

Published in final edited form as:

IEEE Pervasive Comput. 2010 ; 9(1): 48. doi:10.1109/MPRV.2010.7.

Human Activity Recognition and Pattern Discovery

Eunju Kim, Sumi Helal, and Diane Cook

Activity recognition is an important technology in pervasive computing because it can be applied to many real-life, human-centric problems such as eldercare and healthcare. Successful research has so far focused on recognizing simple human activities. Recognizing complex activities remains a challenging and active area of research. Specifically, the nature of human activities poses the following challenges:

- Recognizing concurrent activities

People can do several activities at the same time [3]. For example, people can watch television while talking to their friends. These behaviors should be recognized using a different approach from that for sequential activity.

- Recognizing interleaved activities

Certain real life activities may be interleaved [3]. For instance, while cooking, if there is a call from a friend, people pause cooking for a while and after talking to their friend, they come back to the kitchen and continue to cook.

- Ambiguity of interpretation

Depending on the situation, the interpretation of similar activities may be different. For example, an activity “open refrigerator” can belong to several activities, such as “cooking” or “cleaning”.

- Multiple residents

In many environments more than one resident is present. The activities that are being performed by the residents in parallel need to be recognized, even if the activity is performed together by the residents in a group.

Human activity understanding encompasses activity recognition and activity pattern discovery. The first focuses on accurate detection of the human activities based on a predefined activity model. Therefore, an activity recognition researcher builds a high-level conceptual model first, and then implements the model by building a suitable pervasive system. On the other hand, activity pattern discovery is more about finding some unknown patterns directly from low-level sensor data without any predefined models or assumptions. Hence, the researcher of activity pattern discovery builds a pervasive system first and then analyzes the sensor data to discover activity patterns. Even though the two techniques are different, they both aim at improving human activity technology. Additionally, they are complementary to each other - the discovered activity pattern can be used to define the activities that will be recognized and tracked.

Activity recognition

The goal of activity recognition is to recognize common human activities in real life settings. Accurate activity recognition is challenging because human activity is complex and highly diverse. Several probability-based algorithms have been used to build activity models. The Hidden Markov Model and the Conditional Random Field are among the most popular modeling techniques. We describe these two techniques in the context of an *eating* activity example.

The Hidden Markov Model (HMM)

Simple activities can be modeled accurately as Markov Chains. However, complex or unfamiliar activities are often difficult to understand and model. For example, a researcher studying activities of daily living for a person with dementia will have a difficult time fitting a model unless she is an expert in dementia and understands its related behavioral science. Fortunately, observing signals stemming from complex or unfamiliar activities can be utilized to indirectly build a model of the activity. Such a model is called a Hidden Markov Model or HMM. By observing the effects of an activity, HMM is able to gradually construct the activity model, which can be further tuned, extended and reused in similar studies.

HMM is a generative probabilistic model, which is a model that is used for generating hidden states from observable data [2]. Specifically, a main goal of this model is to determine the hidden state sequence $(y_1 y_2 \dots y_t)$ that corresponds to the observed output sequence (x_1, x_2, \dots, x_t) . Another important goal is to learn model parameters reliably from the history of observed output sequences. Figure 1 shows a graphical representation of an HMM which is composed of 5 hidden states and 4 observable variables.

HMM requires two independence assumptions for tractable inference:

- The 1st order Markov assumption of transition

$$P(y_t | y_1, y_2, y_3, \dots, y_{t-1}) = P(y_t | y_{t-1})$$

The future state depends only on the current state, not on past states [2]. That is, the hidden variable at time t , y_t , depends only on the previous hidden variable y_{t-1} .

- Conditional independence of observation parameters

$$P(x_t | y_t, x_1, x_2, \dots, x_{t-1}, y_1, y_2, y_3, \dots, y_{t-1}) = P(x_t | y_t)$$

The observable variable at time t , x_t , depends only on the current hidden state y_t . In other words, the probability of observing x while in hidden state y is independent of all other observable variables and past states [2].

To find the most probable hidden state sequence from an observed output sequence, HMM finds a state sequence which maximizes a joint probability $p(x, y)$ of the transition probability $p(y_{t-1} | y_t)$ and the observation probability $p(x_t | y_t)$ (that is the probability that outcome x_t is observed in state y_t [2]).

$$p(x, y) = \prod_{t=1}^T p(y_t | y_{t-1}) p(x_t | y_t)$$

When an HMM is used in activity recognition, activities are the hidden states and observable output is sensor data. Figure 2 shows an HMM for the example *eating* activity.

