Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Recognizing Complex Activities by a Probabilistic Interval-Based Model

Li Liu[1], **Li Cheng**[1,2], **Ye Liu**[1], **Yongpo Jia**[1], **David S. Rosenblum**[1]
[1]School of Computing, National University of Singapore, Singapore 117417
[2]Bioinformatics Institute, A*STAR, Singapore 138671
{dcsliuli,liuye.cis}@gmail.com, chengli@bii.a-star.edu.sg, jiayongpo@nus.edu.sg, david@comp.nus.edu.sg

Abstract

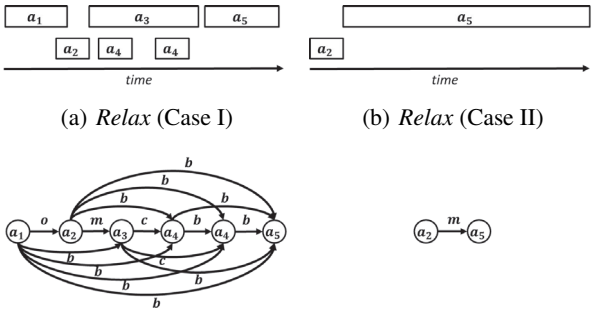
A key challenge in complex activity recognition is the fact that a complex activity can often be performed in several different ways, with each consisting of its own configuration of atomic actions and their temporal dependencies. This leads us to define an atomic activity-based probabilistic framework that employs Allen's interval relations to represent local temporal dependencies. The framework introduces a latent variable from the Chinese Restaurant Process to explicitly characterize these unique internal configurations of a particular complex activity as a variable number of tables. It can be analytically shown that the resulting interval network satisfies the transitivity property, and as a result, all local temporal dependencies can be retained and are globally consistent. Empirical evaluations on benchmark datasets suggest our approach significantly outperforms the state-of-the-art methods.

Introduction

Activity recognition has become an important research field, given its role in facilitating a broad range of applications in areas such as healthcare, sports, smart homes and product recommendations (Padoy et al. 2008). Current techniques are becoming mature to recognize basic actions and movements from cameras or other sensors (Frank, Mannor, and Precup 2010; Gupta and Mooney 2010; Lara and Labrador 2013; Bulling, Blanke, and Schiele 2014; Yürüten, Zhang, and Pu 2014). For example, actions like *drink from cup* can be inferred by sensors attached to the user's arms and the cups, while movements like *walk*, *lie down* and *sit* can be inferred by an accelerometer placed on the user's waist. The *intervals* of these actions and movements can also be obtained as the period of time over which the corresponding sensor states remain unchanged. These actions and movements describe low-level activities that can be inferred from sensors and cannot be further decomposed under application semantics (Saguna, Zaslavsky, and Chakraborty 2013), and are referred to as *atomic activities* here. The main focus of this paper is on complex activities, where a *complex activity* is a collection of temporally related atomic activities. For example, *relax* is a complex activity that may contain six atomic activities, i.e. *walk*, *reach lazychair*, *sit*, *lie down*, and *drink from cup*. As illustrated in Fig. 1(a), modeling complex activities naturally requires the characterization of their temporal dependencies among atomic activities. A complex activity recognition model should also represent uncertainties associated with individual atomic activities as well as their temporal dependencies. It is well known that each individual often possesses a unique style of performing the same complex activity, which may differ noticeably from the others. To further complicate the matter, the same person might perform differently at a different time or location. One such example is provided in Fig. 1(b), where *relax* can also be performed in an alternative manner with only two atomic activities involved. This kind of inherit uncertainty or variability of complex activities usually manifest themselves in terms of the types of the underlying atomic activities and their temporal relationships (Kim et al. 2015).

Despite being a very challenging problem, in recent years there has been a rapid growth of interest in modeling and recognizing complex activities. Semantic-based models (Ryoo and Aggarwal 2009; Morariu and Davis 2011; Helaoui, Riboni, and Stuckenschmidt 2013) have gained attention in recent years for addressing complex activity recognition problems, but they often lack the expressive power to capture and propagate the uncertainties associated with their temporal dependencies. Moreover, formulae and their weights need to be carefully hand-crafted by domain experts, which could be rather difficult in many practical scenarios. On the other hand, the most popular modeling paradigm might be that of the graphical models, which include techniques such as hidden Markov models, Bayesian networks, and conditional random fields (Bui et al. 2008; Hospedales et al. 2011; Liu et al. 2015). While these graphical model-based approaches are capable of managing uncertainties, they are unfortunately rather limited in characterizing rich temporal relationships among activities. In fact, as these models are mostly time-point based, only three relations (i.e. precedes, follows, equals) can be sufficiently captured. Moreover, the time-point based graphical models are computationally expensive when the number of overlapping activities grows (Pinhanez 1999).

