Improving the Recognition of Interleaved Activities

Joseph Modayil
modayil@cs.rochester.edu

Tongxin Bai
bai@cs.rochester.edu

Henry Kautz
kautz@cs.rochester.edu

Department of Computer Science
University of Rochester
Rochester, NY 14627

ABSTRACT

We introduce Interleaved Hidden Markov Models for recognizing multitasked activities. The model captures both inter-activity and intra-activity dynamics. Although the state space is intractably large, we describe an approximation that is both effective and efficient. This method significantly reduces the error rate when compared with previously proposed methods. The algorithm is suitable for mobile platforms where computational resources may be limited.

ACM Classification Keywords

I.5.1 [Pattern Recognition]: Models—Statistical, J.3 [Life and Medical Sciences]:Health

Author Keywords

Hidden Markov Model, Activity Recognition, Interleaved Activities

INTRODUCTION

With the availability of cheaper and more ubiquitous sensors, mobile devices are able to continuously observe the way a person interacts with the physical environment. Using these observations, a device has the potential to accurately recognize a person's activities. This in turn supports many emerging applications, for example, assisting individuals with cognitive impairments [11]. This paper presents a method to significantly improve recognition of *interleaved* activities of daily living (ADLs).

Our goal is to recognize activities from observation sequences of object identifiers following the "invisible human" paradigm introduced by Philipose *et al.* [10]. The observations come from a wrist-worn Radio Frequency Identification (RFID) reader and tags [4]. As the user goes through activities of daily living, the reader records which objects were used. A list of activities to be recognized may come from a healthcare provider. The task is to learn a model that recognizes which activities are being performed given a sequence of these observations.

UbiComp'08, September 21-24, 2008, Seoul, Korea.

People often multitask as they perform ADLs, switching frequently between steps of different activities. In order to more accurately model and recognize such behavior, we introduce the interleaved hidden Markov model, which augments the hidden state by recording the last observation for each activity. Although the full state space is intractably large, we show that using approximate inference with this model is both effective and efficient for activity recognition.

The subsequent sections describe the related work, provide the formalism and present the experimental results.

RELATED WORK

We consider the problem of activity recognition for data in which multiple activities are interleaved, which is a more general formulation than approaches that only classify presegmented sequences of activity. Although in other models of concurrent activities a single observation can be explained by multiple activities [14], this is not common when the observation sequence is coming from objects that are near a user's hand.

There have been many advances in activity modeling. Early work [5] provided logical underpinnings to plan recognition but was not grounded in physical observations. Techniques for activity recognition from sensor data include fixed-length feature-vector classifiers [6], dynamic Bayes Nets [3], and stochastic grammars [7]. Previous research [10] on recognizing activities of daily living with RFID has shown that isolated activities can be recognized with a hidden Markov model with a limited number of states.

Difficulties can arise when a user is multitasking between multiple activities in natural environments. Patterson and colleagues [9] collected data from a variety of morning activities using RFID tags and readers. They compared multiple activity recognizers and found that an HMM with one state per activity performed well, but increasing model complexity did not improve the recognition performance. Closely related work to what we are proposing is presented by Duong and colleagues [1]. They recognize several activities of daily living using a hierarchical hidden semi-Markov model. Their algorithm can perform accurate recognition for some ADLs using observations from cameras that track a user's location.

PROBLEM FORMULATION

Given a sequence of observations, the activity recognition task is to associate one activity with each observation. One model proposed previously [9] for this task is to use an HMM

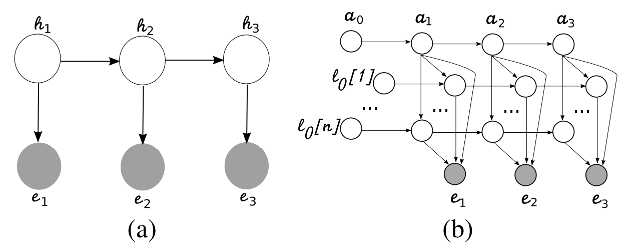

Figure 1. Graphical model representation of the initial timesteps for both the (a) HMM and the (b) IHMM. The shaded nodes are observations and the unshaded nodes form the hypothesis space.

with one state per activity. More complex models face two computational difficulties. First, training is difficult for models with a large number of parameters. Second, models with large state spaces can become intractable for exact inference algorithms.

We have discovered an effective approximation algorithm for activity recognition. The implementation requires only a small modification of the HMM described above, and remains computationally efficient.