To increase the accuracy of an HMM, training is possible with prior knowledge of some aspects of the model [2]. Also, individually trained HMM can be combined to construct a larger HMM model (e.g., of a complex activities with clear sub-activities structure). Training is sometimes necessary to “induce” all possible observation sequences that are needed to find the $p(x, y)$ joint distribution of the HMM.

The Conditional Random Field (CRF)

Even though simple and popular, HMMs have serious limitations, most notably its difficulty in representing multiple interacting activities (concurrent or interwoven) [3]. An HMM is also incapable of capturing long-range or transitive dependencies of the observations due to its very strict independence assumptions (on the observations). Furthermore, without significant training, an HMM may not be able to recognize all of the possible observation sequences that can be consistent with a particular activity.

In practice, many activities may have non-deterministic natures, where some steps of the activities may be done in any order. In practice, also, many activities are concurrent or interwoven. A conditional random field (CRF) is a more flexible alternative to the HMM that addresses such practical requirements. It is a discriminative and generative probabilistic model for the dependence of a hidden variable y on an observed variable x [2]. Both HMMs and CRFs are used to find a hidden state transition from observation sequences. However, instead of finding a joint probability distribution $p(x,y)$ as the HMM does, a CRF attempts to find only the conditional probability $p(y|x)$. A CRF allows for arbitrary, non-independent relationships among the observation sequences, hence the added flexibility. Another major difference is the relaxation of the independence assumptions, in which the hidden state probabilities may depend on the past and even future observations [2]. A CRF is modeled as an undirected acyclic graph, flexibly capturing any relation between an observation variable and a hidden state [2]. Figure 3 shows a CRF equivalent of the eating activity HMM model shown in Figure 2.

A CRF uses a potential function instead of a joint probability function [2]. Suppose there are hidden variables $Y=(y_1, y_2, \dots, y_{t-1}, y_t)$ and observation variables $X=(x_1, x_2, \dots, x_k)$. The two probabilities, transition and observation, of the HMM are replaced by what is called a transition feature function $r(y_{t-1}, y_t)$ and the state feature function $s(y_t, X)$, respectively [2]. Both feature functions return 1 in the simplest case if there is a correlation between its variables. The potential function $p(Y|X)$ is computed by the following equation [2].

$$p(Y|X) = \frac{1}{Z(X)} \exp \left(\sum_i \lambda_i \sum_{t=1}^n f_i(y_{t-1}, y_t, X, t) \right)$$

where, $f_i(y_{t-1}, y_t, X, t) = r(y_{t-1}, y_t)$ or $s(y_t, X)$

and where λ_i is a weight of correlation which represents the actual potential. In the eating example, the λ_i values are shown in Figure 3-a, which are values estimated from training data. $Z(X)$ is a normalization factor to convert a potential value to a probability value between 0 and 1 [2].

The Skip Chain Conditional Random Field (SCCRF)

The above potential function is very simple and leads to what is known as a Linear-Chain CRF. While Linear-Chain CRFs are more flexible than HMMs (due to relaxed independence assumptions), they both share the limitation of being suitable only for activities that are sequential in nature. To model complex activities with concurrent and interleaved sub-activities, more sophisticated potential functions are needed to capture long-range (skip chain) dependencies. In this case, the CRF is called a Skip Chain CRF or SCCRF. Given that a Skip Chain is essentially a Linear Chain with a larger distance between two variables, it is easy to define SCCRF potential functions as products of multiple Linear Chains [2].

Such potential products are expensive computationally, especially for high numbers of Skip Chains between activities [2].