The interval temporal Bayesian network (ITBN) (Zhang et al. 2013) is the first interval-based graphical model that combines the Allen relations with the probabilistic description of Bayesian network. This model can capture 13 tem-



(a) *Relax* (Case I) (b) *Relax* (Case II)

(c) Interval network (Case I) (d) Interval network (Case II)

Figure 1: Two examples of the complex activity *relax* and its corresponding interval networks. a_1 =*walk*, a_2 =*reach lazychair*, a_3 =*sit*, a_4 =*drink from cup*, a_5 =*lie*.

poral interval relations among activities. However, since Bayesian network structure is a directed acyclic graph, the ITBN has to remove some temporal relations from the training dataset in order to maintain temporal consistency, which nevertheless would result in information loss. Besides, checking temporal consistency of such triangle relationships and evaluating all possible network structures (i.e. which relation should be ignored or not) are computationally expensive, and would end up being intractable with the growth of the network size. Moreover, in ITBN the network size is fixed to be the same as the number of atomic activities. Subsequently, while it can handle situations where an atomic activity occurs no more than one time during a complex activity, it is nonetheless difficult for ITBN to manage the repetitive occurrences of atomic activities. In other words, in the event that an atomic activity occurs more than once, ITBN faces the awkward situation where each repetitive occurrence of this atomic activity has to be treated as a novel atomic activity.

To address these issues in complex activity recognition, we present a generative probabilistic model based on Allen interval relations. In particular, our approach considers a principled way of dealing with the inherit structural variability in complex activities. Briefly speaking, to describe a complex activity such as *relax*, we propose to introduce an additional latent variable generated from the *Chinese Restaurant Process* or *CRP* (Pitman 2002). Now each resulting table from the *CRP* contains its unique set of atomic activities that together with the corresponding Allen interval relations, characterize a certain cluster of instances that possess similar atomic activities and their temporal dependencies, i.e. a particular *style* of this complex activity. Note the *CRP* also enables our model to depict repetitive occurrences of atomic activities. In addition, an *interval relation generation constraint* is introduced to ensure temporal consistency during the network generation procedure without loss of internal relations. In this way, our generative model is more capable of characterizing the inherit structural variability in complex activities when comparing to existing methods such as ITBN, which is also verified during empirical evaluations to be detailed in later sections.

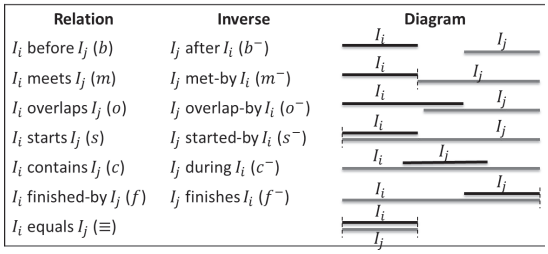


Figure 2: Allen's 13 temporal relations between two intervals. Note that f often refers to *finish* in other literatures.

Problem Formulation

Given a dataset $\mathcal{C}$ of N records from a set of M complex activities, a complex activity recognition model is generated with respect to the interval relations among atomic activities. Each record is a sequence of atomic activity intervals ordered by start-time, i.e. $\langle I_1, I_2, \ldots, I_k \rangle$. An *atomic activity interval* (or *interval* in shorthand) I_i is denoted by a triplet $I_i = (t_i^-, a_i, t_i^+)$. Here t_i^- and t_i^+ refer to the start-time and end-time of the atomic activity a_i, respectively, with $t_i^- < t_i^+$. In the seminal work (Allen 1983), Allen provides an enumeration of 13 feasible temporal relations between two intervals, i.e. $\{b, m, o, s, c, f, \equiv, b^-, m^-, o^-, s^-, c^-, f^-\}$, as summarized in Fig. 2.

An interval network can be used to represent the temporal relationships between atomic activities within a complex activity, where a node represents an interval (i.e. an instantiation of an atomic activity) and a directed link describes the Allen's temporal relationship of the two involved intervals. Each such link is associated with one and only one interval relation that is in the subset of the Allen's temporal relations without negative superscript, i.e. $\mathcal{R} = \{b, m, o, s, c, f, \equiv\}$. It can be verified that the resulting interval network is a directed acyclic graph, whereas the temporal relations on links shall be consistent. Taken an example illustrated in Fig. 1(c), since a_1 *overlaps* a_2 and a_2 *meets* a_3, the temporal relation *before* will appear on the link from a_1 to a_3. More formally, given any two temporal relations $r_{i,j}$ and $r_{j,k}$, where $r_{i,j}$, $r_{j,k}$ refer to the relations on the links from a_i to a_j and from a_j to a_k, respectively, and $r_{i,j}, r_{j,k} \in \mathcal{R}$, the interval relation $r_{i,k}$ on the link from a_i to a_k shall follow the transitivity properties as listed in Table. 1. Take another example, if $r_{i,j} = r_{j,k} = o$, then $r_{i,k} \in \{b, m, o\}$. Meanwhile, the set of the seven relations is closed under *composition* operation, denoted by $r_{i,j} \circ r_{j,k}$. That is to say, if any $r_{i,j}, r_{j,k} \in \mathcal{R}$, then $r_{i,j} \circ r_{j,k} \subseteq \mathcal{R}$. An interval network is *consistent* if and only if the temporal relations on every triangle $\triangle ijk$ in the network satisfy the transitivity properties. Notice that the interval network is different from the interval algebra network (Allen 1983) where each link are labeled with the union of all possible interval relations. In this paper, the term *network* refers to the interval network in our definition, and *relation* refers to the interval relation in $\mathcal{R}$.

