Duration Abnormality Detection in Sequences of Human Activity

Sebastian Lühr† Svetha Venkatesh† Geoff West† Hung H. Bui⋆

†Dept. of Computing, Curtin University of Technology, GPO BOX U1987, Perth, Western Australia

⋆Artificial Intelligence Center, SRI International, 333 Ravenswood Ave, Menlo Park, CA

† {luhrs, svetha, geoff}@cs.curtin.edu.au ⋆bui@ai.sri.com

Abstract

Activity duration is an essential element in the accurate modelling of human behaviour. The application of a standard hidden Markov Model (HMM) for the detection of abnormality in sequences of human activity can create a situation in which highly unusual duration less than or greater than the duration normally observed can fail to be detected. In this paper[1]*, we show how the application of the explicit state duration HMM can alleviate this problem, enabling us to distinguish between sequences of activity in which the order of observations is identical but the duration of activities is different and to identify the presence of abnormal activity duration. Experimental results highlight the improvement over the standard HMM for both abnormality detection and classification in certain anomalous situations.*

1 Introduction

The deterioration of our cognitive and physical abilities as we age is as inevitable as our resulting dependency on third party care. Nursing homes and day carers traditionally fulfil this need if our relatives are unable, or unwilling, to look after us in our old age. The Intelligent Housing Project at Curtin University of Technology seeks to create an intelligent environment that is capable of caring for its occupants without the need for intrusive and costly third party carers. The system aims to learn its occupants' behavioural patterns in order to provide cognitive support and to detect abnormalities, potentially dangerous situations or emergencies in order to take appropriate action.

One of the prerequisites to this task is the learning and recognition of typical sequences of activities performed throughout the day. Much of the current work in human behaviour modelling concentrates on activity recognition, recognising actions and events through pose, movement, and gesture analysis. Our work focuses on learning and detecting abnormality in higher level behavioural patterns. The hidden Markov model (HMM) [9] is one approach for learning such behaviours given a vision tracker recording observations about a subject's activity.

In this paper, we show how implicit state duration in the HMM can create a situation in which highly abnormal deviation as either less than or more than the usually observed activity duration can fail to be detected and how the explicit state duration HMM (ESD-HMM) [10, 3, 9] helps alleviate the problem. Duration of human activity is an important consideration if we are to accurately model a person's behavioural patterns. We show that duration modelling enables us to differentiate between activity sequences in which the order of the observations is identical yet the duration of the activities is varied. Although the explicit state duration HMM has been used extensively in the field of speech recognition we believe its application to the domain of human activity recognition is novel.

The organisation of this paper is as follows. In Section 2 we provide a brief overview of related work in the recognition of human behaviours. A synopsis of state duration HMM theory is given in Section 3. Section 4 and Section 5 discuss our experimentation methodology and results respectively. Conclusions drawn from this work are presented in Section 6.

2 Related Work in Modelling Behaviour

A hidden Markov model approach to learning the behavioural patterns of people in an office environment from visual blob features was presented in [1]. An entropy minimisation technique was applied to learning the parameters of the HMM. The resulting model was shown to have greater discriminative capability in detecting anomalous behaviour in the order, tempo and timing of events than conventionally trained HMM.

Coupled hidden Markov models (CHMMs), extensions of the HMM for modelling independent yet interacting processes, were used in a vision system to model the interactions of people in an outdoor scene [8]. Models first trained

[1]This technical report is an extended version of [5]. TR-2004/02.

on data from synthetic agents and then updated using real world data were able to detect and classify human interactions with good results.

Nguyen et al [7] introduced a behaviour recognition system based on analysis of trajectories using the abstract HMM, an extension of the HMM in which the Markov chain is replaced by a hierarchy of Markov chains. The system was able to label behaviours at different levels of resolution. On subsequent work the model was extended to allow recognition of sequences [6].

A stochastic context free grammar (SCFG) parser for identifying high level events in a car park was discussed in [2]. A vision based tracker monitored the movement of pedestrians and vehicles throughout the scene. Significant changes in the spatio-temporal features of the tracked objects were mapped to discrete events and passed to the grammar parser for labelling.

