Human Behavior Recognition with Generic Exponential Family Duration Modeling in the Hidden Semi-Markov Model

Thi V. Duong [1], Dinh Q. Phung [1], Hung H. Bui [2], Svetha Venkatesh [1]
Department of Computing, Curtin University of Technology, Perth, WA 6102
[1] *{duong,phungquo,svetha}@cs.curtin.edu.au*
[2] *bui@ai.sri.com*

Abstract

The ability to learn and recognize human activities of daily living (ADLs) is important in building pervasive and smart environments. In this paper, we tackle this problem using the hidden semi-Markov model. We discuss the state-of-the-art duration modeling choices and then address a large class of exponential family distributions to model state durations. Inference and learning are efficiently addressed by providing a graphical representation for the model in terms of a dynamic Bayesian network (DBN). We investigate both discrete and continuous distributions from the exponential family (Poisson and Inverse Gaussian respectively) for the problem of learning and recognizing ADLs. A full comparison between the exponential family duration models and other existing models including the traditional multinomial and the new Coxian are also presented. Our work thus completes a thorough investigation into the aspect of duration modeling and its application to human activities recognition in a real-world smart home surveillance scenario.

1 Introduction

Activity recognition has become an increasingly important research problem in building smart environments. Our current application is the construction of a safe and smart house for the aged that facilitates automatic monitoring and support of its occupants. Essential to this end is a system that can learn and automatically build a model of the occupant's activities of daily living (ADLs) through observing what the occupant usually does during the day and use it in the future to monitor and predict an unseen activity. Most of existing work on activity recognition has focused on representing and learning sequential and temporal characteristics in activity sequences using the well-known hidden Markov model (HMM). While suitable and efficient for learning simple sequential data, the HMM is unable to address complex behaviors, which exhibit long-term temporal correlation that is difficult to deal with under the strong Markov assumption. ADLs, on the other hand, possess inherent duration knowledge (e.g., it is unrealistic to have the same durations spent at stove for cooking a hearty dinner and for making a quick snack), and thus the importance of duration modeling has been proven valuable in a number of recent research attempts [6, 2]. In these work, the duration information, which implies long-range correlation, is added to the standard HMM to result in what commonly known as the hidden semi-Markov model (HSMM). In brief, the HSMM generalizes the HMM by allowing its state duration to have any arbitrary distribution rather than geometric as in the HMM.

Central to the use of the HSMM is the choice of distribution family for the state duration. Most commonly, the multinomial is used (e.g., [6]) or recently the Coxian [2]. In [5, 7], the authors investigated the HSMM for the problem of speech recognition. The exponential family was also discussed in their work, in which the gamma was used as a representative. The work of [7] can be seen as closest to ours, but we claim the following points that are different from this work: (1) to our knowledge, none of recent papers have addressed the problem of human activity recognition using the generic exponential family duration modeling and compared them with other modeling choices, (2) we address the problem of inference and learning in the HSMM through a graphical representation for the model and thus provide intuition and enjoy available tools and techniques in graphical models; e.g., EM (Expectation-Maximization) learning in the presence of latent variables becomes a sub-learning problem in a generic Bayesian network.

With respect to ADLs domain, our experimental setting is a real-world home monitoring scenario in which we attempt to learn and recognize a set of complex behaviors using their tracked trajectories from vision modules. We

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

investigate both the discrete and continuous densities[1] in our experiment and provide theoretical comparative analysis and empirical comparison to other existing duration modeling choices.

The remainder of this paper is organized as follows. Section 2 discusses the exponential family state duration HSMM and presents its generic DBN structure. Section 3 focuses on modeling and learning of the exponential family duration parameters. Section 4 details the application in the ADLs domain. Finally, conclusions follow in section 5.

