Non-stationary Hidden Semi Markov Models in Activity Recognition

Einat Marhasev*
Caesarea Rothschild Institute for
Interdisciplinary Applications of Computer Science
University of Haifa
Haifa, 31905 Israel
einatm@cs.haifa.ac.il

Meirav Hadad and **Gal A. Kaminka**
Department of Computer Science
Bar-Ilan University
Ramat-Gan, 52900 Israel
{hadad, galk}@cs.biu.ac.il

Abstract

Activity recognition is a process by which the ongoing observed behavior of an agent is tracked and mapped to a given model, explaining the behavior and accounting for hidden or unobservable state (e.g., goals or beliefs of the observed agents). Various methods for activity recognition exist. A popular family of such methods rely on Hidden Markov Models HMMs and variants for recognition. These models, however, do not account for changes in *transition* probabilities based on the duration an agent has spent in a given state. This paper investigates Markov models that go beyond existing models, to explicitly model the dependency of transition probabilities on state duration. In particular, we propose the use of *Non-stationary Hidden Semi Markov Models* (NHSMMs) in activity recognition. We present the NHSMM model, and compare its performance in recognizing normal and abnormal behavior, using synthetic data from an industry simulator. We show that for relatively simple activity recognition tasks, both HSMMs and NHSMMs easily and significantly outperform HMMs. In more complex tasks, the NHSMMs also outperform the HSMMs, and allow significantly more accurately recognition.

Introduction

Activity recognition is a process by which the ongoing observed behavior of an agent is tracked and mapped to a given model, explaining the behavior and accounting for hidden or unobservable state (e.g., goals or beliefs of the observed agents). It is an important task in surveillance (Bui, Venkatesh, & West 2002; Hongeng & Nevatia 2003; Duong *et al.* 2005), assistive technology, and user-modeling.

Various methods for activity recognition exist. A popular family of such methods rely on Hidden Markov Models HMMs and variants for recognition. Recently, as it became clear that the recognition of many activities has a strong temporal component, HMMs variations which take duration explicitly into account have become more popular, e.g., Hidden Semi-Markov Models (HSMMs) (Hongeng & Nevatia 2003) and Switching Hidden Semi-Markov Models (Duong *et al.* 2005). These allow explicit modeling of the duration an observed agent is expected to be in a particular recognized state. For instance, HSMMs allow modeling the fact that an observed airline passenger may spend a long time at the security check area, and much less time moving from this area to the flight boarding gate.

These models, however, do not account for changes in *transition* probabilities based on the duration an agent has spent in a given state. For instance, an airline passenger spending a short time at the check-in counter is more likely to have gone through the check-in process without glitches, and thus more likely to head towards the gate. On the other hand, a passenger spending a long time at the check-in area is more likely to have encountered problems, and thus the likelihood of the passenger heading towards a supervisor is greater. Existing models encounter difficulty modeling such distinctions, and are therefore insufficient for the complex set of recognition problems found in real-world applications.

This paper investigates Markov models that go beyond existing models, to explicitly model the dependency of transition probabilities on state duration. In particular, we propose the use of *Non-stationary Hidden Semi Markov Models* (NHSMMs) in activity recognition. Variants of NHSMMs have been previously applied in speech recognition and handwritten character recognition (Vaseghi 1995; Sin & Kim 1995), but their use in activity recognition is novel. Just as HSMMs expand HMMs to include state duration (to more accurately model realistic observed behaviors), so do NHSMMs expand HSMMs to model the transition probabilities' dependence on state duration.

We present the NHSMM model, and compare its performance in recognizing normal and abnormal behavior, using synthetic data from an industry simulator. We show that for relatively simple activity recognition tasks, both HSMMs and NHSMMs easily and significantly outperform HMMs. In more complex tasks, the NHSMMs also outperform the HSMMs, and allow significantly more accurately recognition.