A network however characterizes only a possible style of a complex activity. Fig. 1(c) and Fig. 1(d) show two networks that represents the same complex activity *relax* in two

Table 1: The transitivity table for the interval relation $r_{i,k}$ adapted from (Allen 1983).

$r_{i,j} \circ r_{j,k}$	b	m	o	s	c	f	$\equiv$
b	b	b	b	b	b	b	b
m	b	b	b	m	b	b	m
o	b	b	bmo	o	$bmocf$	bmo	o
s	b	b	bmo	s	$bmocf$	bmo	s
c	$bmocf$	ocf	ocf	ocf	c	c	c
f	b	m	o	o	c	f	f
$\equiv$	b	m	o	s	c	f	$\equiv$

different ways of composing atomic activities and their temporal dependencies. On the other hand, a complex activity can be considered as an instantiation of one such network sampled with certain probability that can be decomposed following the network structure. This inspires us to present in what follows a generative probabilistic model where these interval-based networks can be systematically constructed to characterize the complex activities of interests.

Our Model

Let us consider a dataset $\mathcal{C}$ of N records over M complex activities. For any complex activity m ($1 \leq m \leq M$), denote $\mathcal{C}_m \subseteq \mathcal{C}$ the corresponding subset of N_m records. Here each record $c \in \mathcal{C}_m$ is an instance of the m-th complex activity, and is associated with a set of K atomic activities $\mathcal{A} = \{A_1, A_2, \ldots, A_K\}$ and the set of seven non-negative superscripted relations $\mathcal{R}$ stated previously. The number of intervals involved in the record c is denoted as $|c|$.

Denote $G_c = (V_c, E_c)$ the corresponding network of c, with V_c and E_c being the sets of vertices and edges respectively. A network G_c contains $|c|$ nodes $v_{c,1}, \ldots, v_{c,|c|}$ as well as a set of links. Each node needs to be assigned with an atomic activity in $\mathcal{A}$, and each link is to be assigned with a temporal relation in $\mathcal{R}$. We first consider a simple model that contains only the links between two neighbouring nodes, which specifies a set of $|c|-1$ links $e_{c,1,2}, \ldots, e_{c,|c|-1,|c|}$, with $e_{c,i,i+1}$ ($1 \leq i \leq |c|-1$) representing the link of $v_{c,i} \rightarrow v_{c,i+1}$. Similarly, we can obtain a second model with fully connected links that is furnished with the set of all pairwise links with a fixed direction from past to future. It can be seen that any network constructed by both variants of our model are consistent because no inconsistent triangle exists. In what follows, these two variants are referred to as **GPA** (abbreviation of **G**enerative **P**robabilistic model with **A**llen's relations), and **GPA-F** (where **F** denotes that the network is fully connected), respectively.

We further introduce a set of latent tables drawn from the *Chinese Restaurant Process*. In a restaurant with possibly an infinite number of tables, each node (analogous to a customer) is associated with a table, and without loss of generality assume each table contains K atomic activities (analogous to dishes) with associated probabilities. We assume a group of nodes (customers) from the same complex activity (preferring the same cuisine) is more likely to gather on a set of tables where the atomic activities (dishes) relating to the complex activity (preferred cuisine) are served with higher probabilities. The basic process is specified as follows. The first node (customer) chooses the first table. The n-th subsequent node (customer) chooses a table drawn from the following distribution:

$$\begin{cases} \frac{nt_i}{n+\alpha-1} & \text{if choose a non-empty table } t_i, \\ \frac{\alpha}{n+\alpha-1} & \text{if choose a new table,} \end{cases}$$

where nt_i is the number of previous $n-1$ nodes at table t_i, with $\sum_i nt_i = n-1$, and $\alpha > 0$ is a tuning parameter. Note that there is no a priori distinction between the empty tables, and the *CRP*-induced distribution over table assignments is exchangeable and invariant under permutation. Besides, Dirichlet processes extend this construction by serving each table a set of different, independently chosen atomic activities (dishes). In this way, a complex activity (i.e. a class of networks) can be characterized by a unique set of tables and their distributions over atomic activities. Also, networks of various sizes and repetitive occurrences of the same atomic activity can be generated through the *CRP*.