Interleaved HMM Formalism

Denote the activities by A and the objects by O.

$$A = \{1, \ldots, n\} \qquad O = \{o_1, \ldots, o_m\}$$

The evidence stream $E = \{e_1, \ldots, e_T\}$ consists of one object reading per time step $e_t \in O$. For a given hypothesis space of states H, a hidden Markov model [12] is given by (π, D, E) consisting of the starting state distribution $\pi = P(h_1)$, the transition probabilities $D_{ij} = P(h_{t+1} = j | h_t = i)$, and the emission probabilities $E_j(o_k) = P(e_t = o_k | h_t = j)$. Following [12], the probability of the most likely state sequence given the evidence is computed using the following recursive formula.

$$\begin{aligned}\delta_{t+1}(j) &= \max_{h_1,\ldots,h_t} P(h_1, \ldots, h_{t+1} = j, e_1, \ldots e_{t+1}) \\ &= \max_{h_1,\ldots,h_{t-1}} P(e_{t+1} | h_{t+1} = j) \times \\ &\qquad \max_{i \in H} P(h_{t+1} = j | h_t = i) \times \\ &\qquad\qquad P(h_1, \ldots, h_t = i, e_1, \ldots e_t) \\ &= E_j(e_{t+1}) \max_{i \in H} D_{ij} \delta_t(i)\end{aligned} \tag{1}$$

For the hypothesis spaces considered in this paper, there is a natural projection $H \to A$. This projection is used to infer the activities from the hidden state sequence.

The first HMM (Model I) has one state per activity.

$$H_1 = A$$

This model has $|A|^2$ parameters for transition probabilities and $|A||O|$ parameters for emission probabilities.

We now define the interleaved HMM model, where each state consists of a current activity and a record of the last object observed while performing each activity. The state space is

$$H_2 = A \times L,$$

where $L = O^{|A|}$ is a Cartesian product of $|A|$ copies of O, which is shown as a graphical model in Figure 1. The hypothesis at time t is given by $h_t = <a_t, l_t>$. We denote an element $l \in L$ by $(l[1], \ldots, l[n])$ where $l[i]$ indicates the last object observed in activity i. The emission probabilities are deterministic,

$$P(e|\, <a, l>) = id(e, l[a]),$$

where we use the identity function $id(x, y)$, which is one if $x = y$ and zero if $x \neq y$. The transition probability from $<a, l>$ to $<a', l'>$ is given by

$$P(a', l' | a, l) = P(a'|a) P(l'[a'] | l[a'], a') \prod_{i \neq a'} id(l'[i], l[i]) \tag{2}$$

This equation provides the constraint that only one item in the record l changes between time steps. The first factor above is the transition probability between activities ($|A|^2$ parameters). The second factor is the probability of the next object to be observed in the new activity, given the last object observed in the new activity ($|O|^2|A|$ parameters). The third factor prohibits changes in the record of the last object observed except for the new activity.

The number of free parameters decreases slightly as probabilities must sum to one, and increases slightly as a "None" observation is added to O for initialization. Hence, the number of parameters is constrained to approximately $|A||O|^2 + |A|^2$, but the size of the state space ($|A||O|^{|A|}$) prohibits exact computation.

Since the state space H_2 can not be explored completely at each time step, we use a beam search [8] to define a likelihood update equation over a beam $B \subset H_2$ in the search space. Equation 1 is modified in the following set of recursive equations to perform a maximization over the beam instead of the entire hypothesis space. In the following equations, B_t is the beam at time t, which is defined at time zero to simplify the update equations.

$$\begin{aligned}B_0 &= \{<a, (None, \ldots, None)> \ | \ a \in A\} \\ \delta_{t+1}(q') &= \max_{h_1,\ldots,h_t} P(h_1, \ldots, h_{t+1} = q', e_1, \ldots e_{t+1}) \\ &\approx \max_{h_1,\ldots,h_{t-1}} P(e_{t+1} | h_{t+1} = q') \times \\ &\qquad \max_{q \in B_t} P(h_{t+1} = q' | h_t = q) \times \\ &\qquad\qquad P(h_1, \ldots, h_t = q, e_1, \ldots e_t) \\ l_{t+1}^a &= \arg\max_{l \in L} \delta_{t+1}(<a, l>) \\ B_{t+1} &= \{<a, l_{t+1}^a> \ | \ a \in A\}\end{aligned} \tag{3}$$

In spite of the more complex equations, the interleaved HMM algorithm is implemented efficiently as a variation of model I. The IHMM maintains a beam B_t with a single hypothesis in L for each possible current activity in A, namely $\forall a \in A, \exists!(a, l) \in B$. Hence iteration over L in the above equations is performed over A. The record l_{t+1}^a is the most likely record for activity a at time $t + 1$.