Related Work

Hidden Markov Models (HMMs) and its many variants have been successfully applied to a number of scientific and engineering problems (Rabiner 1989). However, their use in activity recognition is fairly recent (Oliver, Horvitz, & Garg 2002). Bui et al. used a variant AHMMs for vision-based ac-

*This research was supported in part by IRST Trento-CRI Haifa.

tivity recognition (Bui, Venkatesh, & West 2002). Yin et al. used them to track users moving in a wireless environment (Yin, Chai, & Yang 2004). In conventional HMMs, the duration of states is not modeled explicitly. This has been shown to be problematic in recognizing real-world activities.

Hidden Semi-Markov Models have been developed to allow explicit modeling of duration probabilities. In conventional hidden semi-Markov models, the probability state duration is established by a distribution function which may be a continuous probability density function suggested by Ferguson (Ferguson 1980) or a parametric distribution such as Gaussian, Poisson or Gamma distributions (e.g, (Levinson 1986)). Yu and Kobayashi (Yu & Kobayashi 2003) propose a model for missing data and multiple observation sequences which is applied to mobility tracking in wireless networks. Hongeng and Nevatia (Hongeng & Nevatia 2003) apply HSMM for recognizing events in a video stream. Duong et al. (Duong *et al.* 2005) introduce Switching HSMM for activity recognition and abnormality detection. All of these previous works concentrate on HSMMs in which the duration distribution is not combined with the state transition probabilities.

Non-stationary Hidden Semi-Markov Models (NHSMMs) have been applied in speech technology (Vaseghi 1995) and the related field of handwritten character recognition (Sin & Kim 1995). However, the models have not been used in activity recognition, our domain of study. Moreover, their use in activity recognition requires at least some modifications from the original models.

Kaminka et al. (Kaminka, Pynadath, & Tambe 2002) have investigated overhearing by plan recognition in cooperative agent teams. They have utilized an ad-hoc recognition model that is reminiscent of hierarchical non-hidden semi-Markov models. The system used average duration of plan-step as an input to a Gamma distribution, to estimate the duration in which an agent executed, unobserved, any given step. Kaminka et al. report that the use of this model produced very poor results—due to the unmodeled variance of actual durations—and use this fact to motivate their work on using multi-agent heuristics to improve recognition rates.

Recently, several statistical models based on Bayesian filters (e.g., (Fox *et al.* 2003; Pynadath & Wellman 2000; Charniak & Goldman 1993)) have been applied for activity recognition. These works do not use statistics on state duration in their computations but highlight the necessity for methods which optimize the probabilistic recognition models for goals and activities.

The Models

We begin with a short survey of familiar models (HMMs and HSMMs), before providing a full discussion of the NHSMM model and its use in activity recognition.

Background

An HMM is defined in (Rabiner 1989) as:

1. A set of N states $S = \{S_1, S_2, \ldots, S_N\}$. A state at time t is denoted as q_t.
2. The state transition probability distribution $A = \{a_{ij}\}$, where $\{a_{ij}\}$ is the probability to transition from state S_i to state S_j. $a_{ij} = P[q_{t+1} = S_j | q_t = S_i] \qquad 1 \leq i, j \leq N.$
3. The initial state probability distribution $\pi = \{\pi_i\}$ where $\pi_i = P[q_1 = a_i] \quad 1 \leq i \leq N.$
4. A set of M observation symbols $V = \{v_1, v_2, \ldots, v_M\}$.
5. The observation symbol probability distribution in state j, $B = \{b_j(k)\}$, where

$$b_j(k) = P[s_k \text{ at } t | q_t = a_j], \quad 1 \leq j \leq N$$
$$1 \leq k \leq M.$$

The parameters of HMM are represented by $\lambda = (A, B, \pi)$. Given a sequence of observations $O = O_1 O_2 \ldots O_T$ and λ, the probability of the observing sequence can be evaluated as an especially simple form:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{1}$$

where, for any $t \in [1, T]$, the forward probability α is defined by:

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(O_t) \tag{2}$$

As noted in previous work (Levinson 1986), in the standard HMM, the transition probability distribution a_{ij} is constant in time. At each time step, the model transfers to the next state according to the transition probabilities and the symbol observed. This self transition, which is a transition from a certain state into the same state, occurs when the same symbol is observed. The self-transition probability a_{ii} for that state may be relatively high. Therefore, the standard model does not differentiate between a self transition and a transition to another state when calculating the probability of an observation sequence causing exponential probability when remaining in a certain state for a number of consecutive observations (see Figure 1).