Emerging Patterns (EP)

An emerging pattern is a feature vector of each activity that describes significant changes between two classes of data [3]. Consider a dataset D that consists of several instances, each having a set of attributes and their corresponding values. Among the available attributes, there are some which support a class more than the others. For instance, a feature vector {location@kitchen, object@burner} is an EP of a *cooking* activity and {object@cleanser, object@plate, location@kitchen} is an EP of a *Cleaning a dining table activity* [3]. To find these attributes, *support* and *GrowthRate* are computed for every attribute X [3].

$$Supp(X) = \frac{\text{the number of instances containing } X \text{ in } D}{\text{the number of instances in } D}$$

Given two different classes of datasets $D1$ and $D2$, the growth rate of an item set X from $D1$ to $D2$ is defined as follows:

$$GrowthRate(X) \begin{cases} 0 & \text{if } Supp_1(X)=0 \text{ and } Supp_2(X)=0 \\ \infty & \text{if } Supp_1(X)=0 \text{ and } Supp_2(X)>0 \\ \frac{Supp_2(X)}{Supp_1(X)} & \text{otherwise} \end{cases}$$

EPs are item sets with large growth rates from $D1$ to $D2$. These EP are mined from sequential sensor data and applied to recognize interleaved and concurrent activities [3].

Other Approaches

The number of machine learning models that have been used for activity recognition varies almost as greatly as the types of activities that have been recognized and types of sensor data that have been used. Naïve Bayes classifiers have been used with promising results for activity recognition [6]. These classifiers identify the activity that corresponds with the greatest probability to the set of (assumed conditionally independent) sensor values that were observed. Other researchers have employed decision trees to learn readable, logical descriptions of the activities [7].

Comparison of activity recognition methods

Table 1 compares the presented activity recognition methods based on three key aspects. The Skip-Chain CRF and Emerging Pattern methods recognize concurrent and interleaved activities. Except for Emerging Patterns, all methods require supervised learning, which in turn require training data for real activity recognition – a major limitation in performing automatic labeling. When the sensor environment is changed, every model is influenced and must also be changed accordingly.

Activity pattern discovery

Complementary to the idea of activity recognition is to automatically recognize activity patterns in an unsupervised fashion. Tracking only pre-selected activities ignores the important insights that other discovered patterns can provide on the habits of the residents and the nature of the environment. In addition, if automatically-discovered activities are

recognized and tracked, data will not need to be pre-labeled and used to training the recognition algorithms.

Topic model based daily routine discovery

One intuitive way to find a daily pattern is to build a hierarchical activity model. The lower-level activities, such as sitting, standing, eating, and driving, are recognized using a supervised learning algorithm. The higher level of the model discovers combinations of the lower-level activities that represent more complex activity patterns.

In the topic model approach to activity pattern discovery [4], activity patterns are recognized in a fashion similar to how topics can be pulled from a document using a bag-of-words approach. A mixture of topics can be modeled as a multinomial probability distribution $p(z|d)$ over topics z . The importance of each word for topic z is also modeled as a probability distribution $p(w|z)$ over words in a vocabulary. The equation below then shows the word distribution that is expected for a set of topics:

$$p(w|d) = \sum_{z=1}^T p(w|z)p(z|d), \text{ d:document, w:word, z:topic}$$

Activity patterns can be similarly discovered and topic-word distributions, where words correspond to recognized activities and daily routines correspond to topic activation. As an example, the process was applied to sensor data collect for one subject. Figure 4 (top) shows the intuitive description of the activities as noted by the subject and Figure 4 (bottom) shows the activities that were automatically discovered with the corresponding $p(w|z)$ values.

Activity data pattern discovery

An alternative approach to activity pattern discovery is to visually observe activities and to extract individual poses from the video data. Activities can then be represented by constructing probabilistic context-free grammars using the poses as the grammar alphabet [5]. Figure 5 shows example poses that correspond to collect video key frame sequences.

Next, specific classes of rules are extracted from the data that represent repeating sequences of poses and complex combinations, as shown in Figure 6. For example, the “kick” action is composed of three pairs of poses, denoted as q1 through q3 in Figure 6 (left). If the “kick” activity is combined with a recognized object such as “football”, the combination corresponds to a more specific activity that can be recognized such as “play football”.

In addition to clustering several primitive activities, they are combined with object and build more specific composite activities. For example, if activity “kick” is combined with an object “football”, it generates a more specific activity like “play football”.

Once activities are discovered, they can provide the basis for a model to recognize the activity, track its occurrence, and even use the information to assess an individual’s well-being or provide activity-aware services. These activity discovery and recognition technologies are thus valuable for providing pervasive assistance in an individual’s everyday environments.

References

1. Bose, R.; Helal, A. Observing Walking Behavior of Humans using Distributed Phenomenon Detection and Tracking Mechanisms. Proceedings of the International Workshop on Practical Applications of Sensor Networks; 2008.