2 The hidden semi-Markov model (HSMM)

Figure 1 shows the state transition diagrams of the familiar HMM and the HSMM. While the HMM's state durations are strictly geometric (parameterized by the diagonal entries in the transition matrix), a state in the HSMM can have any arbitrary duration distribution. Formally, an HSMM is fully defined by a state space Q, an observation set V, and a parameter set $\theta = \{\pi, A, D, B\}$. We number the state sequentially and thus the state space is $Q = \{1, \dots, |Q|\}$. The initial probability π_i specifies the probability of state $i \in Q$ being activated when the HSMM is initialized. The transition probability A_{ij} is the probability of reaching state j given the previous state i has completed its activation. Self-transition is not allowed, thus, $A_{ii} = 0$, $\forall i \in Q$. For each state i, a duration distribution D^i is specified, which governs the number of times it remains unchanged. At each time point, an observation $v \in V$ is generated with probability $B_{v|i}$ where i is the current state. Finally, stochastic constraints require: $\sum_i \pi_i = 1$, $\sum_j A_{ij} = 1$, and $\sum_v B_{v|i} = 1$.

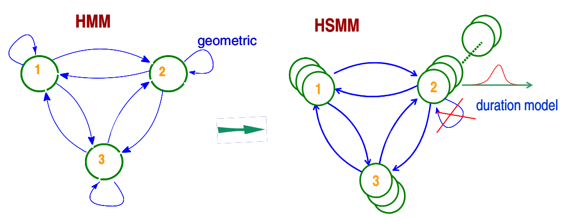


Figure 1. From HMM to HSMM.

Figure 2 shows a generic DBN representation unrolled T times for the HSMM whose state durations can be modeled by any distribution in the exponential family. At each time slice, we maintain a set of three variables $\mathcal{V}_t = \{x_t, m_t, y_t\}$ where x_t is the current state, m_t is the duration variable of the current state, and y_t is the current observation. In most of the cases (and also the most useful case) is when only y_t is observed and the rest is hidden. To keep the tradition, we use clear nodes for hidden variables and shaded nodes for observed variables.

It is noteworthy to mention that the duration node m_t is a counting-down variable, which specifies not only how long the current state will last, but also acts like the context defining how the next time slice is evolved from the current. When $m_t > 1$, the state $x_t = i$ is still in its execution and thus remains unchanged to the next time slice (i.e., $x_{t+1} = i$); at the same time the duration is counted down by 1 (i.e., $m_{t+1} = m_t - 1$). When m_t is counted down to 1, it signals the completion of the current state and therefore the next state $x_{t+1} = j$ is set with a transition probability A_{ij}. The duration variable m_{t+1} is then set to some value m with a probability of D^j_m where D^j is the duration model defined for state $x_{t+1} = j$.

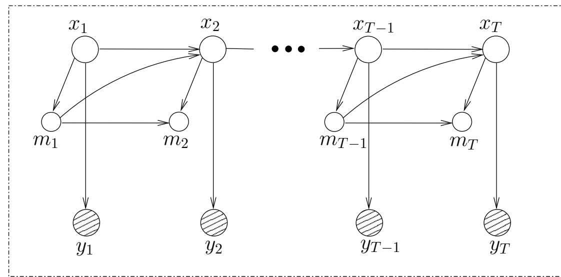


Figure 2. DBN representation of the HSMM.

3 Learning the duration parameters

We use the well-known EM algorith[1] to learn the parameters for the HSMM. EM algorithm is an iterative optimization method that returns the maximum likelihood estimation by optimizing the likelihood lower bound. It hill-climbs in the parameter space and guarantees to converge to a local maximum. EM consists of two main steps: the E-step closes the gap between the (log) likelihood and its lower bound (equivalently to expressing the complete log-likelihood and computing the expected sufficient statistics (ESS) for the parameters), followed by the M-step, which returns the estimated parameters that maximize the lower bound. We focus on learning duration parameters modeled by exponential family in this paper[2].

In the expression of the complete log-likelihood $\mathcal{L} = P(x_{1:T}, m_{1:T}, y_{1:T} \mid \theta)$ the terms associated with the duration parameters are given by:

$$\mathcal{L}_D = \log \prod_{t=1}^{T} \Pr(m_t \mid x_t, m_{t-1}) = \sum_{t=1}^{T} \sum_{m,i} \delta^{(m)}_{m_t} \delta^{(i)}_{x_t} \delta^{(1)}_{m_{t-1}} \log D^i_m \qquad (1)$$

[1] For convenience, we will refer both 'probability density' in continuous case and 'probability mass' in discrete case as 'density'.

