Editor: Sumi Helal ■ University of Florida ■ helal@cise.ufl.edu

Human Activity Recognition and Pattern Discovery

Eunju Kim, Sumi Helal, and Diane Cook

In principle, activity recognition can be exploited to great societal benefits, especially in real-life, human-centric applications such as eldercare and healthcare. Successful research, however, 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,[1] such as watching television while talking to friends. These behaviors should be recognized using a different approach from that for sequential activity.
- *Recognizing interleaved activities*. Certain real-life activities can be interleaved.[1] For instance, if a friend calls while you're cooking, you'd talk to your friend for a while, while you continue to cook.
- *Ambiguity of interpretation*. Similar situations can be interpreted differently. For example, an "open refrigerator" can belong to several activities, such as "cooking" or "cleaning."
- *Multiple residents*. More than one resident can be present in many environments. A smart space—for example, a smart house—needs to recognize the activities residents perform in parallel, even when a group performs them.

Human activity understanding encompasses activity recognition and activity pattern discovery. The first focuses on accurate detection of 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 unknown patterns directly from low-level sensor data without any predefined models or assumptions. Thus, an activity pattern discovery researcher builds a pervasive system first and then analyzes the sensor data to discover activity patterns. Although the two techniques are different, both aim to improve human activity technology. In addition, they're complementary to each other—the discovered activity pattern can help define activities that can be later recognized and tracked.

An activity pattern discovery researcher builds a pervasive system first and then analyzes the sensor data to discover activity patterns.

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. Researchers have used several probability-based algorithms to build activity models. The hidden Markov model (HMM) and the conditional random field (CRF) are among the most popular modeling techniques.

Hidden Markov Model

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 without expert knowledge of dementia and its related behavioral science. Fortunately, a researcher can observe signals stemming from complex or unfamiliar activities to indirectly build an HMM. By observing an activity's effects, an HMM can gradually construct the activity model, which the researcher can further tune, extend, and reuse in similar studies.

An HMM is a generative probabilistic model that's used for generating hidden states from observable data.[2] Its main goal is to determine the hidden state sequence ($y_1, y_2, \ldots, y_t$) that corresponds to the observed output sequence ($x_1, x_2, \ldots, x_t$). Another important goal is to learn model parameters reliably from the history of observed output sequences. Figure 1 shows a graphical

 Published by the IEEE CS ■ 1536-1268/10/$26.00 © 2010 IEEE

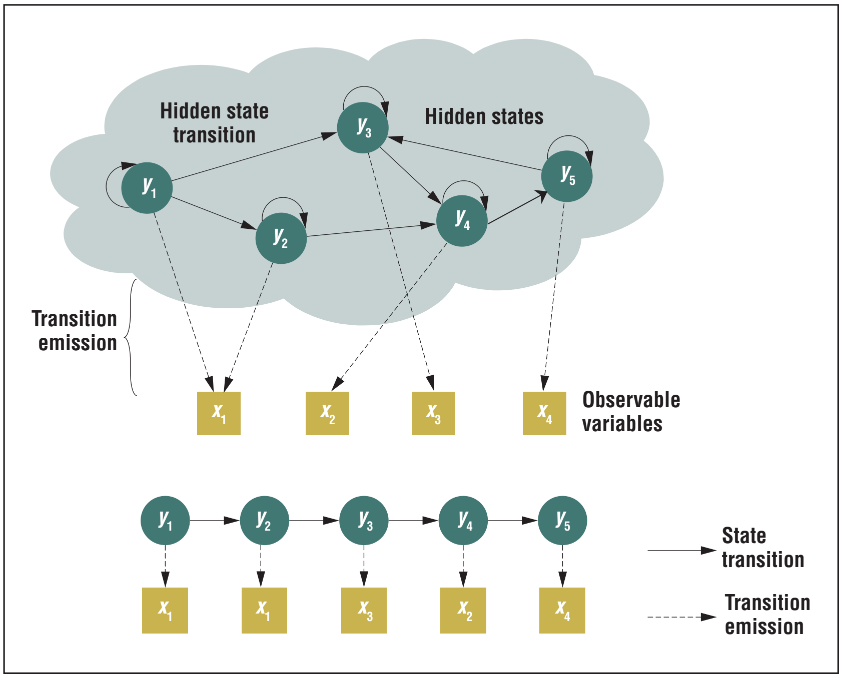

Figure 1. The graphical representation of a hidden Markov model (HMM). By observing sequences of sensor data (observable variables), the HMM discovers hidden state transitions.