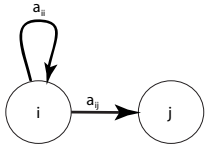


Figure 1: HMM with exponential state duration

Since it means that there is no accounting for the amount of time spent in a certain state, for many applications, this exponential state remaining probability is inappropriate. For this reason, the HSMM was devised to broaden the HMM by eliminating self transitions (see Figure 2). Formally, A is the state transition probability distribution. $A = \{a_{ij}\}$, where $\{a_{ij}\}$ is the probability to transition from state S_i to state S_j. $1 \leq i, j \leq N$, $i \neq j$ and a_{ii} is therefore replaced with duration probability density function $d_j(\tau)$ which denotes

the probability of staying at least of length duration τ in state S_j, $1 \leq \tau \leq D_j$. The rest of the HMM definitions apply for the HSMM.

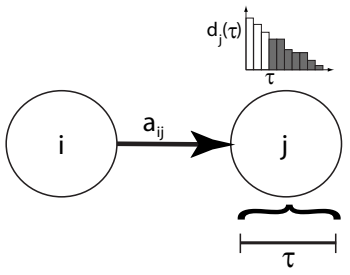


Figure 2: HSMM with explicit state duration

For any $t \in [1, T]$ and $\tau \leq t$, the forward probability α is defined by:

$$\alpha_t(j) = \sum_{\tau=1}^{D_j} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_{t-\tau}(i) a_{ij} d_j(\tau) \prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}) \quad (3)$$

The backward probability β is extended in a similar manner (see (Sin & Kim 1995)). The probability of observing a sequence, in HSMM, can be evaluated by equation1.

The Non-stationary Hidden Semi-Markov Model

In our work we consider an additional expansion of the model which includes not only the modifications of HSMM to HMM but a refinement to the state transition probability to include time dependency (see Figure 3). Formally, we define the state transition probability distribution of an NHSMM as follows. $A = \{a_{ij}(\delta)\}$, where $\{a_{ij}(\delta)\}$ is the transition probability from state S_i to state S_j after having remained in state S_i exactly $\delta - 1$ length duration. This definition of the transition probability reflects the non-stationary property we wish to capture. By assigning the transition probability to be the exact probability of spending a specific length duration before the transition to a certain state, we take into account the concept that different durations in a state affect the choice of where to transition to. As explained by the airline passenger example in the introduction, the NHSMM takes into account not only the explicit duration of a state but also the choice of where to transition to according to that exact duration. In doing so, the NHSMM improves the accuracy of the recognition abilities of the model.

Taking in consideration the time-dependent transition probability, we define the forward probability α of NHSMM as:

$$\alpha_t(j) = \sum_{\tau=1}^{D_j} \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\delta=1}^{D_i} \alpha_{t-\tau-\delta+1}(i) a_{ij}(\delta) d_j(\tau) \prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}) \quad (4)$$

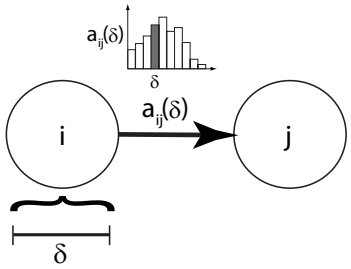


Figure 3: NHSMM with state transition probability distribution

$$\text{where} \sum_{j=1}^{N} \sum_{\delta=1}^{D_i} a_{ij}(\delta) = 1, i = 1 \ldots N$$

and we initiate the computation for each state by the probabilities to begin at that state and remain there until D_i. We denote this initiation as $\alpha_t^{\tau}(j)$:

$$\alpha_t^{\tau}(j) = \pi_j d_j(\tau - 1) \prod_{\theta=1}^{\tau} b_j(O_{t-\tau+1}) \quad (5)$$

$$j = 1 \ldots N \, , t = 1 \ldots D_j.$$

where D_j denotes the maximum length duration of being in state S_j.