2. Sutton, C.; McCallum, A. An introduction to conditional random fields for relational learning. In: Getoor, L.; Taskar, B., editors. *Introduction to Statistical Relational Learning*. MIT Press; 2006.
3. Gu, T.; Wu, Z.; Tao, X.; Pung, HK.; Lu, J. epSICAR: An Emerging Patterns based approach to sequential, interleaved and Concurrent Activity Recognition. *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*; 2009.
4. Hu`ynh, T.; Fritz, M.; Schiele, B. Discovery of Activity Patterns using Topic Models. *Proceedings of the Tenth International Conference on Ubiquitous Computing*; 2008. p. 10-19.
5. Ogale, A.; Karapurkar, A.; Aloimonos, Y. View-invariant modeling and recognition of human actions using grammars. *Proceedings of the International Workshop on Dynamical Vision*; 2005.
6. Munguia-Tapia E, Intille SS, Larson K. Activity recognition in the home using simple and ubiquitous sensors. *Proceedings of PERVASIVE 2004*:158–175.
7. Maurer, U.; Smailagic, A.; Siewiorek, D.; Deisher, M. Activity recognition and monitoring using multiple sensors on different body positions. *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*; 2006. p. 99-102.

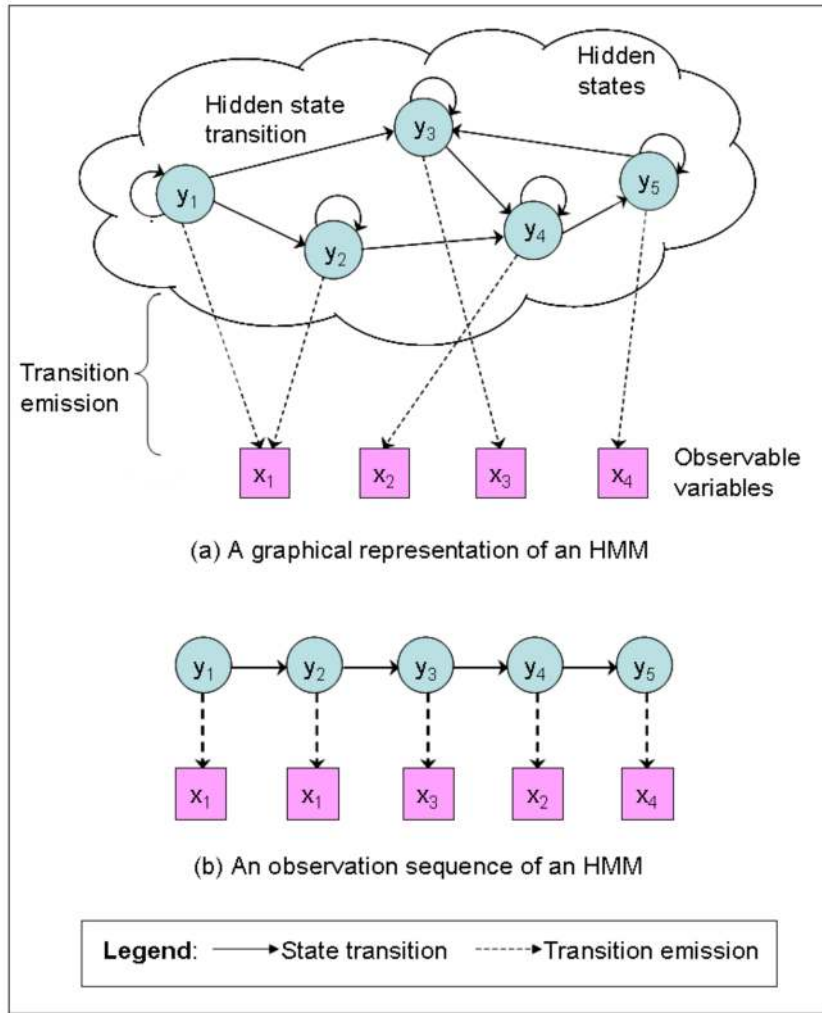


Figure 1.
The graphical representation of a HMM

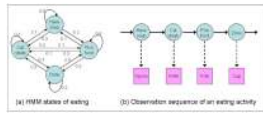


Figure 2.
Example of HMM for eating activity

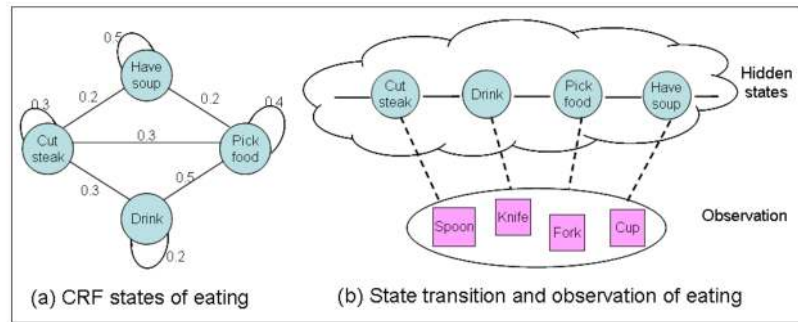


Figure 3.
Example of CRF for eating activity

