Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model

Thi V. Duong†, Hung H. Bui‡, Dinh Q. Phung†, Svetha Venkatesh†
† Department of Computing, Curtin University of Technology
GPO Box U1987, Perth, Western Australia 6845
Email: {duong, phungquo, svetha} @cs.curtin.edu.au
‡ Artificial Intelligence Center, SRI International
333 Ravenswood Ave, Menlo Park, CA 94025, USA
Email: bui@ai.sri.com

Abstract

This paper addresses the problem of learning and recognizing human activities of daily living (ADL), which is an important research issue in building a pervasive and smart environment. In dealing with ADL, we argue that it is beneficial to exploit both the inherent hierarchical organization of the activities and their typical duration. To this end, we introduce the Switching Hidden Semi-Markov Model (S-HSMM), a two-layered extension of the hidden semi-Markov model (HSMM) for the modeling task. Activities are modeled in the S-HSMM in two ways: the bottom layer represents atomic activities and their duration using HSMMs; the top layer represents a sequence of high-level activities where each high-level activity is made of a sequence of atomic activities. We consider two methods for modeling duration: the classic explicit duration model using multinomial distribution, and the novel use of the discrete Coxian distribution. In addition, we propose an effective scheme to detect abnormality without the need for training on abnormal data. Experimental results show that the S-HSMM performs better than existing models including the flat HSMM and the hierarchical hidden Markov model in both classification and abnormality detection tasks, alleviating the need for presegmented training data. Furthermore, our discrete Coxian duration model yields better computation time and generalization error than the classic explicit duration model.

1 Introduction

Activity recognition is an important aspect in building pervasive environments. Our motivating application is the construction of a safe and smart house for the aged that facilitates automatic monitoring and supporting its occupants. There are two main problems in building such a system. First, the system needs to learn, understand, and automatically build a model of the occupant's *activities of daily living* (ADL) through observing what the occupant usually does during the day. Second, the system needs to be able to use its learned knowledge to monitor the person's current activity, and to detect if there is any deviation from the normal activity patterns and alert the caregiver if necessary.

Traditionally, most activity recognition work has focused on representing and learning the sequential and temporal characteristics in activity sequences. This has led to the widespread use of dynamic models such as the hidden Markov model (HMM) [23, 22]. While the HMM is a simple and efficient model for learning sequential data, its performance tends to degrade when the range of activities becomes more complex, or the activities exhibit long-term temporal dependency that is difficult to deal with under the Markov assumption.

To get around these limitations, two classes of extension to the HMM have been proposed. The first extension introduces models that supplement the basic HMM with a hierarchical structure, aiming to exploit the natural hierarchical organization of human behaviors. Examples of these models include the Abstract HMM [2], the Hierarchical HMM [3, 8, 1], and the Layered HMM [18]. Long-term dependency is captured in these models via the additional layers designed to model higher-level activities evolving at slower timescales.

The second extension adopts the semi-Markov model and introduces its hidden variants [15], including explicit duration HMMs [20, 21] and segmental HMMs [4]. In these models, a state is assumed to remain unchanged for some random duration of time before its transit to a new state. For each state a duration distribution is given to char-

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

acterize the length of its duration. The hidden semi-Markov model (HSMM) has been an active research topic since the late 1980s, driven by applications in the field of speech processing and recognition. It addresses the violation of the Markov assumption arising from having states whose duration distributions are nonexponential (or nongeometric if time is discrete).

A classic approach is to model the duration explicitly via the multinomial distribution [20, 21, 4, 10]; however, its drawback is in the large number of free parameters needed, which requires more training data and incurs extra computation cost in both training and classification. Existing approaches to overcome this problem typically use a more compact parametric duration model such as the continuous *gamma* distribution [7], or an integer-valued distribution in the exponential family [11]. All these methods require additional approximation when applied to a discrete-time domain resulting in longer learning/classification time, and an approximate numerical method in the M-step of the Expectation Maximization (EM) procedure for parameter reestimation, with complexity depending on the maximum possible duration length. Using a Coxian distribution for duration modeling, as proposed in this work, yields an elegant solution to these problems: when applied to the discrete domain, its exponential phase components, as shown later, are simply replaced by the geometric distributions; furthermore, it is flexible enough, yet remains computationally efficient and avoids the problem of having to determine the maximum possible duration length in advance.

We argue here that in the domain of modeling ADL, it is beneficial to exploit both the inherent hierarchical organization of activities and their typical duration. Despite the fact that the duration of any activity is unlikely to have an exponential or geometric distribution (hence requiring a semi-Markov model), there have been only a few recent attempts at duration modeling in activity recognition. A Gaussian duration distribution is used in [19] but parameter learning is not supported, while an explicit multinomial duration HMM is used in [9]. Previous work [6] has also recognized the need for combining both the hierarchical and semi-Markov extensions to form a hierarchical hidden semi-Markov model. However, there has been no attempt at formalizing such a model or demonstrating its usefulness empirically over other existing models.