Thus, given equation 4 above, for any $t \in [1, T]$ the likelihood function $P(O|\lambda)$ of NHSMM can be evaluated by equation 1.

Examining the computation involved in the calculation of the likelihood function $P(O|\lambda)$ of NHSMM shows that $\alpha_t(j)$ requires the order of $O(ND^2 + T)$ calculations where $D = argmax_{D_i}$ and therefore the computation of $P(O|\lambda)$ requires $O(N^2TD^2)$. The space involved is in the order of $ND^2 + TN$ calculations. (TN is space needed for the α values table). In the following section, we compare the performance of each model in recognizing normal and abnormal behavior. We also compare CPU times for calculating the likelihood function over time by each model. We show through experimental results that even though the computation time of HMM is significantly better than the both HSMM and NHSMM, these two models easily and significantly outperform HMM. In addition, the CPU time for calculating the likelihood function of NHSMM is slightly higher than HSMM but NHSMM outperforms HSMM and produces significantly more accurately recognition.

Experimental Analysis

To test and compare the discussed models for the task of recognizing agents' activities, we utilized a simulator which simulates up to dozens of passengers moving about a large area according to pre-defined paths (with noise in the movement). This simulator is part of an industry effort in surveillance. A screen-shot is reproduced in Figure 4.

In our experiments, the agents can move within a rectangular shaped 200 $\times$ 200 grid. Each grid cell represents a

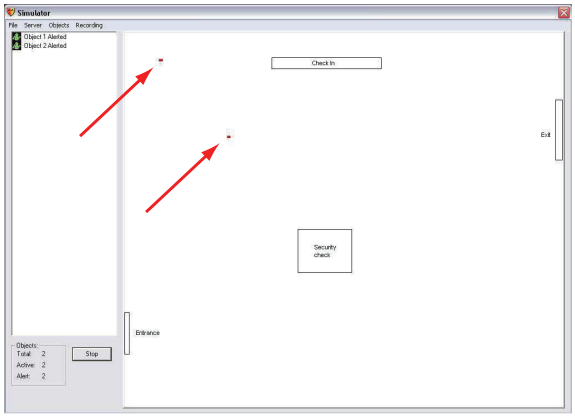


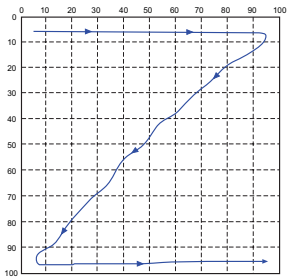
Figure 4: The dialog window of the simulator. The two arrows are pointing at two objects moving in the quadrangular area which is mapped to a XY-grid.

unique single position defined by two coordinates , X and Y, both ranging from 0 to 199. An agent's movements, which is represented by an (x, y) position in the grid followed by the specific time a movement is observed, are recorded into a file. A series of detected movements with their appropriate occurrence times represent a single path. Thus, each file includes a path which the agent has taken. We denote the time marking each movement in the recorded files as time steps. In these experiments a single second is divided into five time steps.

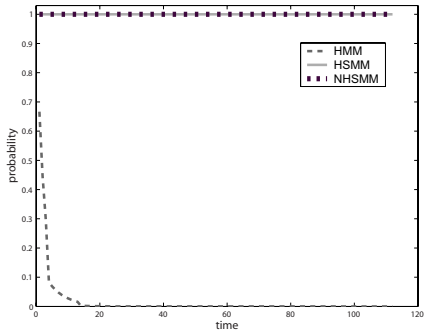
We generated output files to build training sets for our models and then we tested the models' ability to detect paths which are consistent with the training sets. The training set for our models is a group of paths which the model receives. Using the training set, the system learned all paths in the training set, and calculated the probabilities associated with the model, e.g., transition probabilities from one state to another, state-duration probabilities, etc. Since the goal of the experiments is to compare between the models' performance, we ignore the observation probability of a state which is identical in all the models (i.e., the observation probability of each state is 1).