The hierarchical HMM, a special case of SCFG, has similarly been applied to learning and recognising simple sequences of human activity by observing a person's proximity to areas of interest in a home scenario [4].

None of the above tackles the issue of duration.

3 Explicit State Duration HMM

In the standard HMM, state duration is implied as a function of a state's self transition probability. Given a state i and its self transition probability a_{ii}, we can show that the likelihood of remaining in the state for d consecutive time steps is exponential:

$$(a_{ii})^{d-1} \cdot (1 - a_{ii}) \tag{1}$$

Explicit state duration [10, 3, 9] introduces into the HMM the duration variable $p_i(d)$ such that $1 \leq d \leq D$ where D constrains the maximum duration. In its non-parametric form p_i is a vector of discrete duration probabilities such that $\sum_{d=1}^{D} p_i(d) = 1$. The self transition probabilities are tied so that $a_{ii} = 0$.

The model re-estimation formulas are presented in equations 6–9 given the path variables in equations 2–5 and the sequence of observations $O_1, O_2 \ldots O_{t-1}, O_T$ produced over time t such that $1 \leq t \leq T$ and $O_t \in V$ where $V = \{v_1, v_2 \ldots v_M\}$ is the set of possible observation symbols.

For clarity, we use the notation $\lambda = (A, B, P, \pi)$ given the state transition distribution $A = \{a_{ij}\}$, the observation likelihood distribution $B = \{b_i(O_k)\}$, the state duration distribution $P = \{P_i(d)\}$ and the prior distribution $\pi = \{\pi_i\}$.

$$\alpha_t(i) = \Pr(O_1, O_2 \ldots O_t, S_i \text{ ends at } t|\lambda) \tag{2}$$

$$\alpha_t^*(i) = \Pr(O_1, O_2 \ldots O_t, S_i \text{ begins at } t+1|\lambda) \tag{3}$$

$$\beta_t(i) = \Pr(O_{t+1} \ldots O_T|S_i \text{ ends at } t, \lambda) \tag{4}$$

$$\beta_t^*(i) = \Pr(O_{t+1} \ldots O_T|S_i \text{ begins at } t+1, \lambda) \tag{5}$$

$$\hat{\pi}_i = \frac{\pi_i \beta_0^*(i)}{P(O|\lambda)} \tag{6}$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \alpha_t(i) a_{ij} \beta_t^*(j)}{\sum_{j=1}^{N} \sum_{t=1}^{T} \alpha_t(i) a_{ij} \beta_t^*(j)} \tag{7}$$

4 Experimentation Methodology

We recorded 150 video sequences of normal activity in a kitchen scenario using a single camera, each recording belonging to one of five normal classes of activity sequences one might observe in a kitchen; preparing cereal, making toast for breakfast, preparing or reheating dinner and cooking a bacon and eggs breakfast.

Motion in the room was segmented using a robust tracker [11] and a Kalman filter was employed to track moving objects between frames. The position of a subject's feet as approximated by taking the centre of the bottom of the tracker supplied bounding box was used to calculate the proximity of the person to defined areas of interest in the room; the stove, the kitchen bench, the sink, a fridge and the door. The discrete observations `stove`, `bench`, `sink`, `fridge` and `door` mapping to these areas were recorded if the subject was in close proximity else an `undefined` observation was logged. One such observation is recorded every one and a half seconds from the 25fps video. The duration of the video sequences ranged from 30 to 300 seconds with an average length of circa 90 seconds. Although a richer set of features from multiple camera angles would be beneficial in a real world deployment, we found these observations adequate for demonstrating the ideas behind this work.

4.1 Normal Activity Sequences

The five classes of normal behaviour were designed to highlight the importance of modelling duration given the limitations of the tracking system. That is, using an impoverished observation set, the classes would have the same sequence of observations but would differ in the duration spent in a location.