In what follows, we describe the process of generating a network G_c: For each node $v_{c,n}$, a table $\mathbf{t}_{c,n}$ is chosen from *CRP*(α). The first node always chooses the first table t_1 with probability 1. With a table t_k ($k = 1, 2, \ldots$), an atomic activity $\mathbf{a}_{c,n}$ on node $v_{c,n}$ is chosen from a multinomial distribution *Multinomial*(ξ_k) over all atomic activities, and ξ_k is chosen a priori from *Dirichlet*(β), with β being a hyperparameter. Provided with a pair of atomic activities (A_i, A_j), the relation $\mathbf{r}_{c,n-1,n}$ on link $e_{c,n-1,n}$ is chosen from a *Multinomial*($\theta_{i,j}$) over the seven relations in $\mathcal{R}$, where $\theta_{i,j}$ is the parameter vector of the multinomial distribution conditioned on the pair (A_i, A_j), with $1 \leq i, j \leq K$. Under such construction, a consistent network G_c can be associated with atomic activities and relations. Now, given a set of networks associated with $\mathcal{C}_m$ and the maximum number of tables ℓ, our model assumes the following generative process:

---

1 Choose a distribution $\xi_j \sim$ *Dirichlet*(β), for each $j = 1, 2, \ldots, \ell$;

2 For each complex activity record c ($c \in \mathcal{C}_m$),

2.1 Set $\mathbf{t}_{c,1} = t_1$;

2.2 Choose an atomic activity $\mathbf{a}_{c,1} \sim$ *Multinomial*($\xi_{\mathbf{t}_{c,1}}$);

2.3 For each node $v_{c,n}$ ($2 \leq n \leq |c|$),

(1) Choose a table $\mathbf{t}_{c,n} \sim$ *CRP*($\mathbf{t}_{c,1}, \ldots, \mathbf{t}_{c,n-1}; \alpha$);

(2) Choose an atomic activity $\mathbf{a}_{c,n} \sim$ *Multinomial*($\xi_{\mathbf{t}_{c,n}}$);

(3) Choose a relation $\mathbf{r}_{c,n-1,n} \sim$ *Multinomial*($\theta_{\mathbf{a}_{c,n-1},\mathbf{a}_{c,n}}$);

Denote $V = \sum_{c \in \mathcal{C}_m} |c|$ and $W = V - N_m$. The joint distribution of a set of V variables $\mathbf{a}$ for atomic activity assignments, a set of V variables $\mathbf{t}$ for table assignments, and a set of W variables $\mathbf{r}$ for relation assignments, is given by:

$$P(\mathbf{a}, \mathbf{t}, \mathbf{r}; \alpha, \beta, \theta) = \prod_c (P(\mathbf{a}_{c,1}; \beta) \prod_{n=2}^{|c|} P(\mathbf{t}_{c,n} \mid \mathbf{t}_{c,1}, \ldots, \mathbf{t}_{c,n-1}; \alpha) P(\mathbf{a}_{c,n} \mid \mathbf{t}_{c,n}; \beta) P(\mathbf{r}_{c,n-1,n} \mid \mathbf{a}_{c,n-1}, \mathbf{a}_{c,n}; \theta_{\mathbf{a}_{c,n-1},\mathbf{a}_{c,n}})). \quad (1)$$

Notice that although the *CRP* can generate an infinite number of tables, there are at most $\max\{|c|\}$ tables given the training dataset $\mathcal{C}_m$. So we can often set $\ell = \max\{|c|\}$. A

$\ell > \max\{|c|\}$ can also be set to generate a new class of networks with the size greater than any training complex activities. In such cases, atomic activities are chosen with the same probability (i.e. $\frac{1}{K}$) after the $\max\{|c|\}$-th node.

Parameter Estimation

There are two independent parameter estimation problems: the estimation of ℓ node distribution associated parameters $\{\xi_1, \xi_2, \ldots, \xi_\ell\}$ and the estimation of $K \times K$ relation distribution associated parameters $\{\theta_{i,j} : 1 \leq i, j \leq K\}$.

Since the probability distribution of the relation $\mathbf{r}_{c,i,j}$ only relies on the pair of atomic activities (A_i, A_j), the maximum likelihood estimate (MLE) can be used here to learn parameters for the given training dataset $\mathcal{C}_m$. The parameters include the conditional probability for each pair $(A_i, A_j) \in \mathcal{A} \times \mathcal{A}$. According to our generative model, the relations are independently distributed given a pair of atomic activities and the conditional probability distribution is a multinomial over the seven relations in $\mathcal{R}$. Assuming the relations are also identically conditioned on each pair of (A_i, A_j), the likelihood of the parameter $\theta_{i,j}$ for $P(\mathbf{r}_{c,i,j} \mid A_i, A_j)$ with respect to the training dataset becomes:

$$\mathcal{L}(\theta_{A_i,A_j}; \mathcal{C}_m) = \prod_c P(\mathbf{r}_{c,i,j} \mid A_i, A_j; \theta_{i,j}) = \prod_{r=1}^{7} \theta_{i,j,r}^{nr_{i,j,r}}, \quad (2)$$

where $\theta_{i,j} = \{\theta_{i,j,1}, \ldots, \theta_{i,j,7}\}$, $\theta_{i,j,r}$ is the parameter for the r-th relation in $\mathcal{R}$, and $nr_{i,j,r}$ is the number of times the link $A_i \rightarrow A_j$ is labeled with the r-th relation in the training dataset. By taking the logarithm of the likelihood function, applying a Lagrange multiplier to satisfy $\sum_{r=1}^{7} \theta_{i,j,r} = 1$, and setting the partial derivatives to zero, we can get the maximum likelihood estimate for $\theta_{i,j}$ as

$$\hat{\theta}_{i,j,r} = \frac{nr_{i,j,r}}{\sum_{r'=1}^{7} nr_{i,j,r'}}. \quad (3)$$