We call our variant of beam search an interleaved hidden Markov model or IHMM (model II). The interleaved HMM is a variant of switching HMMs [2] as shown in Figure 1. In this interpretation, the component a is the state variable of an HMM that is used to switch between the output of $|A|$ activity HMMs, one HMM for each component of l. The transitions are constrained so that the only activity HMM permitted to change its state is the one selected by the switch.

Defining the best state sequence

For a given HMM, there are multiple ways to define the best state sequence for the evidence. The most popular is the most likely activity sequence given all the evidence (the Viterbi path).

$$Viterbi = \arg\max_{h_1,...,h_T} P(h_1, ..., h_T, e_1, \ldots, e_T),$$

The Viterbi path is efficiently computed using the Viterbi algorithm. We can also estimate the current state for the most likely path from the evidence seen up to time t.

$$Filter(t) = \arg\max_{<a,l>} \delta_t(<a, l>)$$

The filtering option is desirable when the inferred states are required in real time.

EXPERIMENTS

We simulated data for a simple scenario to demonstrate how the IHMM improves activity recognition. There are three activities (drinking a glass of water, making a stir-fry and making jello) and six tagged objects: a drinking glass, a jello packet, a stir-fry packet, a spoon, a serving bowl, and a dessert dish. In each session, the user first makes a stir-fry (sequentially using the stir-fry packet, the spoon, and the serving bowl) and then makes jello (using the jello packet, the spoon, and the dessert dish). Both activities have 5 minutes of stirring with the spoon, and the stirring is interrupted every minute when the user takes a drink of water for 5 seconds. After each drinking interruption, the user resumes stirring. At this point, the standard HMM performs at chance for discriminating between making jello and making a stir-fry because both activities are equally likely to have a period of stirring. The interleaved HMM performs better, because the IHMM state represents which activity was previously using the spoon. In this simulated scenario, the accuracy of the HMM is 66% while the IHMM acheives 100% accuracy.

The IHMM also provides improved recognition accuracy in real data. The data set collected by Patterson *et al.* [9] consists of 11 interleaved morning activities involving 43 object classes. The mean duration of each run was 27 minutes and the mean time between switching activities was 74 seconds. The activities are shown in Table 1, along with some representative objects for each activity. They evaluated Model I along with more complex models, but there was no significant difference in accuracy between Model I and the more complex models. We measured different performance numbers than those reported in their paper since our time steps correspond to RFID events instead of physical time.

The activity recognizers were trained and tested on this data. As labels are available for all the training data, probabilities are computed by event counting. All probability tables are initialized with a small uniform probability which provides a Dirichlet prior for the multinomial distribution.

1	Clear Table	cups, spoons
2	Eat Breakfast	plates, spoons, mug
3	Front Door	door
4	Make Espresso	espresso steam knob, mug, fridge
5	Make Juice	juice pitcher, juice, faucet
6	Make Oatmeal	oatmeal, saucepan, stove control
7	Make Eggs	fridge, saucepan
8	Make Tea	stove control, kettle, faucet
9	Set Table	cupboard, mug
10	Use Bathroom	toilet lid, toilet flush handle
11	Use Phone	phone

Table 1. Breakfast activities and some of the related objects

Model	Algorithm	Error Rate
I	HMM (Viterbi)	$4.5\%(\sigma = 1.21)$
II	IHMM (Viterbi)	$1.2\%(\sigma = 0.71)$
III	IHMM (Filtering)	$2.9\%(\sigma = 0.51)$
IV	Instance	$13.4\%(\sigma = 1.95)$

Table 2. A comparison of activity recognition algorithms on the breakfast data set, where the accuracies are reported for the most likely sequence. The data set is quite large, consisting of over 23,000 object readings from 11 activities collected over 10 days. The accuracy and standard deviation are computed from leave-one out cross validation. The interleaved HMM outperforms the rest.

The results for activity recognition are shown in Table 2. The interleaved HMM outperforms the other recognizers. The accuracy of filtering with the IHMM is also slightly better than Model I. For comparison to the sequence based methods, we include an instance based classifier, which classifies each object to its most probable activity, independent of the sequence information. This model performs considerably worse than Model I.

To explain the improvement provided by the IHMM, we measure the quality of the search space approximation by how often the record variable contains the correct last object used on activity transitions. When the activity is unchanged across time steps, the record for the activity is correct by construction. Given the *true* sequence of activities and observations, $\{(a_t, o_t)\}$, for the time t and activity i we define $\bar{l}(i, t)$ to be the last object observed in activity i; formally $\bar{l}(i, t) = o_q$ where $q = \max\{s : 1 \leq s \leq t \wedge a_s = i\}$. Let T' denote the time points of activity transitions, namely $T' = \{t \in [1, \ldots, T] | a_t \neq a_{t+1}\}$.