The first two classes, preparing cereal and making toast for breakfast, are identical in the order that the areas of interest in the room are visited by the person under observation. Hence, given that the order of observations returned

$$\hat{b}_i(k) = \frac{\sum_{\substack{t=1\\ \text{s.t. } O_t=k}}^{T} \left[\left(\pi_i \beta_0^*(i) + \sum_{\tau=1}^{t-1} \alpha_\tau^*(i) \beta_\tau^* \right) - \left(\sum_{\tau=1}^{t-1} \alpha_\tau(i) \beta_\tau(i) \right) \right]}{\sum_{k=1}^{M} \sum_{\substack{t=1\\ \text{s.t. } O_t=v_k}}^{T} \left[\left(\pi_i \beta_0^*(i) + \sum_{\tau=1}^{t-1} \alpha_\tau^*(i) \beta_\tau^* \right) - \left(\sum_{\tau=1}^{t-1} \alpha_\tau(i) \beta_\tau(i) \right) \right]} \tag{8}$$

$$\hat{p}_i(d) = \frac{\pi_i p_i(d) \beta_d(i) \prod_{s=1}^{d} b_i(O_s) \quad + \sum_{t=1}^{T-d} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s)}{\sum_{d=1}^{D} \left[\pi_i p_i(d) \beta_d(i) \prod_{s=1}^{d} b_i(O_s) + \sum_{t=1}^{T-d} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s) \right]} \tag{9}$$

by the tracker are [`door`, `fridge`, `bench`, `sink`, `bench`, `fridge`, `door`] it is only possible to distinguish between the two classes by observing the time spent at the kitchen bench, the act of making toast taking considerably longer than the preparation of a bowl of cereal.

Similarly, the dinner preparation and reheating classes consist of the activities [`door`, `fridge`, `bench`, `stove`, `door`], the classes differing only in the duration spent standing by the stove. The fifth class is made up of the activities [`door`, `fridge`, `bench`, `sink`, `bench`, `stove`, `fridge`, `door`] and was included because it differs to the other classes in both the activity duration and the order in which the activities are performed. The recorded video sequences were evenly distributed among the five classes.

4.2 Abnormal Activity Sequences

A further 24 sequences of abnormal behaviour were recorded. The abnormal sequences differ from the normal only in terms of activity duration, either shorter or longer than the durations seen in the normal classes, not in the order or type of activities seen.

4.3 Model Selection

Each normal class was modelled using a standard fully connected HMM, a left-right HMM, a fully connected explicit state duration HMM (ESD-HMM) and a left-right ESD-HMM. The left-right models were chosen to investigate how constraining the state transitions would affect classification and abnormality detection by preventing the models from treating duration as a cyclic activity. The HMM was selected as a baseline for comparison. The models were trained on a random sample of 60% of the normal activity sequences and tested on the remainder. To keep the comparison fair, an optimal number of states for each model was empirically selected based on classification accuracy.

A single Gaussian distribution was used to estimate the duration probabilities in the ESD-HMM case, the model otherwise requiring an unrealistic amount of training data to accurately estimate the state duration probabilities.

5 Results

First we compare the classification accuracy of the various models. Duration abnormality detection is then discussed. Finally, we examine how the models function under varying degrees of duration abnormality.

5.1 Sequence Classification

Classification accuracy and the optimal number of states for each of the four models are presented in Figure 5. The models were trained on 60% of the normal activity sequences and tested on the remainder.

The HMM was found to be the weakest model for classification. Its low score is attributed to dynamic time warping, a property which renders it unsuitable for use as a classifier given the type of observation sequences used in this experiment. This is also evident in its relatively poor classification of the original training data.

Forcing the HMM to be a left-right model, that is $A_{ij} = 0$ for all $j < i$, appears to improve classification accuracy with near perfect results. Although the two state left-right HMM performed well empirically, the limited number of parameters is inadequate to properly encode the sequences and hence discriminate between classes. Confusion between the similar activity classes is shown in Figure 2(a).

The ESD-HMM, in contrast, appears to perform well given no state transition restrictions, providing no room for further improvement when left-right constraints are imposed. Explicit duration allows the model to clearly differentiate between all classes. The confusion matrix for the

Model	# States	Training	Testing
HMM	12	93.75%	81.43%
Left-right HMM	2	98.75%	97%
ESD-HMM	3	100%	100%
Left-right ESD-HMM	2	98.75%	100%

Figure 1. The optimal number of states and classification accuracy for both the training and the test sequences for each of the four model types.