Unlike the relation associated parameter leaning, it is unfortunately intractable to perform an exact learning for parameters $\{\xi_1, \xi_2, \ldots, \xi_\ell\}$. Instead we develop an approximate inference procedure based on Gibbs sampling. More specifically, for each node, we estimate the posterior distribution on latent table $\mathbf{t}$ based on the following conditional probabilities, which can be derived by marginalizing the above joint probabilities in Eq.(1). The probability of assigning the node $v_{c,n}$ to the table t_ζ $(1 \leq \zeta \leq \ell)$ is shown as

$$P(\mathbf{t}_{c,n} = t_\zeta \mid \mathbf{t}_{c,-n}, \mathbf{a}, \mathbf{r}; \alpha, \beta, \theta) = \begin{cases} \frac{na_{\zeta,\mathbf{a}_{c,n},-n}+\beta}{\sum_{k'=1}^{K} na_{\zeta,k',-n}+\beta K} \times \frac{nt_{c,\zeta}}{n+\alpha-1} & \text{if } \zeta \leq T_{c,n}, \\ \frac{na_{\zeta,\mathbf{a}_{c,n},-n}+\beta}{\sum_{k'=1}^{K} na_{\zeta,k',-n}+\beta K} \times \frac{\alpha}{n+\alpha-1} & \text{if } \zeta = T_{c,n}+1, \end{cases} \quad (4)$$

where $na_{\zeta,k}$ is the number of times node has been assigned to the atomic activity A_k on the ζ-th table t_ζ, $nt_{c,\zeta}$ is the number of times the previous $n-1$ nodes in G_c are assigned to the table t_ζ. $T_{c,n}$ is the number of non-empty tables occupied by the previous $n-1$ nodes in G_c, and $\sum_{i=1}^{T_{c,n}} nt_{c,i} = n-1$. The suffix $-n$ of na means the count that does not include the current assignment of table for the node $v_{c,n}$. The detailed derivations and procedures of the Gibbs sampling are provided in the supplementary material.

With the sampled tables available, we can readily estimate the distributions of ξ_y $(1 \leq y \leq \ell)$ as

$$\xi_{y,k} = \frac{na_{y,k} + \beta}{\sum_{k'=1}^{K} na_{y,k'} + \beta K}. \quad (5)$$

Now we are ready to evaluate the probability $P(\mathbf{a}, \mathbf{r}; \mathcal{C}_m)$ of the occurrence of a set of atomic activities $\mathbf{a}$ and their relations $\mathbf{r}$ given the m-th complex activity by integrating out the latent table $\mathbf{t}$, which gives

$$P(\mathbf{a}, \mathbf{r}; \mathcal{C}_m) = \prod_k (\sum_y \xi_{y,k}) \times \prod_{i,j,r} \theta_{i,j,r}. \quad (6)$$

GPA-F: The Variant with Fully Connected Links

So far we have described GPA, a simple variant of our approach with only relations between two temporally neighbouring nodes. This model may potentially lead to the loss of relations. For example, in Fig. 1(c), a_3 contains a_4 and a_4 is before a_5. There are five possible relations between a_3 and a_5. The previous model can only depict the possibilities of relations between neighbouring nodes rather than between any pair of nodes.

As a remedy of this issue, we extend the model by generating networks with links between any pair of nodes, which gives rise to the GPA-F variant considered in this section. The generation of atomic activities remains the same as in the previous model. To generate a relation, we have to ensure network consistency. We define an *interval relation generation constraint* $\mathbf{I}_{c,n',n} \subseteq \mathcal{R}$ for each link $e_{c,n',n}$ $(1 \leq n' < n \leq |c|)$, where

$$\mathbf{I}_{c,n',n} = \begin{cases} \bigcap_{u=n'+1}^{n-1} (\mathbf{r}_{c,n',u} \circ \mathbf{r}_{c,u,n}) & \text{if } n > n'+1, \\ \mathcal{R} & \text{if } n = n'+1, \end{cases}$$

where $n = n'+1$ indicates $v_{c,n'}$ and $v_{c,n}$ are neighbouring nodes. Each $\mathbf{r}_{c,n',n}$ can only be chosen from the set $\mathbf{I}_{c,n',n}$. Given a pair of atomic activities (A_i, A_j) and a constraint $I_z \subseteq \mathcal{R}$ $(z = 1, 2, \ldots)$ on $e_{c,n',n}$, $\mathbf{r}_{c,n',n}$ is chosen from a multinomial distribution *Multinomial*$(\varphi_{i,j,z})$ over all possible relations in I_z, where $\varphi_{i,j,z}$ is the parameter vector of the multinomial distribution. The generative process for a network with fully connected links is given as follows:

. . . (the same as GPA) . . .