[2] Learning the rest of parameters has been addressed in several existing papers, e.g.,[9, 6, 2].

where $\delta_a^{(b)}$ is an identity function, i.e., returns 1 if $a = b$ and $= 0$ otherwise. Given (1), it can be shown that in the expression of the expectation of $\mathcal{L}$ over $\Pr(\text{hidden} \mid \text{observed}) = \Pr(x_{1:T}, m_{1:T} \mid y_{1:T})$ in the E-step, the term associated the duration parameters are:

$$\langle \mathcal{L}_D \rangle = \sum_{m,i} \text{Ess}\left(D_m^i\right) \log D_m^i \tag{2}$$

where $\text{Ess}\left(D_m^i\right) = \sum_{t=1}^T \Pr(x_t^i, m_t^m, m_{t-1}^1 \mid y_{1:T})$ is the ESS for the parameter D_m^i, and we write $\langle \cdot \rangle$ to mean the expectation operator.

In the M-step, we need to maximize the expected log-likelihood $\langle \mathcal{L}_D \rangle$ in equation (2) with respect to D_m^i and the method depends on the choice of duration distributions. We note that the method is identical to every state $i \in Q$, and thus i is henceforth dropped out in equation (2) for brevity:

$$\langle \mathcal{L}_D \rangle = \sum_{m} \text{Ess}\left(D_m\right) \log D_m \tag{3}$$

In the following section, we will address the M-step when the duration is modeled as a multinomial, a Poisson, or an Inverse Gaussian.

3.1 Multinomial Duration

In non-parametric modeling[3], the duration is modeled as multinomial, i.e., $D \sim Mult\left(D_1, D_2, \ldots, D_M\right)$. Lagrange multiplier method is used in the M-step:

$$\langle \mathcal{L}_D \rangle = \sum_{m} \text{Ess}\left(D_m\right) \log D_m + \lambda \left(1 - \sum_{m'} D_{m'}\right) \tag{4}$$

Setting the derivative $\frac{\delta \langle \mathcal{L}_D \rangle}{\delta D_m} = 0$, and with some manipulation, we obtain the re-estimated formula:

$$\hat{D}_m = \frac{\text{Ess}\left(D_m\right)}{\sum_{m'} \text{Ess}\left(D_{m'}\right)} \tag{5}$$

3.2 Modeling state durations by the Exponential Family

The duration D is now modeled by a distribution in the exponential family, which includes a rich set of distributions such as binomial, Poisson, Gaussian, Inverse Gaussian, gamma, and many more. The probability of a state having a duration m takes the following form:

$$D_m \triangleq \Pr(m \mid \mathbf{w}) \triangleq h(m) \exp\left(\mathbf{w}^T S(m) - A(\mathbf{w})\right) \tag{6}$$

where the function $h(m)$ is not of fundamental importance as it plays no role in the M-step. Of rather more importance are the natural parameter $\mathbf{w}$, the sufficient statistics $S(m)$, and the log partition function $A(\mathbf{w})$. The log partition function is a log normalization factor: $A(\mathbf{w}) = \log \sum_m \exp\left(\mathbf{w}^T S(m)\right)$ for discrete m. From equations (3) and (6), it follows:

$$\langle \mathcal{L}_D \rangle = \sum_{m} \text{Ess}\left(D_m\right) \mathbf{w}^T S(m) - \sum_{m} \text{Ess}\left(D_m\right) A(\mathbf{w}) \tag{7}$$

In maximizing $\langle \mathcal{L}_D \rangle$, we consider two popular and useful distributions from the exponential family: the Poisson (discrete) and Inverse Gaussian (continuous).

3.2.1 The Poisson duration distribution

The Poisson duration distribution was used to model durations in speech recognition [10] and reported with good results. We thus employ it in our work for the problem of human activity recognition. Its density is given as:

$$D_m \triangleq \Pr(m \mid \lambda) = \frac{\lambda^m \exp(-\lambda)}{m!} \propto \exp\left(m \log \lambda - \lambda\right) \tag{8}$$

Thus, it belongs to exponential family with the natural parameter $w = \log(\lambda)$, the sufficient statistics $S(m) = m$, and the log partition function $A(w) = \exp\left(w\right)$. Therefore, it follows from equation (7) that:

$$\frac{\delta \langle \mathcal{L}_D \rangle}{\delta w} = \sum_{m} \text{Ess}\left(D_m\right) m - \sum_{m} \text{Ess}\left(D_m\right) \exp(w) \tag{9}$$

Setting $\frac{\delta \langle \mathcal{L}_D \rangle}{\delta w} = 0$ leads to the estimation of λ:

$$\hat{\lambda} = \exp\left(\hat{w}\right) = \frac{\sum_m \text{Ess}\left(D_m\right) m}{\sum_m \text{Ess}\left(D_m\right)} \tag{10}$$

The above expression is intuitive as the Poisson parameter λ has the meaning of the duration mean.