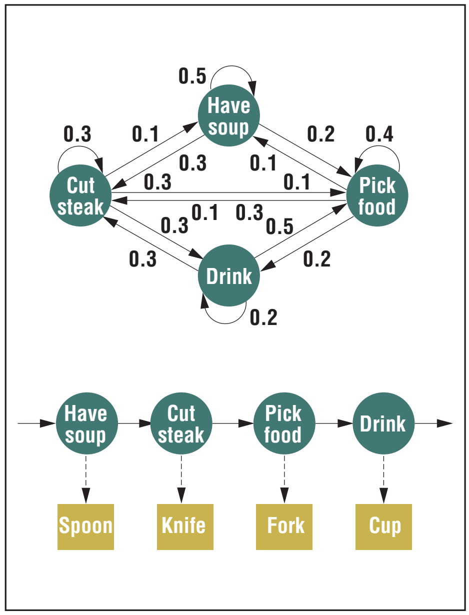

Figure 2. Example of an HMM for an eating activity. Based on the objects used (spoon, knife, fork, or cup, which are the observable variables), we can infer the HMM states and their transitions.

representation of an HMM composed of five hidden states and four observable variables.

An HMM requires two independence assumptions for tractable inference. The first-order Markov assumption of transition is

$$P\left(y_t \mid y_1, y_2, y_3, \ldots, y_{t-1}\right) = P\left(y_t \mid y_{t-1}\right).$$

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}.

The second assumption is conditional independence of observation parameters:

$$P\left(x_t \mid y_t, x_1, x_2, \ldots, x_{t-1}, y_1, y_2, y_3, \ldots, y_{t-1}\right) = P\left(x_t \mid y_t\right).$$

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

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

When an HMM is used in activity recognition, activities are the hidden states and the 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, a researcher can combine individually trained HMM to construct a larger HMM model—for example, of complex activities with a clear subactivities structure. Training is sometimes necessary to "induce" all possible observation sequences that could be needed to find the $p(x,y)$ joint distribution of the HMM.

The Conditional Random Field

Although simple and popular, HMMs have serious limitations, most notably their difficulty in representing multiple interacting activities (concurrent or interwoven).[1] An HMM is also incapable of capturing long-range or transitive dependencies of the observations due to its strict independence assumptions on the observations. Furthermore, without significant training, an HMM might not recognize all the possible observation sequences consistent with a particular activity.

In practice, many activities may have nondeterministic natures in which some steps can be performed in any order. In practice, many activities are concurrent or interwoven. A conditional random field is a more flexible alternative to the HMM that addresses such practical requirements. It's

STANDARDS & EMERGING TECHNOLOGIES

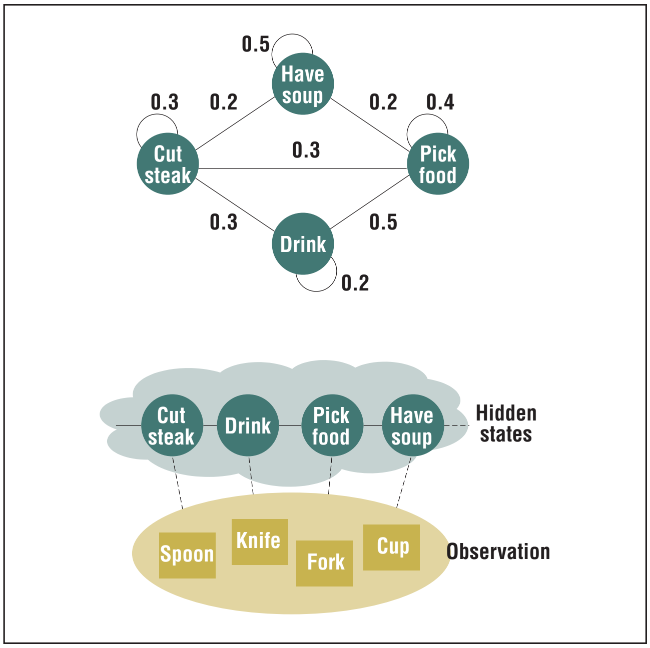

Figure 3. Example of a conditional random field for an eating activity. Observations aren't randomly generated, and hidden states depend on global observations.

