–198, 1999.

[3] N. T. Nguyen, H. H. Bui, S. Venkatesh, and G. West. Recognising and monitoring high-level behaviours in complex spatial environments. 2003.

[4] N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *ICMI '02*, page 3, 2002.

[5] S. Osentoski, V. Manfredi, and S. Mahadevan. Learning hierarchical models of activity. In *International Conference on Robots and Systems*, 2004.

[6] B. Sallans. Learning factored representations for partially observable markov decision processes, 2000.

[7] V. Stanford. Using pervasive computing to deliver elder care. *IEEE Pervasive Computing*, 1(1):10–13, 2002.

[8] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home setting using simple and ubiquitous sensors.

[9] D. T. Tran, D. Q. Phung, H. H. Bui, and S. Venkatesh. Factored state-abstract hidden markov models for activity recognition using pervasive multi-modal sensors. In *ISSNIPS 2005, Melbourne*, 2005.

[10] D. H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *PERVASIVE 2005*, pages 62–79, 2005.