2 For each complex activity c $(c \in \mathcal{C}_m)$,

. . . (the same as GPA) . . .

2.3 For each node $v_{c,n}$ $(2 \leq n \leq |c|)$,

. . . (the same as GPA) . . .

(3) Choose $\mathbf{r}_{c,n-1,n} \sim$ *Multinomial*$(\varphi_{\mathbf{a}_{c,n-1},\mathbf{a}_{c,n},\mathcal{R}})$;

(4) for each node $v_{c,n'}$ $(n-2 \geq n' \geq 1)$,

(4-1) Set $\mathbf{I}_{c,n',n} = \bigcap_{u=n'+1}^{n-1} (\mathbf{r}_{c,n',u} \circ \mathbf{r}_{c,u,n})$;

(4-2) Choose $\mathbf{r}_{c,n',n} \sim$ *Multinomial*$(\varphi_{\mathbf{a}_{c,n'},\mathbf{a}_{c,n},\mathbf{I}_{c,n',n}})$;

Theoretical analysis is also carried out to ensure the networks generated by our model are temporally consistent and relation lossless. This is summarized in the two theorems below, with the proofs relegated to the supplementary material.

Theorem 1 ***(Network Consistency)***
A network G_c constructed by the above generative process is consistent.

Theorem 2 ***(Lossless)***
Any possible combination of relations in a consistent network can be constructed through the above generative process.

We further consider the interval relation generation constraints. From the 7 relations in $\mathcal{R}$, naively the set of possible composition relations between intervals would be $2^7 = 128$ when including the empty relation. Fortunately with the above generative process and the transitivity table in Table 1, there are only 12 feasible unions for any constraint $\mathbf{I}_{c,n',n}$, as displayed in Table 2. Denote $\mathcal{I} = \{I_z : 0 \leq z \leq 11\}$ the set of all 12 possible unions, with $\mathbf{I}_{c,n',n} \in \mathcal{I}$ being any interval relation generation constraint. In particular, $\mathbf{I}_{c,n-1,n} = I_{11} = \mathcal{R}$ and $\varphi_{\mathbf{a}_{c,n-1},\mathbf{a}_{c,n},\mathcal{R}} = \theta_{\mathbf{a}_{c,n-1},\mathbf{a}_{c,n}}$. For a union containing only one relation (i.e. $I_0 - I_7$), the probability of choosing that relation is always one. If all the constraints in a network (except for $\mathbf{I}_{c,n-1,n}$) are one of these unions, the network amounts to be the same as our first variant GPA. Notably, although GPA-F is lossless, it can produce a network with empty relation $\emptyset$ on links. To remedy this issue, $\emptyset$ is considered as a special relation on links in our context with always zero probability. In practice, the empty relation never appears in the training dataset.

Table 2: The 12 possible interval composition relations.

$I_0 = \{\emptyset\}, I_1 = \{b\}, I_2 = \{m\}, I_3 = \{o\},$
$I_4 = \{s\}, I_5 = \{c\}, I_6 = \{f\}, I_7 = \{\equiv\},$
$I_8 = \{b, m, o\}, I_9 = \{o, c, f\}, I_{10} = \{b, m, o, c, f\},$
$I_{11} = \{b, m, o, s, c, f, \equiv\}$

Moreover, the total number of the set of variables $\mathbf{r}$ for relation assignments is $\sum_c \frac{|c|\times(|c|-1)}{2}$. Now, the joint distribution of our GPA-F model with fully connected links is given by:

$$P(\mathbf{a}, \mathbf{t}, \mathbf{r}; \alpha, \beta, \varphi) = \prod_c (P(\mathbf{a}_{c,1}; \beta) \prod_{n=2}^{|c|} P(\mathbf{t}_{c,n} \mid \mathbf{t}_{c,1}, \ldots, \mathbf{t}_{c,n-1}; \alpha)$$
$$P(\mathbf{a}_{c,n} \mid \mathbf{t}_{c,n}; \beta) \prod_{n'=n-1}^{1} P(\mathbf{r}_{c,n',n} \mid \mathbf{a}_{c,i}, \mathbf{a}_{c,n}; \varphi_{\mathbf{a}_{c,n'},\mathbf{a}_{c,n},\mathbf{I}_{c,n',n}})).$$