3.2.2 The Inverse Gaussian duration distribution

We choose the Inverse Gaussian (IG) as a representative of continuous distribution in the exponential family because it is restricted to the positive domain, and has been used to model patients' staying time in hospital [11] with successful results reported. The IG distribution can be expressed in exponential form as:

$$IG(m, w_1, w_2) = \frac{m^{-\frac{3}{2}}}{\sqrt{2\pi}} \exp\left(\frac{w_1}{m} + w_2 m - A(w_1, w_2)\right) \tag{11}$$

where the log partition function $A(w_1, w_2) = -2\sqrt{w_1 w_2} - \frac{1}{2}\log(-2w_1)$, and clearly, the sufficient statistics is $S(m) = [1/m, m]$. When applying the IG into discrete domain, we have to introduce an additional normalization term $N(w_1, w_2) = \sum_m \exp\left(\mathbf{w}^T S(m) - A(w_1, w_2)\right)$.

$$D_m \triangleq \Pr(m \mid w_1, w_2) = \frac{IG(m, w_1, w_2)}{N(w_1, w_2)} \tag{12}$$

Compared to the standard result in equation (7), we have an extra (and problematic!) term

[3]The multinomial in fact belongs to the exponential family distribution, however, in duration modeling we tend to view it as a non-parametric model.

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

$\sum_m \text{Ess}\,(D_m) \log\left(N(w_1, w_2)\right)$ to optimize, which will require an approximation solution. Theoretically we can choose any black-box optimization method (e.g., steepest descent, conjugate gradient, etc.) here. In this work, to avoid the costly computation of Jacobian or Hessian, we choose to optimize it by the Nelder-Mead method [4], which requires neither the first nor the second derivative, but only function evaluations.

As a final note, the above analyses show that the complexity of re-estimating the duration parameters in both the Poisson and Inverse Gaussian is in order $O(M)$ with M being the maximum allowed state duration length, which is the same as the multinomial case, and different from the Coxian case where the complexity depends on the Coxian size [2, 3][4].

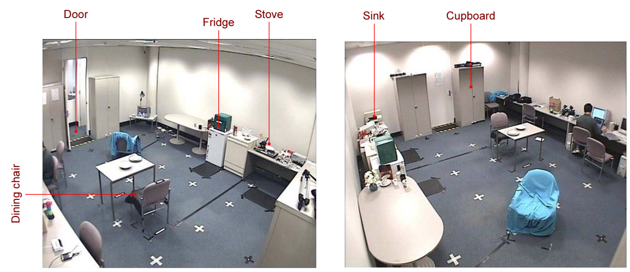
4 Experimental Results

Our interest is the construction of a smart home for the aged, that can learn and recognize occupant's activities performed daily. There are several common categories of ADLs in the house such as cooking-meal, washing-dishes, ironing-clothes, reading-newspaper, etc. Activities of the same category generally follow the same standard procedures. For example, cooking-meal would include: *taking-food-from-fridge - washing-vegies/cutting-meat - seasoning-food - cooking*. However, the sub-activities within a given category may have different durations. For example, time spent at stove for cooking-breakfast would be less than that for cooking-dinner. The problem of learning and distinguishing ADLs of the same category (such as cooking-breakfast vs. cooking-dinner) is more challenging as the discriminative power lies mainly in the differences in the duration patterns of their sub-activities.

In this paper, we experiment with the exponential family duration HSMMs including the (Poisson) Ps-HSMM and the (Inverse Gaussian) Ig-HSMM, in learning and recognizing three different routines of the meal-preparation-&-consumption category, and compare their performance with other existing duration models including the (Multinomial) Mu-HSMM, the (Coxian) Cx-HSMM (presented in [3]), and also the standard HMM.