Evaluating the models

For the first experiment, we aimed to evaluate and compare the models' performance in the case of a simple learned path, without stopping along the course and with no variance within the group. For this task, some alterations in the modeling of the XY-grid were necessary. The modeling of the XY-grid mentioned above was not appropriate for this experiment, even with the agent moving as fast as the simulator allows, suggests 5 to 10 observations relating to a single state in the model, resulting in a large duration for each state entered. We solved this problem by modeling the 200×200 XY-grid to 400 non overlapping blocks, where the size of each block is 10×10. By reducing the size of the block, the number of observed positions within a single block is lessened and therefore resulting in short durations



(a) A sketch of the simple snake-like path comprising the training set. The grid in this experiment was divided in to 10×10 size blocks.



(b) A graph showing the resulting probability values for each model, as time progresses. Results are averaged over 30-some different trials.

Figure 5: First experiment: Settings and Results.

for states entered in the models.

Figure 5(a) shows the path we chose for this first training set, a snake-like path, measuring over 120 time steps. The objects following this path were to make no stops, i.e. move continuously through the designated coordinates on the XY-grid, and we set the deviation from the path to be very small. We obtained the training set from the simulator and then used it to train each of the three models. Each model calculated the probabilities associated with it according to the given training set.

We tested the models on paths matching the training set and compared their results. Figure 5(b) shows the comparison between the three models when given observation sequences which are consistent with the simple training set over time. First, these results clearly show that the probability over time in the case of the HMM decreases exponentially—even when the data is consistent with the learned model. This is due to the exponential nature of

HMM probabilities, as discussed in preceding sections. Second, these results show that in a group carrying out a simple stationary path with no variation, HSMM and NHSMM perform equally well and give very high results. This can be easily explained by the explicit state duration property in both models and also by the fact that the data tested in this experiment is very simple.

In the next experiment, we compare the three models for their ability to manage data that is still stationary but is more complex than that of an entire group following the same single path. For this and the rest of the experiments, we used a 64-cell tessellation of the XY-grid with 25×25 size blocks for the simulated data.

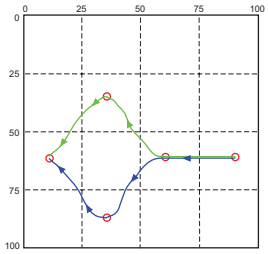
In the second experiment, the training set was composed of two types of behavior groups. At first, the two groups start at the same point and move at the same rate to the next designated coordinate on the XY-grid. At some point, the two groups separate and each group turns to a different direction. After a period of time where each group is moving in a different direction, both groups change directions and eventually meet at the same position on the XY-grid. Figure 6(a) provides a sketch of the two paths.

We trained the three models using the two-behavioral training set over 160 time steps. Figure 6(b) shows the resulting probability values of the models tested on paths matching the two types of groups within the training set. The comparison graph indisputably shows the weakness of HMM versus the other two models, HSMM and NHSMM. Again, the HMM decreases exponentially over time (with a drop in the probabilities, which will be discussed in the following paragraph) while both HSMM and NHSMM remain very high up to a certain point, then the probabilities of all three models drop, and remain constant around the same value till the end of the testing.