In this paper, we introduce the Switching Hidden Semi-Markov Model (S-HSMM), a special case of the hierarchical model with only two layers. The top layer is a Markov sequence of *switching* variables, while the bottom layer is a sequence of concatenated HSMMs whose parameters are determined by the switching variable at the top. Thus, the dynamics and duration parameters of the HSMM at the bottom layer are not time invariant, but are "switched" from time to time, similar to the way linear Gaussian dynamics are "switched" in a switching Kalman filter [12].

Mapping to the ADL modeling problem, our intention is to use the bottom layer in our model to capture atomic activities such as spending time at the cupboard, stove, fridge, or moving between these designated places. Several of these atomic activities then form high-level activities in the house such as making breakfast, eating breakfast, making coffee, or washing dishes, each represented by a state at the top layer in our model. Transition from one top-level state to another represents sequences of high-level activities that are typical in a human daily routine. We note that only the duration of the atomic activity is represented in our framework using the semi-Markov model at the bottom layer. Since a high-level activity is made of a sequence of atomic activities, its total duration is implicit from the duration of the atomic activities.

We provide a formal definition for the S-HSMM including two different probabilistic models for the duration using the multinomial and the discrete Coxian parameterization. We then develop formalisms for inference and maximum-likelihood parameter estimation based on the conversion of this model into an equivalent dynamic Bayesian network (DBN) [13]. We apply the S-HSMM to the problem of recognizing high-level activities and detecting abnormalities in durations of low-level activities in a typical routine sequence and compare its performance with two existing approaches: a flat HSMM without information about activity hierarchy, and a two-layer hierarchical HMM without duration modeling. Our experimental results demonstrate that the S-HSMM outperforms these existing models and confirm our belief that both hierarchy and duration information are needed to build accurate activity models in the home. We also demonstrate that using the discrete Coxian distribution for duration modeling improves both the computational time and the generalization error. Furthermore, we derive an effective method for detecting abnormality in activity duration based on inverting the learned duration model.

The layout of this paper as follows. Sec.2 develops the S-HSMM framework including definition, inference and learning. Sec.3 presents the experimental results for activity and duration abnormality detection. Finally, our conclusions follow in Sec.4.

2 The Switching Hidden Semi-Markov Model

In this section, we provide a formal definition for the S-HSMM, together with methods for inference and parameter estimation. We start with a 2-layer hierarchical HMM and then describe how a semi-Markov extension can be added to this model. Methods for inference and learning are then derived by viewing the model as a dynamic Bayesian network.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

2.1 Model definitions and parameters

Let us consider a 2-layer hierarchical HMM [3, 1] defined as follows. The state space is divided into the set of states at the top level $Q^* = \{1, \ldots, |Q^*|\}$ and states at the bottom level $Q = \{1, \ldots, |Q|\}$. Our convention is to use the letters p, q to refer to elements of Q^* and i, j to refer to elements of Q. The parameters π^*_p and A^*_{pq} are the initial probability and the transition matrix of a Markov chain defined over the states in Q^*. For each top-level state p, $\mathrm{ch}(p) \subset Q$ is the set of children of p (it is possible that two different parent states might share some common children). A transition to p in the top-level Markov chain will initiate a Markov chain at the bottom level over the states in $\mathrm{ch}(p)$. The parameters of this p-initiated chain are given by π^p_i,A^p_{ij},$A^p_{i,end}$ where π^p_i,A^p_{ij} are the initial and transition probabilities as usual, and $A^p_{i,end}$ is the probability that this chain will terminate after a transition to i. At each time point, an observation y in the alphabet Y is generated with probability B_{iy}, where i is the current state at the bottom level.

In this 2-layer HHMM, the duration d for which a bottom state i remains the same has a geometric distribution: $d \sim \mathit{Geom}(1 - A^p_{ii})$. In activity modeling, geometric distributions are often too restricted to model realistic durations of activities. Thus, we would like d to follow a more general discrete distribution $d \sim D^{p,i}(d)$. To state this in a more precise way, the p-initiated chain at the bottom level is now a semi-Markov sequence with π^p_i,A^p_{ij}, $D^{p,i}(d)$ being the initial, transition, and duration probabilities, respectively (A^p_{ii} must be zero). The termination and observation probabilities remain the same as in the 2-layer HHMM. We term this the Switching Hidden Semi-Markov Model (S-HSMM) since it can be viewed as the concatenation of many HSMMs, each initiated by a different "switching" state p.

2.2 Duration model

The usual choice for D is the multinomial distribution $\mathit{Mult}(\mu_1, \ldots, \mu_M)$, $\mu_i \geq 0, \sum_i \mu_i = 1$ [20, 10]. However, modeling duration in this way becomes very inefficient when M, the maximum duration length, is large. Such a situation is often encountered in activity modeling, especially when some types of activities are considerably longer than others. We thus propose the use of the *discrete Coxian* distribution [16] as follows.