4.1 Environment and Data Descriptions

The smart environment used in this experiment is a laboratory kitchen set up as shown in figure 3. The scene is captured by two cameras mounted at two opposite ceiling corners, and a multiple-camera tracking module is used to detect movements, returning the list of positions in x-y coordinates visited by the occupant. For modeling convenience, the kitchen is quantized into 28 square cells of $1m^2$, and the returned x-y readings are then converted into cell numbers. The low-level vision tracking module employed in this work is the same as that of [8]. This tracking module, however, sometimes returns a neighboring position instead of the actual position occupied by the person, so an observation model is estimated offline with manually labeled ground truth [8]. This corresponds to estimating the observation model B separately.



(a) camera 1 (b) camera 2

Figure 3. The environment viewed from two cameras.

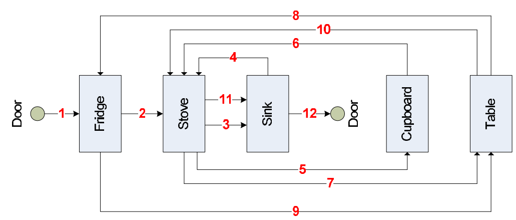


Figure 4. Sequential flow chart for landmarks being visited in three experimenting activities.

We collect a total of 48 sequences for the three following activities in the meal preparation and consumption category: (a.1) a-tea-cake-newspaper-breakfast, (a.2) a-scrambled-egg-on-toast-lunch, and (a.3) a-lasagna-salad-lunch. We consider the extreme case, in which the three activities have exactly the same sequential order of sub-activities, but differ in the durations of these tasks. This is also the hardest scenario since the differences in duration patterns, and not in trajectories makes our task of activity classification more challenging. Figure 4 shows the twelve fixed sequential steps: *1. take-food-from-fridge → 2. bring-food-to-stove → 3. wash-vegetable/fill-water-at-sink → 4. come-back-to-stove-for-cooking → 5. take-plates/cup-from-cupboard → 6. return-to-stove-for-food → 7. bring-food-to-table → 8. take-drink-from-fridge → 9. have-meal-at-table → 10. clean-stove → 11. wash-dishes-at-sink → 12. leave-the-kitchen*. Table 1 shows the ground truth of typical durations spent at special landmarks including fridge, stove, sink, cupboard, and table

[4]For full description of the Coxian duration model, please refer to [3].

of the three activities. For example, duration spent at stove for cooking scrambled eggs on toast, $15-17$(s), is generally longer than for reheating the lasagna, $8-10$(s), or making a cup of tea, $7-9$(s); or having breakfast while reading the morning newspaper, $28-32$(s), usually requires more time at the table than simply having lunch alone, $14-16$(s) or $19-21$(s). In addition, table 1 shows that each landmark may have multiple durations (the first column shows the duration of the first visit, the second column is the duration of the second visit, etc.[5]). In this experiment, we have covered the possibility that an occupant may visit some landmarks several times within an activity, and different activities may sometimes share the same durations at the same places.

Table 1. Typical durations spent (seconds) at landmarks.

	FRIDGE		STOVE				SINK		CUPB	TABLE	
(a.1)	1–2	4–6	1–2	1–2	7–9	1–2	2–4	8–10	8–10	1–2	28–32
(a.2)	6–8	1–2	8–10	15–17	4–6	8–10	6–8	18–20	1–2	3–4	14–16
(a.3)	10–12	1–2	4–6	8–10	2–4	3–5	12–14	12–14	1–2	3–4	19–21

4.2 Training

To ensure an objective result, we employ a *leave-one-out* strategy for training and testing. We sequentially pick out one sequence Y from the data set D for testing, and use the remainder $\{D \setminus Y\}$ for training. In the trained models (Ps-HSMM, Ig-HSMM, Mu-HSMM, 5-phase Cx-HSMM and HMM), we let the number of states $|Q| = 28$, equal to the number of square cells quantized in the kitchen environment (figure 3), and the observation model B is obtained offline [8]. In the Ps-HSMM, Ig-HSMM, and Mu-HSMM, we equate the maximum duration M to the maximum activity length, approximately in the range $[100, 120]$. Except for the constraints mentioned here, all other parameters are randomly initialized. Figure 5 shows the log likelihood learned from the full data set of activities (a.3) by the Ps-HSMM and the Ig-HSMM. The Ps-HSMM quickly converges after the first three iterations since its duration model has only one parameter, and is re-estimated by a closed-form solution. On the contrary, the Ig-HSMM, whose duration models are parameterized by two parameters and learned by an approximation method, needs about twenty iterations to stabilize. However, the learning time for the Ig-HSMM is well paid off as its log likelihood converges to a much higher value than that of the Ps-HSMM.[6]