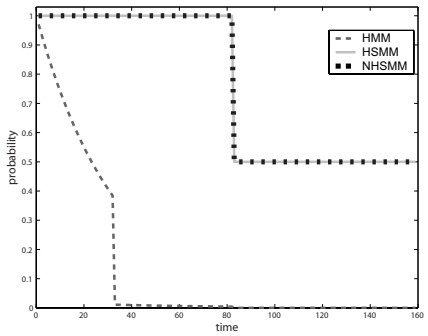
The abrupt drop in likelihood (Figure 6(b)) has different explanations, for the different models. For the graph of HMM, the falling point is at time step 38, where the two type of paths are still moving together in the grid. This drop is due to a long remaining period in a certain state. In this graph, the exact behavior of the HMM graph after this drop is not visible due to the Y-axis scaling range (driven by the probability values in the graphs of the HSMM and NHSMM, which are significantly higher). However, closely observing the results of HMM alone shows additional falling points in the graph. These drops also occur due to long duration within states and due to a decrease in transition probabilities caused by the splitting of the two paths.

The abrupt drop in the graphs of both HSMM and NHSMM is at time step 86, which, according to the two type paths composing the training set, is the point where the two paths split to different directions. This split results in a drop in the probability values for paths matching the training set from a probability of almost 1 to almost 0.5 at the time the splitting occurs. From this point on, the two graphs remain constant around 0.5. Since this was a case of stationary data, there was no effect of time on the transition to the next state. This explains why the two models performed equally well.

In the third test we conducted, we tested an additional type of simulated data that is of a non-stationary nature. The goal

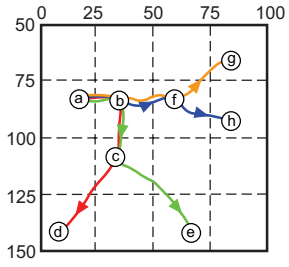


(a) A sketch of the two paths comprising the training set group in the second experiment. The circles represent the states in which the agents lingered (both paths lingered about the same time, i.e., state durations were similar). Movement is right-to-left.

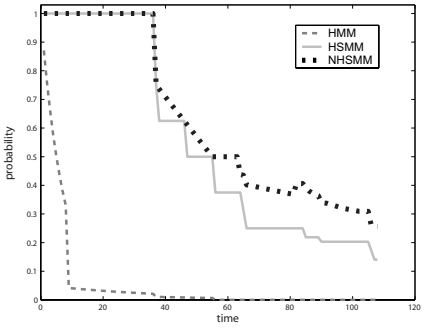


(b) A graph showing the resulting probability values for each models from testing both types of paths of agents moving on the grid according to the paths described over a period. Results are averaged over 30-some different trials.

Figure 6: Second experiment: Settings and Results.



(a) A sketch of the four paths comprising the training set for this experiment. Movement is left-to-right.



(b) A graph showing the resulting probability values for each model from testing the four types of paths of agents moving on the grid according to the paths described over a period. Results are averaged over 30-some different trials.

Figure 7: Third experiment: Settings and Results.

of this experiment was to test the behaviors of HSMM and NHSMM in this type of data and compare their ability to estimate sequences of observation which are compatible with the non-stationary group of paths provided to the models as a training set. In this experiment, the non-stationary training set was devised of four types of behaviors on the grid. Each type differs from the other by its motion rate and by the designated positions along its path.

Figure 7(a) shows a sketch of the types of the four paths. All four paths start at the same position, denoted *a*, and move together on the grid. When they all reach a certain point (coordinate $(60, 60)$ on the XY-grid, marked *b*), where the four paths separate into two different directions, based on the amount of time spent in the state encompassing *b*. The two slower paths keep moving straight towards point *f* after having stayed a longer period of time in the area of *b*, and the two faster paths turn to the right, towards *c*. Further along, another split occurs between each of the two pairs of paths. In *c*, the trajectory turning right (towards *d*) remained a shorter period of time than the trajectory turning left (towards *e*). Similarly, in *f*, the trajectory turning right towards *h* remained in *f* less time than the other trajectory (heading towards *g*).