A discrete Coxian distribution[1] with parameter $\mu = \mu_1, \ldots, \mu_M$ and $\lambda = \lambda_1, \ldots, \lambda_M$, denoted by $\mathit{DCox}(\mu, \lambda)$ where $0 \leq \mu_i \leq 1$, $\sum \mu_i = 1$, $0 < \lambda_i \leq 1$, is defined as the mixture $\mathit{Mix}(\mu_1, S_1; \ldots; \mu_M, S_M)$ where $S_i = X_i + \ldots + X_M$; X_i are independent and have geometric distributions $X_i \sim \mathit{Geom}(\lambda_i)$. Note that the Coxian distribution, as shown later in section 3, will typically has a much smaller M than the multinomial one. The discrete Coxian distribution is a member of the phase-typed distribution family and has the following very appealing interpretation. Imagine a left-to-right Markov chain with $M + 1$ states numbered from 1 to $M + 1$, with the self transition parameter $A_{ii} = 1 - \lambda_i$. The first M states represent the M *phases*, while the last state is absorbing and acts like an end state. The duration of the state (phase) i is $X_i \sim \mathit{Geom}(\lambda_i)$. If we start from state i, $S_i = X_i + \ldots + X_M$ is the duration of the Markov chain before the end state is reached. Thus, $\mathit{DCox}(\mu, \lambda)$ is in fact the distribution of the duration of this constructed Markov chain when μ is the initial state distribution. The discrete Coxian is much more flexible than the geometric distribution: its probability mass function is no longer monotonically decreasing and it can have more than one mode. As a special case, we note that if for all i, $\lambda_i = 1$ thus $X_i \equiv 1$, we recover the multinomial distribution: $\mathit{DCox}(\mu, 1) = \mathit{Mult}(\mu_M, \mu_{M-1}, \ldots, \mu_1)$. Further, note that in the Coxian distribution the number of phases M, as shown later in Sec.3, is typically much smaller than the maximum length M in the multinomial case.

Using the discrete Coxian distribution, we model the duration distribution for states at the bottom level in the S-HSMM as follows. For each p-initiated semi-Markov sequence, the duration distribution of a child state i is $D^{p,i}(d) = \mathit{DCox}(d; \mu^{p,i}, \lambda^{p,i})$. The parameters $\mu^{p,i}$ and $\lambda^{p,i}$ are M-dimensional vectors where M is a fixed constant representing the number of phases of the discrete Coxian. When $M = 1$, the model becomes identical to a 2-layer HHMM.

2.3 Dynamic Bayesian Network representation

Following the idea of representing the HHMM as a DBN [14], Fig. 1 shows a DBN representation of the S-HSMM for two time slices. At each time slice t, a set of variables $\mathcal{V}_t = \{z_t, \epsilon_t, x_t, e_t, m_t, y_t\}$ is maintained. At the top level, z_t is the current top-level state acting as a switching variable; ϵ_t is a boolean-valued variable set to 1 when the z_t-initiated semi-Markov sequence ends at the current time slice. At the bottom level, x_t is the current child state in the z_t-initiated semi-Markov sequence; e_t is a boolean-valued variable set to 1 when x_t reaches the end of its duration[2]. Since we are using the M-phase discrete Coxian to model duration, m_t represents the current phase of x_t. Last, y_t is the observed alphabet.

[1] When considering continuous Coxian, the geometric distribution is replaced by its continuous counterpart, the exponential distribution.

[2] In an HSMM, t is the end of duration of the state x_t iff $x_t \neq x_{t+1}$. However, in an S-HSMM, it is possible that x_{t+1} is actually part of a newly initiated HSMM. Thus $x_{t+1} \neq x_t$ if $e_t = 1$ and $\epsilon_t = 0$, but we can have $x_{t+1} = x_t$ if $e_t = \epsilon_t = 1$.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

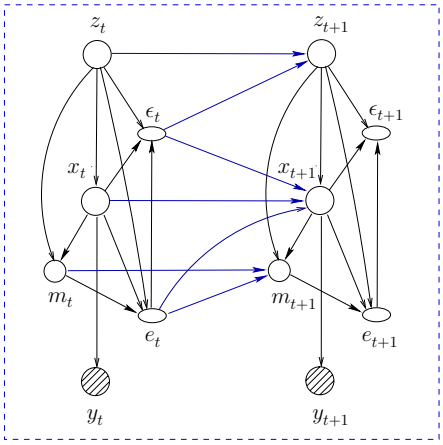

Figure 1. DBN representation of the S-HSMM for two time slices.

The parameters of this DBN are constructed from the parameters of the S-HSMM in a way similar to the hierarchical HMM [1, 14]. Intuitively, the "ending" variables ϵ_t and e_t act like the context defining how the next time slice $t+1$ can be derived from the current time slice t. When $e_t = 0$, the same states at the top and bottom levels carry on to the next time slice. When $e_t = 1$, there are two possibilities: if $\epsilon_t = 0$, the same top-level state carries on to the next time slice, but the semi-Markov sequence at the bottom level transits to a new child state; if $\epsilon_t = 1$, the top-level state "switches" to the next state, and a new semi-Markov sequence is initiated at the bottom level.