[5]For example, for activity (a.1), the occupant first stops at the fridge for 1-2(s) to check out milk and cake, then later returns to the fridge for 4-6(s) (after steeping tea) to take out milk and cake; whereas in activity (a.2), the occupant stops at the fridge the first time for 6-8(s) to take out food and then re-visits the fridge afterwards for 1-2(s) to get a drink.

[6]As it is well known, a more complex model (i.e. having more parameters) will result in a higher log likelihood. However, the significant discrepancy between the two log likelihood curves suggests that the substantial increase in the Ig-HSMM's log likelihood is mainly due to the model's better capability in learning the data.

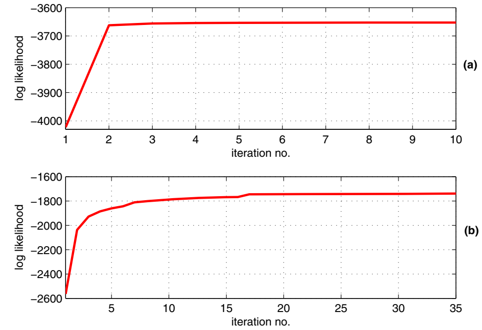


Figure 5. The log likelihood learned from activity (a.3) by: **(a)** the Ps-HSMM, and **(b)** the Ig-HSMM.

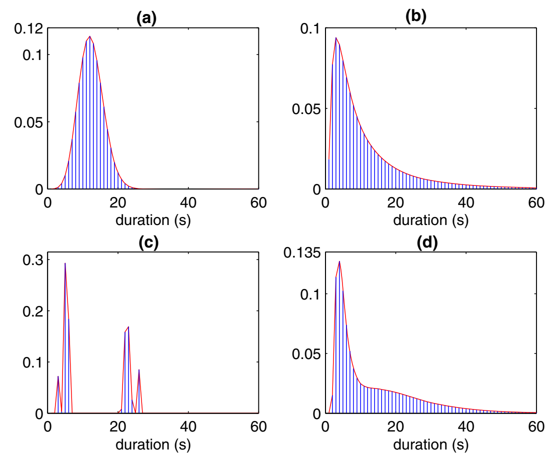


Figure 6. The pmf of the durations spent at table in activity (a.3) learned by: **(a)** the Ps-HSMM, **(b)** the Ig-HSMM, **(c)** the Mu-HSMM, and **(d)** the Cx-HSMM.

4.3 Activity Recognition Results

First, we look at how the different models have learned the state durations. Figure 6 shows the duration spent at table in activity (a.3) learned by the Ps-HSMM, the Ig-HSMM, the Mu-HSMM, and the Cx-HSMM. Except the Ps-HSMM, the rest has relatively captured well the mixture of two typical durations: $3-4$(s), and $19-21$(s). Like the Cx-HSMM, the Ig-HSMM has not fully separated the two peaks, but successfully smoothed out the spikes in the durations in comparison with the Mu-HSMM.

Next, the exponential family duration HSMMs are compared against other models on their performance at activity classification accuracy and early detection rate. For each sequence $y_{1:T}$ left out in the *leave-one-out* training selection, the likelihood $\Pr(y_{1:t} \mid \theta_i)$, for $i = 1, 2, 3$, where θ_i is the

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

model trained with the set of activity (a.i), is computed at each time t and used to label the most likely activity. Classification accuracy is the ratio of activities correctly labeled at time $t = T$ to the total activities tested, while early detection rate is the ratio t_0/T where t_0 is the earliest time from which the activity label remains accurate.