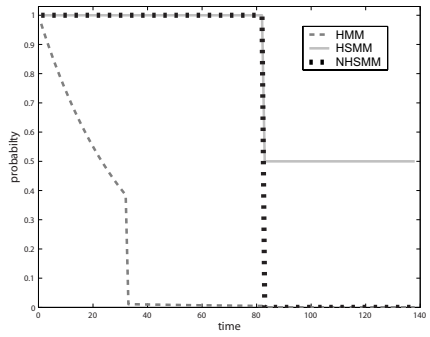
Figure 7(b) shows the results of comparing the ability of the three models to estimate the probability of observation sequences that are compatible with all four types of behaviors comprising the non-stationary data introduced above. Again, we see an exponential decline in the graph of the HMM. For the HSMM and NHSMM, the graph first shows very high probabilities for both models. These are the high probabilities received in the first phase of the training set, where all four paths are moving together on the grid. At time step 42 there is a clear drop in both graphs and further along the graphs appear to gradually decline with a few more noticed drops. These drops in the probabilities of the tested observation sequences are explained as before, by the splitting of the paths causing a descent in the probability of transitioning to each state along each path.

However, from time step 42 where the first split occurs, the graphs of HSMM and NHSMM no longer converge. The NHSMM graph surpasses that of the HSMM throughout the time and significantly outperforms it (paired t-test, p-value is $2.0909 \cdot 10^{-21}$ i.e., lesser than 0.001). This can be explained by the difference in the transition probability functions of the two models. While the HSMM has a constant transition probability, the NHSMM holds an explicit transition probability distribution from state to state allowing the NHSMM to model the non-stationary behavior of the data. These results show the ability of the NHSMM to capture the non-stationary nature of the given data and its superiority over HSMM in cases where the time spent in a certain state has implications on where to proceed to in the tested environment

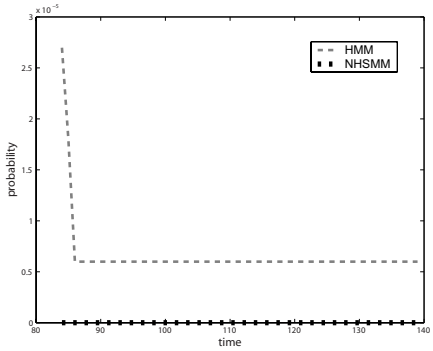
In order to better highlight the difference between HSMM and NHSMMs, we compared their ability to detect non-stationary abnormal behavior. Our goal was to examine, in this forth experiment, how well do the models detect an observation sequence which is not compatible with the training set in terms of duration, but is compatible in terms of the states in the models that it visits.

The training set for this experiment was comprised of two types of behaviors over a period of 140 time steps. The two paths are very similar to those of the second experiment. Both types of behaviors start at the same point in the XY-grid and continue to the next state, where they spend a different amount of time and each turns to opposite directions. In other words, the path spending a short period of time at the splitting state turns right and the path which lingers at the splitting state, turns left. Thus, non-stationary abnormal behavior in this experiment is an observation sequence that remains a short number of time steps at the splitting state but turns left (instead of right).

Figure 8(a) shows the likelihood probability graphs of the non-stationary abnormal behavior for each of the three models. When comparing the resulting graphs of HMM and HSMM in this experiment (see figure 8(a)) with the corresponding resulting graphs in the second experiment (see figure 6(b)), it becomes very clear that these two models do



(a) A graph showing the resulting probability values for each model for their ability to detect non-stationary abnormal behavior.



(b) A closer look at the low values of the HMM and NHSMM graphs in the abnormal behavior detection experiment.

Figure 8: Fourth experiment: Settings and Results. Results are averaged over 30-some different trials.

not detect the non-stationarity of the abnormal behavior and refer to this observation sequence as matching the training set. This is not the case in NHSMM. The NHSMM graph in figure 8(b) clearly shows an abrupt fall at time step 84. This is due to the unexpected turn to the wrong direction. While the transition probability in NHSMM predicted a turn to the right after spending a short time in the splitting state, the path followed a turn to the left, which resulted in very low transition probability in the explicit state transition probability function.