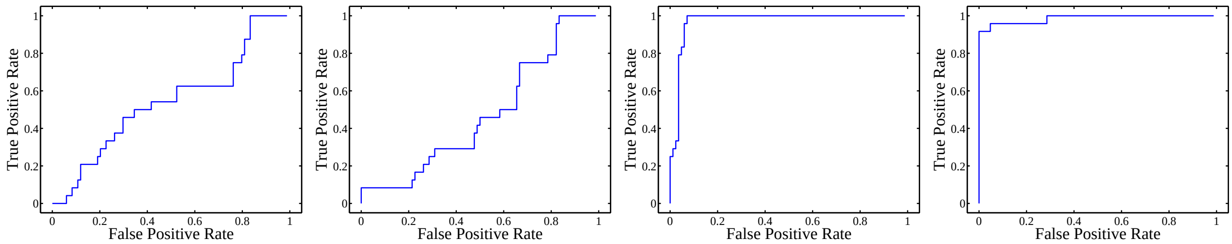
	Cereal	Toast	Cook Dinner	Reheat Dinner	Bacon & Eggs
Cereal	11	3	0	0	0
Toast	0	14	0	0	0
Cook Dinner	0	0	10	4	0
Reheat Dinner	0	0	6	7	0
Bacon & Eggs	0	0	0	0	14

(a)

	Cereal	Toast	Cook Dinner	Reheat Dinner	Bacon & Eggs
Cereal	14	0	0	0	0
Toast	0	14	0	0	0
Cook Dinner	0	0	14	0	0
Reheat Dinner	0	0	0	14	0
Bacon & Eggs	0	0	0	0	14

(b)

Figure 2. Test sequence confusion matrix for (a) the twelve state standard HMM and (b) both the three state ESD-HMM and the two state left-right ESD-HMM.



(a) (b) (c) (d)

Figure 3. ROC curves for (a) the twelve state HMM, (b) the two state left-right HMM, (c) the three state ESD-HMM and (d) the two state left-right ESD-HMM.

explicit state duration models is presented in Figure 2(b).

5.2 Duration Abnormality

We wish to classify unseen observation sequences as either normal or abnormal. We do so by querying each of the trained models on the likelihood of generating a given sequence and then thresholding on the highest log likelihood returned. The log likelihoods are normalised by the total length of the observation sequence so that a global threshold may be applied regardless of sequence length. Receiver operator characteristic (ROC) curves were used to investigate the suitability of each of the models as a detector of abnormality. The observation sequences used in this experiment consisted of the set of unseen normal sequences and the abnormal sequences described in Section 4.

The ROC curves in Figure 3(a) and Figure 3(b) for the HMM and the left-right HMM respectively suggest that neither model is able to reliably differentiate between our normal and abnormal sequences using the thresholding approach.

The ROC curve for the ESD-HMM displays better results, Figure 3(c) showing the true positive rate increasing more rapidly than the false positive rate. The use of explicit state duration has increased the reliability of the HMM in the detection of abnormality due to the presence of unusual activity duration. The main cause of the remaining misclassification was found to be due to the model freely transitioning between states. The model will temporarily enter a state with sub-optimal emission probabilities prior to returning to the original state so as to maximise the state duration likelihoods over the entire length of a given sequence.

Further improvement is seen when the transition constraints of the two state left-right ESD-HMM are imposed as evidenced by the steep ascent of the true positive rate in Figure 3(d).

An analysis of errors showed that two of the normal activity sequences had been misclassified by the explicit state duration models because they contained a noisy observation, uncommon and not present in the training data, in the middle of a typically long activity. The models were forced to make a transition to another state in order to emit the rogue observation, the transition resulting in very low state duration probabilities.