(a)	HMM (avg.68.02%)			Cx-HSMM (avg.91.39%)		
	(a.1)	(a.2)	(a.3)	(a.1)	(a.2)	(a.3)
(a.1)	88.24	0	11.76	100	0	0
(a.2)	0	62.50	37.50	0	87.50	12.50
(a.3)	13.33	33.33	53.33	0	13.33	86.67

(b)	Ps-HSMM (avg.69.05%)			Ig-HSMM (avg.76.53%)			Mu-HSMM (avg.95.56%)		
	(a.1)	(a.2)	(a.3)	(a.1)	(a.2)	(a.3)	(a.1)	(a.2)	(a.3)
(a.1)	58.82	17.65	23.53	100	0	0	100	0	0
(a.2)	0	75.00	25.00	0	56.25	43.75	0	100	0
(a.3)	0	26.67	73.33	0	26.67	73.33	0	13.33	86.67

Table 2. Classification Accuracy (%).

	HMM	Cx-HSMM	Ps-HSMM	Ig-HSMM	Mu-HSMM
(a.1)	9.12	7.26	31.54	7.99	8.97
(a.2)	37.28	20.31	13.89	47.72	11.77
(a.3)	42.57	27.56	43.96	31.96	26.03
Avg.	29.66	18.38	29.80	29.22	15.59

Table 3. Early Detection Rates (%).

Tables 2(a-b) demonstrate that the HSMM with Poisson or Inverse Gaussian duration model has outperformed the standard HMM. The improvement, however, is slightly insignificant for the Poisson case (i.e. increase only 1% from 68.02% to 69.05%). This is possibly because the Poisson is not flexible enough as its shape is controlled by only one free parameter λ. With respect to the Inverse Gaussian, although it has significantly increased the accuracy level to 76.53%, its performance is still worse than the multinomial and the Coxian. This possibly dues to its approximation step to discrete domain and the approximate optimization step during EM learning.

5 Conclusion

This paper has presented an investigation into the use of exponential family distribution in modeling state occupancy in the HSMM and its application to human behavior recognition. A generic DBN structure for the HSMM is presented which consequentially allows us to efficiently address the problem of inference and parameter learning. We then apply the HSMM for the problem of learning and recognizing a set of complex behaviors in a smart-home environment. Different member of exponential family including both discrete and continuous densities are used in our experiment and compared to other duration modeling choices and the standard HMM. The results have shown that the introduction of exponential family duration in the HSMM results in significant recognition improvements as compared to the HMM. They have also indicated that for our ADLs setting, among experimenting members of the exponential family, the multinomial performs best. The Ps-HSMM has been shown to be too simplistic while Ig-HSMM requires a number of approximation steps when applied to discrete domain, and thus despite its significant improvement over the HMM, it is still worse than the multinomial. Finally, this paper has completed a full investigation into a rich set of duration modeling methods for the HSMM in the ADLs domain.

References

[1] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[2] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the Switching Hidden Semi-Markov Model. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 838–845, San Diego, 20-26 June 2005. IEEE Computer Society.

[3] T. V. Duong, D. Q. Phung, H. H. Bui, and S. Venkatesh. Efficient coxian duration modelling for activity recognition in smart environment with the hidden semi-markov model. In *Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, Australia, 5-6 December 2005.

[4] K. D. F. John H. Mathews. *Numerical Methods Using Matlab*. Prentice Hall, Upper Saddle River, NJ 07458, 1999.

[5] S. E. Levinson. Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech and Language*, 1(1):2945, March 1986.

[6] S. Luhr, S. Venkatesh, G. West, and H. H. Bui. Duration abnormality detection in sequences of human activity. Technical report, Department of Computing, Curtin University of Technology, May 2004.

[7] C. D. Mitchell and L. H. Jamieson. Modeling duration in a hidden markov model with the exponential family. In *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages II.331–II.334, Minneapolis, Minnesota, April 1993.

[8] N. T. Nguyen, S. Venkatesh, G. West, and H. H. Bui. Learning people movement model from multiple cameras for behaviour recognition. In *Joint IAPR International Workshops on Structural and Syntactical Pattern Recognition and Statistical Techniques in Pattern Recognition*, pages 315–324, Lisbon, Portugal, August 2004.

[9] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Procs. IEEE*, volume 77, pages 257–286, February 1989.

[10] M. J. Russell and R. K. Moore. Explicit modelling of state occupancy in hidden markov models for automatic speech recognition. In *Proceedings of IEEE Conference on Acoustics Speech and Signal Processing,*, pages 5–8, March 1985.

[11] V. Seshadri. *The Inverse Gaussian Distribution: A Case Study in Exponential Family*. Oxford Science Publications, 1993.

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