Similar to the estimation of the parameter θ in the GPA model i.e. Eq. (2), the parameter φ can be learned by MLE:

$$\hat{\varphi}_{i,j,z,r} = \frac{nr_{i,j,z,r}}{\sum_{r'=1}^{|I_z|} nr_{i,j,z,r'}}. \quad (7)$$

where $\varphi_{i,j,z,r}$ is the parameter for the r-th relation in I_z, and $nr_{i,j,z,r}$ is the number of times the link $A_i \rightarrow A_j$ with constraint I_z is labeled with the r-th relation in the training dataset. Notice that for $1 \leq z \leq 7$, $\mid I_z \mid = 1$ and $\hat{\varphi}_{i,j,z,r} = 1$. The probability of the occurrence of a new set of atomic activities $\mathbf{a}$ and their relations $\mathbf{r}$ given the m-th complex activity is updated as

$$P(\mathbf{a}, \mathbf{r}; \mathcal{C}_m) = \prod_k (\sum_y \xi_{y,k}) \times \prod_{i,j,z,r} \varphi_{i,j,z,r}. \quad (8)$$

Empirical Evaluations

Datasets

Three complex activity recognition datasets are considered in our experiments.

OSUPEL dataset (Brendel, Fern, and Todorovic 2011). This is a video-recorded dataset of actual two-on-two basketball games. The players are tracked and labelled with six atomic activities: *pass*, *catch*, *hold ball*, *shoot*, *jump*, and *dribble*, which are used to form two complex offensive play activities as defined in (Zhang et al. 2013). The numbers of instances for the two offensive play types are 56 and 16, respectively.

Opportunity dataset (Roggen et al. 2010). It contains five complex daily living activities (*relax*, *coffee time*, *early morning*, *cleanup*, and *sandwich time*) performed by four subjects and recorded in a room with 72 sensors of 10 different modalities simultaneously deployed either in objects or on the body. These complex activities involves a total number of 211 atomic activities such as *sit*, *walk*, *reach table*, *open door*, and so on. Overall the dataset contains a total number of 28,976,744 sensor data records with sampling rate of 30Hz.

Composable activities dataset (CAD14) (Lillo, Soto, and Niebles 2014). It is consist of 693 RGB-D video records captured by Microsoft Kinect, with 14 actors performing 16 complex activities such as *talk phone and drink*, *walk while clapping*, *talk phone and pick up*, among others. The total number of atomic activities is 26, including *clap*, *talk phone*, and so on. Each complex activity contains 3 to 11 intervals (i.e. instances of atomic actions).

These three datasets contain unique challenges: The OSUPEL dataset is comprised of a small number of atomic activities and records with simple relations, the Opportunity dataset involves a large number of atomic activities and records with intricate relations, while the CAD14 dataset have a large number of complex activities with diverse forms of relations.

Experimental Set-Ups

The classification performance of our approach is compared against four established graphical model-based methods: HMM (Oliver and Horvitz 2005), skip-chain conditional random field or SCCRF (Hu and Yang 2008), DBN (Oliver and Horvitz 2005) and ITBN (Zhang et al. 2013). *Accuracy* is employed as the evaluation metric, which is computed as the proportion of true results among the total number of records examined. The evaluation procedures for recognizing atomic activities of basketball playing from ordinary videos and those of composable activities from RGB-D videos proposed by (Zhang et al. 2013) and (Lillo, Soto, and Niebles 2014) are adopted here, respectively. The activity recognition chain (ARC) system developed by (Bulling, Blanke, and Schiele 2014) is utilized for atomic activity recognition from sensors. Note the records that are not annotated to any activities are labeled as *null* activity.

To monitor the empirical convergence behaviors of the Gibbs sampling components (Eq. 4) for GPA and GPA-F, we utilized the Raftery and Lewis diagnostic tool (Raftery

and Lewis 1992) to detect the burn-in (i.e. convergence), and we set na and nt to the average values of their first 500 samples collected right after the burn-in stage. Besides, the hyperparameters α and β are generally unknown before the start of Gibbs sampling and therefore need to be estimated. In our experiment, we used the convergent method (Minka 2000) that iteratively updates these hyperparameters by approximately estimating the objective maximum likelihood function values. More details can be found in the supplementary material. Besides, a small smoothing constant S ($S = 0.00001$) is introduced to avoid the numerical issue of division by zero, i.e. $\sum_{r'=1}^{7} nr_{i,j,r'} = 0$ in Eq. (3) or $\sum_{r'=1}^{|I_z|} nr_{i,j,z,r'} = 0$ in Eq. (7).

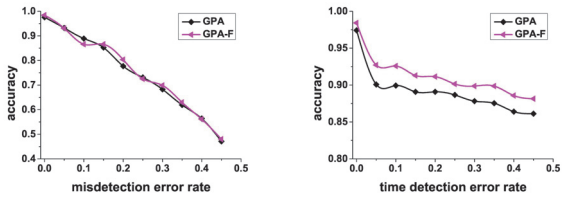
Experimental Results

Table 3 shows the averaged accuracy results over 5-fold cross-validations. The two variants of our approach clearly outperform the other methods with a large margin on all three datasets. This is mainly due to their abilities to take advantage of the rich temporal dependency information between atomic activities. Notably, although ITBN also encodes temporal information in the model, during training it has to remove all the inconsistent relations. This might explain why ITBN gives the worst performance among all comparison methods for the CAD14 dataset where a large amount of inconsistent relations exist.