a discriminative and generative probabilistic model for the dependence of a hidden variable y on an observed variable x.[2] Both HMMs and CRFs can be 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, dependent 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 can depend on past and even future observations.[2] A CRF is modeled as an undirected 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 $Y=(y_1, y_2, \ldots, y_{t-1}, y_t)$ and observation variables $X=(x_1, x_2, \ldots, x_k)$. The two probabilities, transition and observation, of the HMM are replaced respectively by a transition feature function $r(y_{t-1}, y_t)$ and a state feature function $s(y_t, X)$.[2] Both feature functions return 1 in the simplest case if there's a correlation between its variables. The potential function $p(Y \mid X)$ is computed by the equation[2]

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

where

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

and where λ_i is a weight of correlation that represents the actual potential. In the eating example, the λ_i values shown in Figure 3 are 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 CRF

The potential function is very simple and leads to 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 sequential activities. To model complex activities with concurrent and interleaved subactivities, more sophisticated potential functions are needed to capture long-range (skip-chain) dependencies. In this case, the CRF is called a skip-chain CRF (SCCRF). Given that a skip chain is essentially a linear chain with a larger distance between two variables, it's 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

An emerging pattern (EP) is a feature vector of each activity that describes significant changes between two classes of data.[1] Consider a data set D that consists of several instances, each having a set of attributes and corresponding values. Among the available attributes, some support a class more than 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*.[1] To find these attributes, *support* and *GrowthRate* are computed for every attribute X:[1]

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

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

$$GrowthRate(X) \begin{cases} 0 & \text{if } \mathrm{Supp}_1(X) = 0 \text{ and } \mathrm{Supp}_2(X) = 0 \\ \infty & \text{if } \mathrm{Supp}_1(X) = 0 \text{ and } \mathrm{Supp}_2(X) > 0 \\ \dfrac{\mathrm{Supp}_2(X)}{\mathrm{Supp}_1(X)} & \text{otherwise} \end{cases}.$$

EPs are item sets with large growth rates from $D1$ to $D2$. These EPs are mined from sequential sensor data and applied to recognize interleaved and concurrent activities.[1]

Other Approaches

The number of machine learning models that have been

TABLE 1
Comparison of activity recognition models.

	Hidden Markov model (HMM)	Linear chain conditional random field (CRF)	Skip-chain CRF (SCCRF)	Emerging patterns (EP)
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

used for activity recognition varies almost as greatly as the types of recognized activities and sensor data used. Naïve Bayes classifiers have been used with promising results for activity recognition.[3] These classifiers identify the activity that corresponds with the greatest probability to the set of (assumed conditionally independent) observed sensor values. Other researchers have employed decision trees to learn readable, logical descriptions of the activities.[4]

Comparison of Activity Recognition Methods

Table 1 compares three key aspects of the activity recognition methods we've presented. The SCCRF and emerging pattern methods recognize concurrent and interleaved activities. Except for emerging patterns, all methods require supervised learning, which in turn requires training data for real activity recognition—a major limitation in performing automatic labeling. Changing the sensor environment influences every model, which must then also be changed accordingly.

ACTIVITY PATTERN DISCOVERY

Complementing the idea of activity recognition is automatically recognizing activity patterns in an unsupervised fashion. Tracking only preselected activities ignores important insights that other discovered patterns can provide on the residents' habits and the nature of the environment. In addition, recognizing and tracking automatically discovered activities eliminates the need to prelabel data and use it to train 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,[3] activity patterns are recognized similar to the way 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. Where d is document, w is word, and z is topic, this equation shows the word distribution that's expected for a set of topics:

$$p(w \mid d) = \sum_{z=1}^{T} p(w \mid z) p(z \mid d).$$

Activity patterns can be similarly discovered and topic-word distributions, where words correspond to recognized activities and daily routines, correspond to topic activation (see Table 2). As an example, the process was applied to sensor data collected for one subject. Figure 4a shows the subject's intuitive description of the activities, and Figure 4b 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 extract individual poses from video data. Activities can then be represented by constructing probabilistic context-free grammars using the poses as the grammar alphabet.[6] Figure 5 shows example poses that correspond to collected 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* activity 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*.