In addition, the transition of the phase variables m_t follows the parameters of the Coxian duration model. When $e_t = 0$, we have $m_{t+1} \in \{m_t, m_t + 1\}$ and the probability of staying in the same phase is (we use the short-hand notation s_t^k to denote the event $\{s_t = k\}$):

$$\begin{aligned} \Pr(m_{t+1}^m \,|\, m_t^m, x_{t+1}^i, z_{t+1}^p, e_t^0) &= 1 - \lambda_m^{p,i} \text{ for } m < M \\ \Pr(m_{t+1}^M \,|\, m_t^M, x_t^i, z_t^p, e_t^0) &= 1 \end{aligned}$$

When $e_t = 1$, the starting phase for a new semi-Markov sequence is initialized:

$$\Pr(m_{t+1}^m \,|\, x_{t+1}^i, z_{t+1}^p, e_t^1) = \mu_m^{p,i}$$

Finally, $e_t = 1$ only when the m_t is in the last phase M:

$$\Pr(e_t = 1 \,|\, m_t^m, x_t^i, z_t^p) = \begin{cases} 0 \text{ if } m < M \\ \lambda_M^{p,i} \text{ if } m = M \end{cases}$$

2.4 Inference and parameter estimation

When applying the S-HSMM to activity modeling, we would like to learn the parameters of the S-HSMM from training data and then use the learned model for classifying, segmenting and detecting abnormality in new activity sequences. Since we have derived a DBN equivalent to a given S-HSMM, existing learning and inference methods for DBNs can be readily applied to these problems [13].

In the inference task, let $S_t \triangleq \{z_t, \epsilon_t, x_t, e_t, m_t\}$ be the amalgamated hidden state; we are interested in computing the filtering distribution $\Pr(S_t \,|\, y_{1:t})$, and the smoothing distributions $\Pr(S_t \,|\, y_{1:T})$ and $\Pr(S_t, S_{t+1} \,|\, y_{1:T})$. A range of queries regarding the current high-level activity (z_t), the current atomic activity (x_t), and the remaining duration of the current activity can be answered from the marginals of these distributions. Using the familiar forward/backward procedures for HMM [20], the complexity for computing these distributions is $O(|Q|^2|Q^*|^2M^2T)$, or $O(|Q|^2|Q^*|^2M^2)$ for each filtering step. However, since the phase variables are constrained so that $m_{t+1} \in \{m_t, m_t + 1\}$, the full joint probability of m_t and m_{t+1} can be represented in just $O(M)$ space instead of $O(M^2)$. This reduces the overall complexity to $O(|Q|^2|Q^*|^2MT)$, or $O(|Q|^2|Q^*|^2M)$ per filtering step.

On the surface, this complexity term seems to be identical to the complexity of explicit duration HMMs [10]; however, note that for the explicit multinomial duration model, M is the same as the maximum possible duration length L, which in theory can be as large as the length of the observation sequence T. For the discrete Coxian duration model, as we show in section 3, we can choose $M \ll L$, and at the same time avoid the problem of having to determine L in advance.

In the learning task, we note that the set of parameters for the S-HSMM $\theta = \{\pi^*, A^*, \pi, A, B, \mu, \lambda\}$ ties together different parameters for the DBN. The S-HSMM thus can be viewed as a member of the exponential family [5] with parameter θ. Given a sequence of training data of the form $y_{1:T}$, the maximum likelihood parameter $\theta^* = \text{argmax}_\theta \Pr(y_{1:T} | \theta)$ can be estimated iteratively using the EM algorithm. This involves first computing the expected sufficient statistics (ESS) that can be derived from the smoothing distribution. The reestimated parameters are then set to the normalized values of the ESS. Due to space restrictions, we give reestimation formulas only for the parameters of the discrete Coxian duration model below. In the reestimation formula for $\hat{\mu}_m^{p,i}$, note that e_t^1 is true by definition. Further, notice that the number of free parameters for the Coxian duration model is $|Q||Q^*|(2M-1)$, which is usually much smaller than $|Q||Q^*|(L-1)$ for the explicit duration model.

$$\begin{aligned} \hat{\lambda}_m^{p,i} &= \frac{\sum_{t=1}^{T-1} \Pr(m_{t+1}^{m+1}, m_t^m, x_{t+1}^i, z_{t+1}^p, e_t^0 \,|\, y_{1:T}, \theta)}{\sum_{t=1}^{T-1} \Pr(m_t^m, x_t^i, z_t^p, e_t^0 \,|\, y_{1:T}, \theta)}, m < M \\ \hat{\lambda}_M^{p,i} &= \frac{\sum_{t=1}^{T} \Pr(e_t^1, m_t^M, x_t^i, z_t^p \,|\, y_{1:T}, \theta)}{\sum_{t=1}^{T} \Pr(m_t^M, x_t^i, z_t^p \,|\, y_{1:T}, \theta)} \\ \hat{\mu}_m^{p,i} &= \frac{\sum_{t=0}^{T-1} \Pr(m_{t+1}^m, x_{t+1}^i, z_{t+1}^p, e_t^1 \,|\, y_{1:T}, \theta)}{\sum_{t=0}^{T-1} \Pr(x_{t+1}^i, z_{t+1}^p, e_t^1 \,|\, y_{1:T}, \theta)} \end{aligned}$$