5.3 Longer Term Duration Abnormality

To investigate the ability of the models to detect longer term abnormal duration we artificially varied the time spent at a primary activity, standing near the kitchen bench, in a

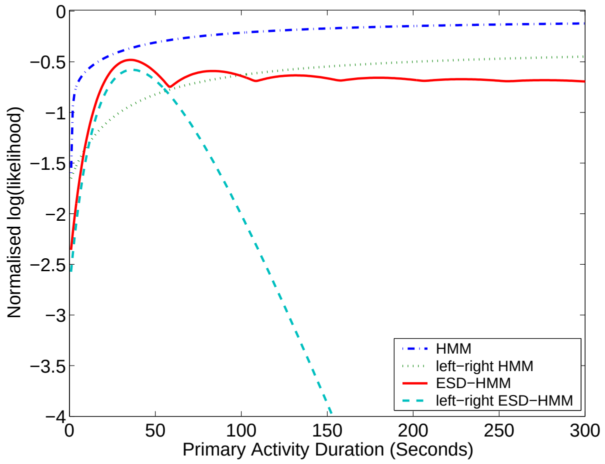


Figure 4. The normalised log likelihood for each of the models as the primary activity in an observation sequence is varied from one second to five minutes. The normal duration for the primary activity is circa 40 seconds. Only the left-right ESD-HMM is able to detect abnormality in a timely manner.

randomly selected test sequence from the first activity class. The time spent at the kitchen bench was varied from one second to five minutes. The usual time for a subject to remain at the kitchen bench is circa forty seconds.

The likelihood, normalised by the length of the observation sequence, of the modified activity sequence being generated by each of the standard HMM, left -right HMM, ESD-HMM and left-right ESD-HMM was plotted over the duration period and is presented as Figure 4.

The figure shows the normalised log-likelihood returned by the HMM and the left-right HMM increasing with the time spent at the primary activity due to dynamic time warping. The HMM and left-right HMM are therefore not suitable for the detection of highly abnormal activity duration.

The ESD-HMM exhibits a similar trend. The lack of transition constraints allows the model to temporarily enter a state with a sub-optimal emission probability in order to maximise the state duration likelihoods.

The left-right ESD-HMM behaves correctly given the intention of our system as the model is unable to explain away highly abnormal duration as a cyclic activity and thus identify them as abnormal. The curve is seen to drop rapidly as the time spent at the kitchen bench increases.

6 Conclusion

Experimental results highlighting the importance of explicit duration modelling for correct classification of sequences of human activity and the reliable and timely detection of duration abnormality have been presented. It has been shown that explicit duration modelling with the addition of left-right transition constraints are necessary if we are to identify abnormality in activity duration using the explicit state duration HMM.

The incorporation of duration in models of human behaviour is an important consideration for intelligent environments seeking to provide cognitive support and to detect deviation in the day-to-day behavioural patterns of the elderly.

References

[1] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):844–851, 2000.

[2] Y. A. Ivanov and A. F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):852–872, August 2000.

[3] S. E. Levinson. Structural methods in automatic speech recognition. *Proc. the IEEE*, 73(11):1625–1650, Nov 1985.

[4] S. Lühr, H. H. Bui, S. Venkatesh, and G. A. W. West. Recognition of human activity through hierarchical stochastic learning. In *IEEE Int'l Conf. Pervasive Computing and Communications*, Texas, USA, March 2003.

[5] S. Lühr, S. Venkatesh, G. W. West, and H. H. Bui. Explicit state duration HMM for abnormality detection in sequences of human activity. In *Proc. 8th Pacific Rim Int'l Conf. Artificial Intelligence*, volume 3157 of *Lecture Notes in Artificial Intelligence*, pages 983–984. Springer-Verlag, August 2004.

[6] N. T. Nguyen, H. H. Bui, S. Venkatesh, and G. West. Recognising and monitoring high-level behaviours in complex spatial environments. In *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Wisconsin, USA, June 2003.

[7] N. T. Nguyen, S. Venkatesh, G. West, and H. H. Bui. Hierarchical monitoring of people's behaviours in complex environments using multiple cameras. *Int'l Conf. Pattern Recognition*, August 2002.

[8] N. M. Oliver, B. Rosario, and A. P. Pentland. A Bayesian computer vision system for modeling human interactions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):831–843, August 2000.

[9] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. the IEEE*, 77(2):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 *Int'l Conf. Acoustics Speech and Signal Processing*, pages 5–8, 1985.

[11] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):747–757, August 2000.