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.

STANDARDS & EMERGING TECHNOLOGIES

TABLE 2
Topics and activities.[5]

Topic	Content activities
1.	Sitting/desk activities (1.0)
2.	Sitting/desk activities (.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 a car (0.33), walking and carrying something (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 down/using computer (1.0)
9.	Unlabeled (0.87)
10.	Watching a movie (1.0)

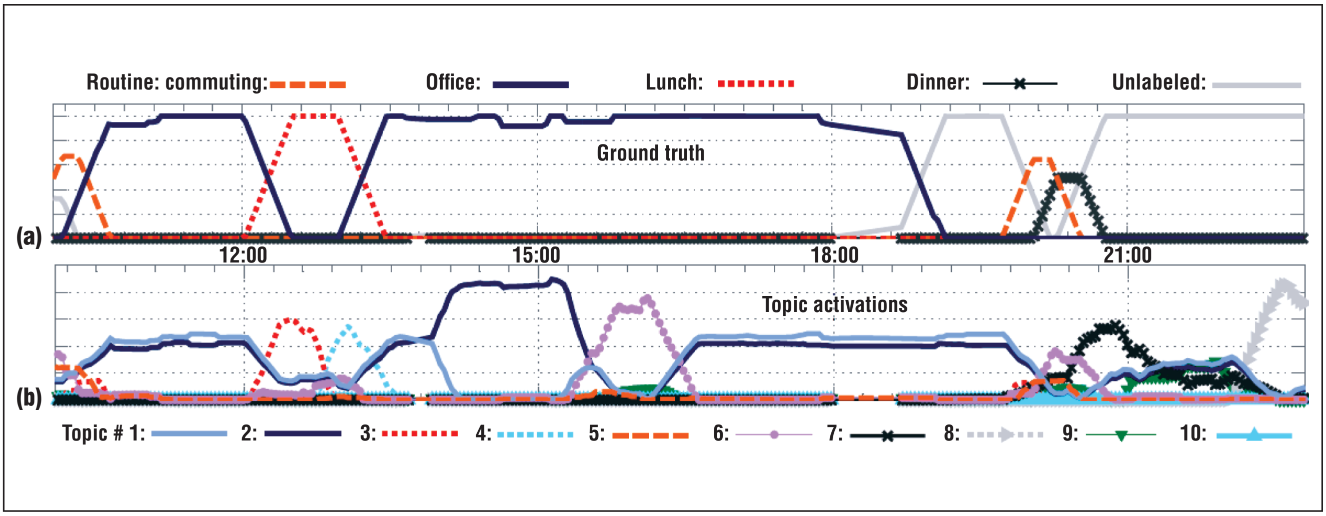

Figure 4. A comparison between ground truth topics and estimated topics. (a) The subject-annotated activity topics for daily routines and (b) the automatically discovered activity topics for a day by a trained system.

Activity and pattern recognition research can benefit significantly from a community effort in which collected activity data is shared and made easily accessible. One effort in this direction is an MIT project hosting longitudinal home activity datasets as a shared resource.[7] The project aims to create a new community database of the various data sets structured as dense, multimodal sensor records of people living in homes that have been instrumented with a sensing infrastructure. One obstacle to sharing is the representation of data sets collected by different researchers. An effort to create a community standard for such representation is the Sensory Dataset Description Language (SDDL) standard spearheaded by the University of Florida.[8] The standard specification is available online.[8]

REFERENCES

1. T. Gu et al., "epSICAR: An Emerging Patterns Based Approach to Sequential, Interleaved and Concurrent Activity Recognition," *Proc. IEEE 7th Ann. Int'l Conf. Pervasive Computing and Comm.* (PerCom 09), IEEE CS Press, 2009, pp. 1–9.

2. C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," L. Getoor and B. Taskar, eds., *Introduction to Statistical Relational Learning*, MIT Press, 2006.

3. E. Munguia-Tapia, S.S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors," *Proc. 2nd Int'l Conf. Pervasive Computing* (Pervasive 04), Springer, 2004, pp. 158–175.

4. U. Maurer et al., "Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions," *Proc. Int'l Workshop on Wearable and Implantable Body Sensor Net-*

works, IEEE CS Press, 2006, pp. 99–102.