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

3 Experimental Results

We apply the S-HSMM to the problem of learning and online classification in sequences of activities in the home. We consider a typical morning routine consisting of six high-level activities: *[1] entering-the-room & making-breakfast*, *[2] eating-breakfast*, *[3] washing-dishes*, *[4] making-coffee*, *[5] reading-morning-newspaper & having-coffee*, and *[6] leaving-the-room*. The routine generally follows the sequence *[1-2-3-4-5-6]* or *[1-2-4-5-3-6]*, depending on whether the person washes the dishes before or after having coffee. The kitchen is quantized into 28 square cells of $1m^2$ each. The six activities and their typical trajectories are shown in Fig. 2. The shaded regular polygons in the walking path imply that the person does not simply walk past the cell, but actually spends some time at the region (the darker the polygons, the longer the time). For example, in the first activity (*entering-the-room & making-breakfast*), the occupant first walks to the fridge from the kitchen door, and spends some time there (5 - 7s) taking out the food, as indicated by a dark polygon in cell number 13, and then goes to the stove and stays there (10 - 15s) cooking breakfast, as illustrated by a darker polygon in cell number 5. The scene is captured by four cameras mounted at the ceiling corners, and a multiple-camera tracking module is used to detect movement and return the list of cells visited by the person. The tracking module sometimes returns a neighboring cell instead of the actual cell occupied by the person, so an observation model is estimated offline with manually labeled ground truth [17]. A total of 40 unlabeled, unsegmented sequences of cells as returned by the tracking module are used for training, with another 40 sequences used for testing. Each sequence consists of six activities with total length of approximately 140 time slices. For evaluation of abnormality detection, we capture 18 abnormal sequences where a person spends too little or too much time at some location. We also assume that the number of activities (six) and their spatial extent (the estimated set of cells visited during the activity) are known.

We use the data to train four different models: a Coxian duration S-HSMM, a multinomial explicit duration S-HSMM, a 2-layer hierarchical HMM, and a multinomial explicit duration flat HSMM. For the S-HSMM and the HHMM, we let $|Q^*| = 6$, $|Q| = 28$, and use the spatial extent of each activity to define the set of children $ch(p)$ for each activity $p \in Q^*$. Each bottom-level state $i \in Q$ thus represents the atomic activity within the i-th cell such as passing through, cooking at the stove, eating at dining table or rummaging through fridge. For the Coxian duration S-HSMM, the number of phases M is 3, and the initial phase distribution is fixed to $\mu_1 = 1, \mu_2 = 0, \mu_3 = 0$; for the explicit duration S-HSMM, the number of phases M is set to 35, which is the maximum time span of any individual

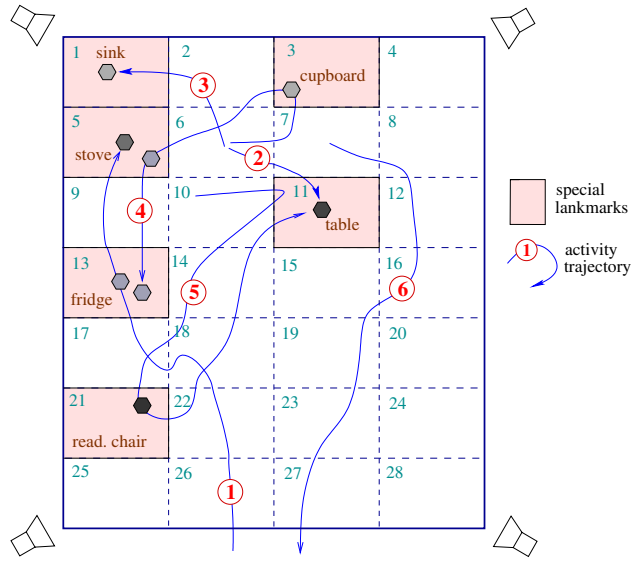

Figure 2. The morning routine consists of activities *[1]-[6]*.

activity (assumed known in advance). The flat HSMM has only a single layer with $|Q| = 28$. All these models have a fixed observation model B obtained offline from the characteristics of the tracking module. Except for the constraints mentioned here, all other parameters of these models are initialized randomly.