Let $l(a, t, b)$ denote the last object used for activity a in the record associated to activity b in the beam at time t. We define the accuracy to be the percent of time that the true history matches the record in the beam.

$$acc = \frac{1}{|T'|} \sum_{t \in T'} id(l(a_{t+1}, t, a_t), \bar{l}(a_{t+1}, t))$$

For this dataset, the record correctly contained the last object observed in the activity for 87% ($\sigma = 8$) of the inter-activity transitions. We conclude that the approximation in IHMM to the full state space search is very effective. While other approximation methods (loopy belief propagation, variational methods) might provide some improvement, the beam search is providing most of the benefit of this extended search space for little added complexity.

DISCUSSION

The IHMM provides a simple yet effective way to improve activity recognition on real data by recording the last object observed for each activity. Despite uncertainty in real sensor data (when irrelevant objects are observed), the probabilistic nature of the model allows graceful recovery. For situations with several potentially confusable activities, the IHMM is an efficient mechanism for focusing computational resources on plausible activity sequences.

While this research presents promising results for accurately recognizing interleaved activities, there are limitations. Data gathered as part of the House_n project [6] has shown some of the limitations of RFID in extended use environments, for example, the tags must be within the range of the reader. However, the interleaved HMM can be applied to other sensor modalities such as motes or vision to provide coverage when RFID is ineffective or undesirable [13].

Another objection to the methodology could be the perceived difficulty of gathering labeled data for training. This may not be a significant difficulty, as even with only one day of training data, the model performs nearly as well as with nine days of training data (3.7% error).

The work in [1] also modeled the activities of daily living, but they use a significantly different style of observations (video tracking of a person's location in a room). Their approach also used a switching model, but their approach restarts activities instead of remembering the last state within an activity model. However, they used hidden semi-Markov models instead of HMMs in order to better model the duration of activities. Combining the techniques might yield even better activity recognition.

CONCLUSIONS

We have described how the interleaved HMM can be used for activity recognition. The model is able to predict transition probabilities better by recording the last object observed in each activity. This approach requires an approximation for the inference, but it is able to achieve very low error rates. While there are still technical challenges to be overcome in reliable feature detection, this work demonstrates that activity recognition using RFID can be very accurate.

We thank the reviewers for their constructive comments. This work is supported by NSF award ISS-0734843, NYSTAR award C020010, and DARPA SBIR award 07SB2-0196.

REFERENCES

1. T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. In *Proc. 2005 IEEE Computer Vision and Pattern Recognition (CVPR-05)*, pages 838–845, 2005.
2. Z. Ghahramani and G. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12(4):831–864, 2000.
3. E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence*, pages 256–265, 1998.
4. A. R. K. Fishkin, M. Philipose. Hands-on RFID: Wireless wearables for detecting use of objects. In *International Symposium on Wearable Computers (ISWC'05)*, 2005.
5. H. Kautz and J. Allen. Generalized plan recognition. In *Proc. 5th National Conf. on Artificial Intelligence (AAAI-86)*, pages 32–37, 1986.
6. B. Logan, J. Healey, M. Philipose, E. Tapia, and S. Intille. A long-term evaluation of sensing modalities for activity recognition. In *UbiComp 2007: Ubiquitous Computing*, pages 483–500, 2007.
7. D. Moore and I. Essa. Recognizing multitasked activities from video using stochastic context-free grammar. In *Proc. 18th National Conf. on Artificial Intelligence (AAAI-2002)*, 2002.
8. H. Ney, D. Mergel, A. Noll, and A. Paeseler. A data-driven organization of the dynamic programming beam search for continuous speech recognition. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'87)*, pages 833–836, 1987.
9. D. Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *Proceedings of the IEEE International Symposium on Wearable Computers*, Osaka, Japan, Oct 2005.
10. M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Hahnel, D. Fox, and H. Kautz. Inferring activities from interactions with objects. *IEEE Pervasive Computing: Mobile and Ubiquitous Systems*, 3(4):50–57, 2004.
11. M. Pollack, L. Brown, D. Colbry, C. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, and I. Tsamardinos. Autominder: An intelligent cognitive orthotic system for people with memory impairment. *Robotics and Autonomous Systems*, 44:273–282, 2003.
12. L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.
13. J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. Rehg. A scalable approach to activity recognition based on object use. In *IEEE Int. Conf. on Computer Vision (ICCV 2007)*, pages 1–8, 2007.
14. T. Wu, C. Lian, and J. Hsu. Joint recognition of multiple concurrent activities using factorial conditional random fields. In *Proc. 22nd Conf. on Artificial Intelligence (AAAI-2007)*, 2007.