5. T. Huynh, M. Fritz, and B. Schiele, "Discovery of Activity Patterns Using Topic Models," *Proc. 10th Int'l Conf. Ubiquitous Computing*, ACM Press, 2008, pp. 10–19.

6. A. Ogale, A. Karapurkar, and Y. Aloimonos, "View-Invariant Modeling and Recognition of Human Actions Using Grammars," *Proc. Int'l Workshop Dynamical Vision*, Springer, 2005, pp. 115–126.

7. S. Intille, "The NSF CRI:CRD Development of Longitudinal Home Activity Datasets as a Shared Resource," 2008; www.mit.edu/~intille/AbstractNSF.html.

8 A. Helal and S. Houssein, "The Sensory Dataset Description Language" project, 2009; www.icta.ufl.edu/persim/sddl.

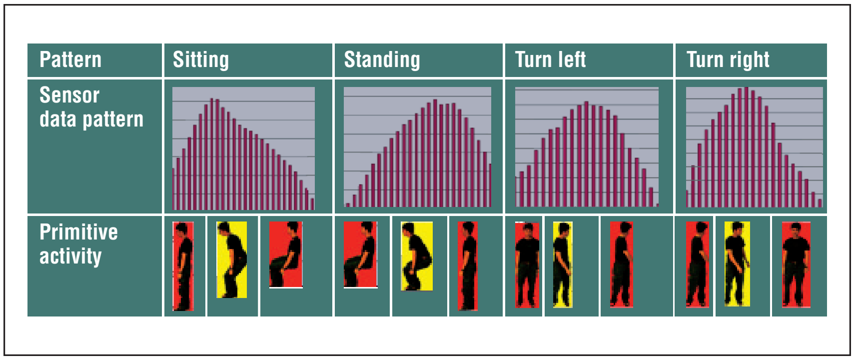

Figure 5. Extraction of motion patterns from sensor data.[6] Each video key frame is composed of three primitive poses.

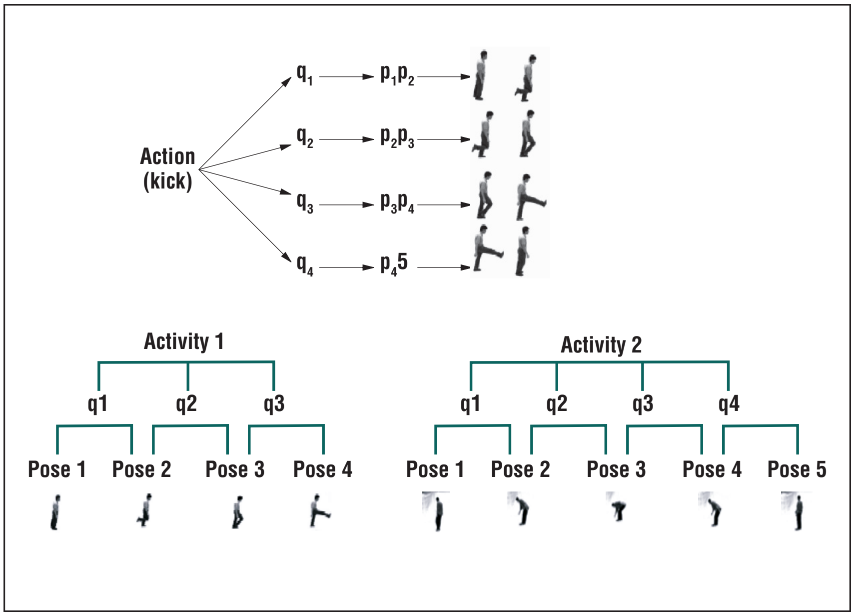

Figure 6. Composite activities.[6] An activity is a combination of multiple poses according to a sequence rule.

Eunju Kim is a PhD candidate in the University of Florida's computer and information science and engineering department. Her research focus is on flexible and scalable activity recognition. Contact her at ejkim@cise.ufl.edu.

Sumi Helal is a professor in the University of Florida's computer and information science and engineering department. His research interests span mobile and pervasive computing systems. Contact him at helal@cise.ufl.edu.

Diane Cook is a Huie-Rogers Chair Professor at Washington State University. Her research interests include machine learning, data mining, graph algorithms, and smart environments. Contact her at cook@eecs.wsu.edu.