3.1 Learning and online classification results

Through examining the learned parameters of these models after training, we find that the S-HSMM variants adequately capture the patterns exhibited in the training data, while the 2-layer HHMM fails to do so. To illustrate, the left matrix below is the transition between the six high-level activities A^*_{pq} obtained from the multinomial S-HSMM (the Coxian S-HSMM yields a similar result), while the right one is obtained from the 2-layered HHMM. While the S-HSMM has learned reasonable transitions (from activity *[2]* to *[3]* or *[4]*; from activity *[3]* to *[4]* or *[6]*), the HHMM fails to capture these transitions.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.18 & 0 & 0.10 & 0.72 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0.05 & 0.19 & 0 & 0.76 \\ 0 & 0.06 & 0 & 0 & 0.07 & 0.87 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0.15 & 0.65 & 0 & 0 & 0 & 0.2 \\ 0.83 & 0 & 0 & 0.17 & 0 & 0 \end{bmatrix}$$

Fig. 3 shows an example of the duration model learned by different S-HSMMs for the state "at-stove" at the bottom level associated with activity *[4]* (*making-coffee*). The Coxian model tends to lean to the left as compared with the multinomial explicit model; however, it does an adequate job at smoothing out the spikes in the multinomial model. For comparison, we also smooth the multinomial

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

duration distribution using a simple moving-window averaging method.

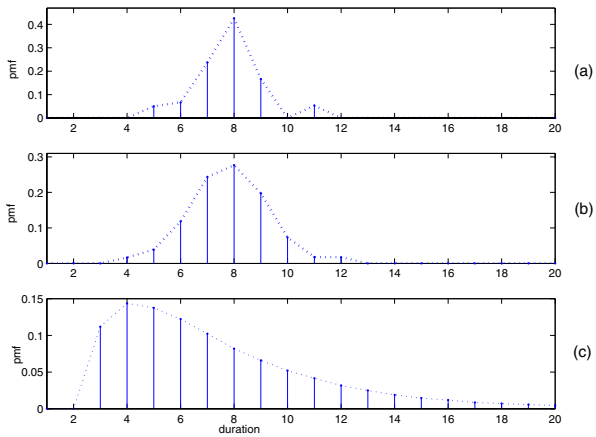

Figure 3. Duration distribution learned for the atomic activity 'at-stove' of the activity *[4]* (making-coffee) by a multinomial S-HSMM: (a) before smoothed, (b) after smoothed by moving-window averaging; and (c) by a 3-phase Coxian S-HSMM.

In the online classification test, we use the learned models for segmenting and classifying segments of the test sequences into the six high-level activities (the flat HSMM is not included in the test since it cannot model the high-level activities). The filtering distributions $\Pr(z_t|y_{1:t})$ and the most likely label z_t are computed for each time t. The label z_t at 95% percent of the length of each true segment is used to measure the classification accuracy. In addition, if a segment starts from t_0 and t is the earliest time from which the label z_t remains accurate, then $(t-t_0)/segmentlength$ is used as the measure for early detection performance.

The complete result on classification accuracy is presented in Table 1. The result shows that modeling the duration as ether multinomial or discrete Coxian works reasonably well. While smoothing the multinomial results in a small improvement, the Coxian model yields the best performance. Since the 2-layer HHMM has not learned an adequate transition model at the high level, it performs very poorly as expected. In early detection (Table 2), the Coxian S-HSMM performs slightly worse in some cases, but all the S-HSMM variants are capable of recognizing the activities earlier than 30% of their life span, which is reasonably adequate for real deployment. Table 2 also shows that it takes longer time to decide the occurances of activities *[3], [4]* and *[6]* than the rest. It is because the morning routine generally follows activities in the order *[1-2-3-4-5-6]*, but sometimes conforms to *[1-2-4-5-3-6]*, thus, each of activities *[3], [4]* and *[6]* can be reached from more than one activity, while the others can be reached from only one defined activity.

An example of the online segmentation process is shown in Fig. 4. While the segmentation obtained by the Coxian S-HSMM is fairly accurate, the result for the 2-layer HHMM is rather inconsistent: it correctly detects activities *[1]*, *[4]* and *[5]*; however, it becomes confused during the remaining major portion of the sequence.

	Avg. accuracy of each activity (%)					
	Act.1	Act.2	Act.3	Act.4	Act.5	Act.6
Cox	100	100	92.5	95	100	97.5
S-mul	100	100	87.5	95	100	97.5
UnS-mul	100	97.5	87.5	95	100	97.5
HHMM	5	0	0	35	95	0

Table 1. S-HSMM: activity recognition at the top level obtained with the S-HSMM when the durations are un-smoothed multinomial (UnS-mul), smoothed multinomial (S-mul) and Coxian (Cox), and with the 2-layer HHMM.

Activity no.	1	2	3	4	5	6
Cox	0	4.67	26.85	25.23	4.70	29.20
S-mul	0	4.82	21.73	24.12	2.29	26.05
UnS-mul	0	5.19	26.96	24.13	2.18	27.22

Table 2. S-HSMM: Early Detection rates (%). Notice that activities with shorter life spans, such as *[3],[4]* and *[6]*, will naturally result in higher early detection rates.

In addition to having a favorable performance result, the Coxian S-HSMM reduces learning and filtering time by a factor of 4. In our MATLAB implementation, the filtering computation per time slice is approximately $3s$ for the multinomial, and only $0.8s$ for the Coxian.