Figure 8(b) shows, in greater detail, the graphs of HMM and NHSMM from time step 84 till 140. The low probability of the HMM graph is due to its exponential probabilities, and this graph is similar to the HMM graph of probability likelihood for an observation matching the training set, whereas the low probabilities of the NHSMM graph are the result of the model's ability to detect abnormal behavior in which the transition from one state to another does not match the duration probability in the prior state.

A comparison between CPU times for calculating the likelihood function over time by each model is presented in the graph in Figure 9. As shown in the graph, there is a substantial increase in computation time when comparing HMM with both HSMM and NHSMM. However, between HSMM and NHSMM, the additional computation in NHSMM appears minor.

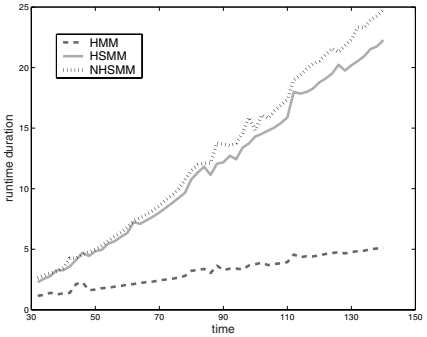


Figure 9: A graph showing the runtime of each model (in seconds) over time

Conclusions

The use of the family of Markov models in activity recognition is fast gaining acceptance. Markov models provide relatively efficient and familiar methods for recognition. However, up until now, the use of such models in activity recognition has been hindered by lack of treatment of the effect of state durations on transition probabilities.

This paper investigates the novel use of Non-stationary Hidden Semi-Markov Model in activity recognition. We discuss the model and its associated algorithms, and evaluate it in comparison with standard HMMs and with Hidden Semi-Markov Models in recognition of normal and abnormal activities. The evaluation reveals the model's strengths in more realistic recognition tasks.

References

Bui, H. H.; Venkatesh, S.; and West, G. 2002. Policy recognition in the abstract hidden Markov model. *Journal of Artificial Intelligence Research* 17:451–499.

Charniak, E., and Goldman, R. 1993. A Bayesian model of plan recognition. *Artificial Intelligence Journal* (64):"53–79".

Duong, T. V.; Bui, H. H.; Phung, D. Q.; and Venkatesh, S. 2005. Activity recognition and abnormality detection with the switching hidden semi-Markov model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*.

Ferguson, J. 1980. Variable duration models for speech. In *Symposium on the Application of Hidden Markov Models to Text and Speech*, 143–179.

Fox, D.; Hightower, J.; Liao, L.; Schulz, D.; and Borriello, G. 2003. Bayesian filtering for location estimation. *IEEE Pervasive Computing*.

Hongeng, S., and Nevatia, R. 2003. Large-scale event detection using semi-HMMs. In *9th IEEE International Conference on Computer Vision (ICCV'03)*.

Kaminka, G. A.; Pynadath, D. V.; and Tambe, M. 2002. Monitoring teams by overhearing: A multi-agent plan recognition approach. *Journal of Artificial Intelligence Research* 17.

Levinson, S. 1986. Continuously variable duration hidden Markov models for automatic speech recognition. *Computation Speech Language* (45):1–29.

Oliver, N.; Horvitz, E.; and Garg, A. 2002. Layered representation for human activity recognition. In *Fourth IEEE International Conference on Multimodal Interfaces (ICMI02)*, 3–8.

Pynadath, D., and Wellman, M. 2000. Probabilistic state-dependent grammars for plan recognition. In *the Conference on Uncertainty in Artificial Intelligence*, 507–514.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In *IEEE*, volume 2, 257–286.

Sin, B., and Kim, J. 1995. Nonstationary hidden Markov models. *Signal Processing* 46:31–46.

Vaseghi, S. V. 1995. State duration modelling in hidden Markov models. *Signal Processing* 41:31–41.

Yin, J.; Chai, X.; and Yang, Q. 2004. High-level goal recognition in a wireless LAN. In *American Association for Artificial Intelligence*.

Yu, S.-Z., and Kobayashi, H. 2003. A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing* 83(2):235–250.