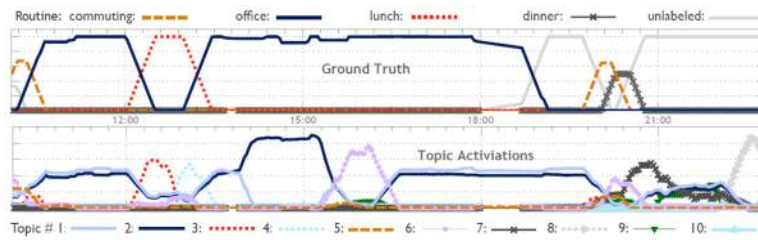


Figure 4.
Discovered topics [4]

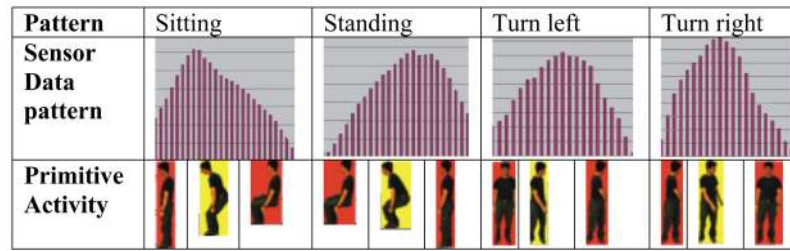


Figure 5.
Extraction of motion patterns from sensor data [5]

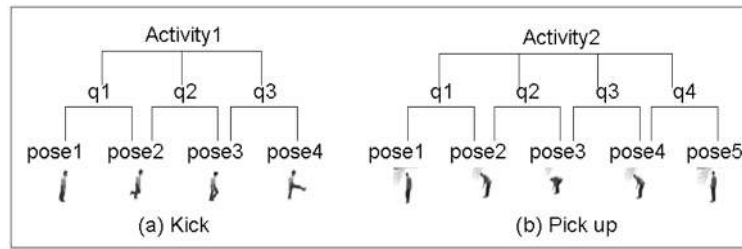


Figure 6.
Composite activities [5]

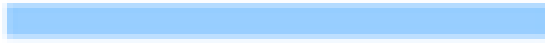









Table 1

Comparison of activity recognition models

	HMM	Linear Chain CRF	Skip Chain CRF	Emerging Pattern
Concurrent & interleaved activity	Not recognized	Not recognized	Recognized	Recognized
Learning method for Labeling	Supervised	Supervised	Supervised	Partially unsupervised
Scalability	Change of HMM graph required	Change of CRF graph required	Change of SCCRF graph required	EP mining required

Table 2

Topics and its activities [4]

Topic	Content Activities of the Ten Topics
1 	sitting/desk activities (1.0)
2 	sitting/desk activities (0.99)
3 	having lunch (0.5), walking freely (0.26), picking up cafeteria food (0.06), queuing in line (0.06), unlabeled (0.04)
4 	standing/using toilet (0.69), walking freely (0.18), queuing in line (0.06)
5 	driving car (0.33), walking & carrying sth (0.21), sitting/desk activities (0.2), walking (0.14), unlabeled (0.08)
6 	using the toilet (0.69), walking freely (0.17), discussing at whiteboard (0.06), sitting/desk activities (0.03)
7 	having dinner (0.77), desk activities (0.1), washing dishes (0.08), unlabeled (0.04)
8 	lying/using computer (1.0)
9 	unlabeled (0.87)
10 	watching a movie (1.0)