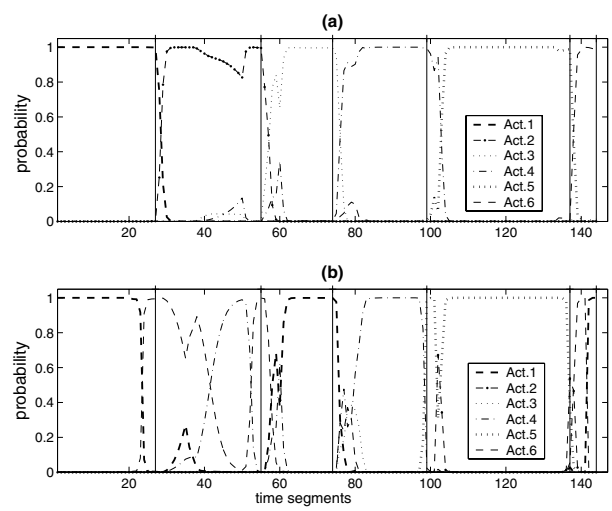

Figure 4. Comparing (a) Coxian S-HSMM with (b) 2 layer HHMM in activity recognition for a sequence comprised of activities in order [1-6]. The vertical solid lines are the true segmentation of activities.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

3.2 Duration abnormality detection

Abnormality in the duration of activities, if detected, can provide an important clue to an alert system. For example, in the elder care domain, an person staying at a location for a longer duration than usually observed might indicate the onset of illness. Since the S-HSMM can capture the normal patterns in duration spent at each location, it can also be used to detect abnormality in new activity sequences.

We implement an online abnormality detection scheme in activity sequence as follows. Suppose that at time t, the online classification algorithm has recognized that p is the winning activity in the period starting from some $t_p \leq t$. The decision to classify p as normal or abnormal is based on examining the likelihood ratio $R^p(t) = \frac{\Pr(y_{t_p:t}|\theta^p)}{\Pr(y_{t_p:t}|\bar{\theta}^p)}$ where θ^p is the parameter of the p-initiated semi-Markov sequence (the learned normal model for p), and $\bar{\theta}^p$ is the abnormal model for p. The abnormal model $\bar{\theta}^p$ is the same as θ^p except that the duration parameter is either replaced by a uniform distribution, or is "inverted", where the inverted distribution of $Mult(\mu_i)$ is $Mult(\bar{\mu}_i)$ where $\bar{\mu}_i = \frac{\max(\mu)-\mu_i}{M*\max(\mu)-1}$. A Coxian distribution model can be inverted by first approximating it using a multinomial distribution. The abnormal model θ^p constructed by only inverting the duration model suffices to capture abnormalities since our aims as mentioned previously focus on detecting a more subtle form of abnormality, which is the abnormalities only in the state duration, and not in the state order, as this form of abnormality often presents in the elder care domain.

The unseen testing data includes 22 normal sequences, and 18 sequences in which some activities contain abnormal duration. Fig. 5 shows the two receiver operating characteristic (ROC) curves resulting from hypothesis testing with the two likelihood ratios: one uses an inverted duration model of the smoothed multinomial S-HSMM, and the other uses a uniform duration model. Both curves, especially the inverted S-HSMM's curve, climb rapidly toward the upper left corner of the graph, indicating high true positive rates and low false positive rates. The optimal choice would be around 84%, and 80% true positive rates for the inverted and uniform S-HSMMs, respectively at the expense of 10% false positive rate. Hence, our inverted S-HSMM outperforms the uniform S-HSMM by about 4%. Given that detection is performed online on unsegmented sequences, the results obtained are promising.

We also compare the use of the S-HSMMs versus a flat HSMM in abnormality detection. Since the HSMM cannot segment the sequence into the six activities, it learns a normal duration model at each cell location for the entire morning routine. This makes the HSMM less flexible and unable to isolate the abnormal segments in a sequence. Fig. 6 shows an example of a sequence where abnormality occurs in the first two activities, while the situations are back to normal in the remaining four activities. While the S-HSMM successfully deals with this scenario, the HSMM continues to label the sequence as abnormal until the the sequence is about to end.

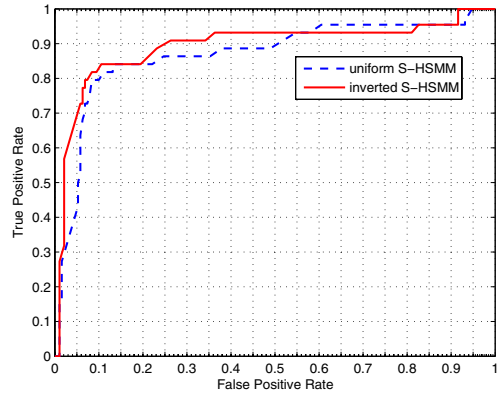

Figure 5. ROC curves for abnormality detection.

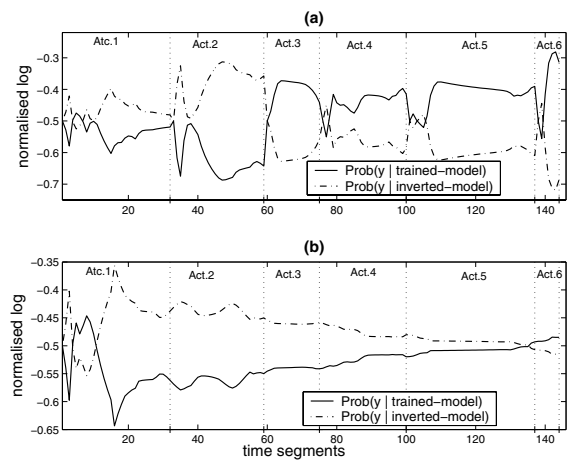

Figure 6. Abnormality detection with (a) the S-HSMM, (b) the flat HSMM, where all, except for the first two activities, are normal.

4 Conclusions

We have presented the formal foundation and experimental validation for the novel switching hidden semi-Markov model. The model can learn what an occupant normally does during the day from unsegmented training data, and then perform online activity classification, segmentation, and abnormality detection in sequences. In addition, the novel use of the discrete Coxian distribution in modeling duration has resulted in an important improvement in comparison with the classic explicit model using multinomial

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

distribution. Furthermore, the experiments also demonstrate the superiority of the model over the existing HHMM and the flat HSMM in online activity recognition and duration abnormality detection tasks.

In other more complex domains, a full hierarchical model might be needed, and the framework presented here can be extended to a full hierarchical hidden semi-Markov model. In addition, the use of the Coxian duration model allows us to control the model complexity by varying the number of phases M of the Coxian. We plan to apply a cross-validation technique and other model selection methods to this interesting problem in the future.

Acknowledgement

Hung Bui is supported by the Defense Advanced Research Projects Agency (DARPA), through the Department of the Interior, NBC, Acquisition Services Division, under Contract No. NBCHD030010.

References

[1] H. H. Bui, D. Q. Phung, and S. Venkatesh. Hierarchical Hidden Markov Models with General State Hierarchy. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*, pages 324–329, San Jose, California, 2004.

[2] H. H. Bui, S. Venkatesh, and G. West. Policy Recognition in the Abstract Hidden Markov Model. *Journal of Artificial Intelligence Research 17*, pages 451–499, 2002.

[3] S. Fine, Y. Singer, and N. Tishby. The Hierarchical Hidden Markov Model: Analysis and Applications. *Machine Learning*, 32(1):41–62, 1998.

[4] M. J. F. Gales and S. J. Young. The Theory of Segmental Hidden Markov Models. Technical Report CUED/F-INFENG/TR133, Cambridge University Engineering Department, June 1993.

[5] D. Geiger. Graphical Models and Exponential Families. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, pages 156–165, San Francisco, 1998.

[6] H. Kautz, O. Etzioni, D. Fox, and D. Weld. Foundations of Assisted Cognition Systems. Technical report, University of Washington, CSE, March 2003.

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

[8] S. Luhr, H. H. Bui, S. Venkatesh, and G. West. Recognition of Human Activity through Hierarchical Stochastic learning. In *International Conference on Pervasive Computing and Communication*, 2003.

[9] S. Luhr, S. Venkatesh, G. West, and H. H. Bui. Duration Abnormality Detection in Dequences of Human Activity. Technical report, Department of Computing, Curtin University of Technology, May 2004.

[10] C. Mitchell, M. Harper, and L. Jamieson. On the Complexity of Explicit Duration HMMs. *IEEE Transactions on Speech and Audio Processing*, 3(3), May 1999.

[11] C. D. Mitchell and L. H. Jamieson. Modeling Duration in a Hidden Markov Model with the Exponential Family. In *Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages II.331–II.334, Minneapolis, April 1993.

[12] K. Murphy. Learning Switching Kalman Filter Models. Technical report, Campaq Cambridge Research Lab, 1998.

[13] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California at Berkeley, Computer Science Division, 2002.

[14] K. Murphy and M. Paskin. Linear-time inference in Hierarchical HMMs. In *Advances in Neural Information Processing Systems*, Cambridge, MA, 2001. MIT Press.

[15] K. P. Murphy. Hidden semi-Markov models (HSMMs), unpublished notes, 2002.

[16] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models*. The Johns Hopkins University Press, Baltimore and London, 1981.

[17] 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, August 2004.

[18] N. Oliver, E. Horvitz, and A. Garg. Layered Representations for Human Activity Recognition. In *Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, pages 3 – 8, October 2002.

[19] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose. Sporadic State Estimation for General Activity Inference. Technical report, Intel Research Seattle and the University of Washington, July 2004.

[20] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286, February 1989.

[21] 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.

[22] T. Starner and A. Pentland. Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. In *Proceedings of SCV'95*, pages 265–270, 1995.

[23] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden Markov model. In *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pages 379–385, 1992.