Table 3: Accuracies on the three evaluation datasets.

	HMM	SCCRF	DBN	ITBN	GPA	GPA-F
OSUPEL	0.53	0.67	0.58	0.69	**0.79**	0.76
Opportunity	0.74	0.94	0.83	0.88	**0.98**	0.96
CAD14	0.93	0.95	0.95	0.51	0.97	**0.98**

Robustness under atomic activity detection errors. We evaluate the performance robustness of comparison methods under various atomic activity recognition errors. They are first evaluated by testing against two types of synthetic errors that are common with atomic activity recognition, i.e. misdetection errors (the correct activity is not detected or is falsely recognized as another activity) and time detection errors (either the start-time or end-time of an activity is falsely detected). Synthetic misdetection errors are simulated by perturbing the true labels under a varying amount of error rates; while time detection errors are simulated by perturbing the start and end time with a varying noise level of 10, 20, and 30 percent of the maximal temporal distance between neighboring intervals, respectively. Besides, the performances are also evaluated under real detected errors caused by three classifiers (i.e. kNN, SVM and Decision Tree) that are specifically built to recognize atomic activities and their time durations from sensor records. Table 4 depicts the comparison results on the Opportunity Dataset. ITBN is relatively more robust to these errors than other comparison methods as being capable to capture interval relations. Moreover, it is clear that the proposed GPA and GPA-F are significantly more robust than ITBN (with around 15% – 100% performance boost) as well as others under various atomic activity detection errors.



(a) under misdetection error (b) under time detection error

Figure 3: Comparisons of GPA and GPA-F on CAD14 dataset.

Table 4: Accuracies under synthetic and detected errors of atomic activity recognition on Opportunity dataset.

atomic activities error rate	classification accuracy					
	HMM	SCCRF	DBN	ITBN	GPA	GPA-F
under synthetic misdetection errors						
0.1	0.31	0.71	0.73	0.79	**0.92**	0.88
0.2	0.29	0.69	0.67	0.74	**0.87**	**0.87**
0.3	0.22	0.69	0.65	0.71	0.83	**0.84**
under synthetic time detection errors						
0.1	0.16	0.65	0.45	0.76	0.83	**0.88**
0.2	0.16	0.65	0.45	0.72	0.83	**0.85**
0.3	0.16	0.65	0.45	0.69	0.82	**0.83**
under real detected errors						
0.165 (kNN)	0.66	0.69	0.62	0.54	**0.91**	0.79
0.242 (SVM)	0.58	0.10	0.54	0.46	**0.83**	0.71
0.315 (DT)	0.16	0.10	0.04	0.38	**0.71**	0.54

GPA v.s. GPA-F. We consider two factors that may affect the performances of the two variants, i.e. misdetection error and time detection error. As suggested in Fig.3 and from the results on the other datasets in supplemental material, it seems GPA-F is more robust to the time detection errors than GPA, but not to the misdetection errors. This may be due to the fact that GPA-F contains more temporal relations.

Runtime. We present three variables that may affect the runtime, i.e. the number of atomic activities, the number of intervals per record and the number of training (or testing) records per complex activity. The empirical runtime is tested on different settings by varying one variable while fixing others. The training and testing stages of each approach were investigated separately. The results of varying the number of atomic activities are shown in Fig. 4, while others are reported in the supplemental material. It can be seen that the GPA and GPA-F variants performs nearly the same, which on average outperform the other methods at both training and testing stages. Theoretically, the time complexities of GPA and GPA-F are the same, which are $\mathcal{O}(MK^2+NT_n\ell^2)$ and $\mathcal{O}(NM\ell)$ at the training and testing stages, respectively, where T_n is the number of iterations in Gibbs sampling.

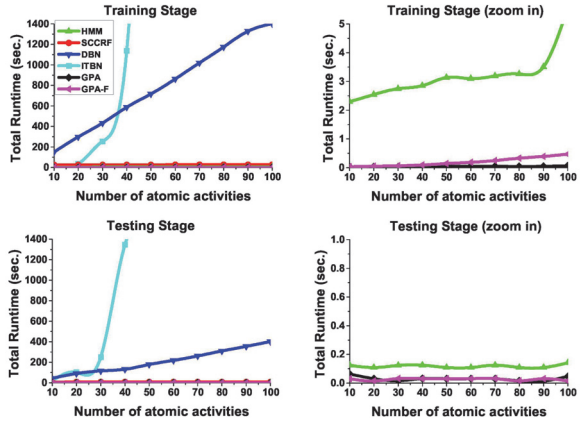


Figure 4: Runtime comparison by varying the number of atomic activities.

Conclusion

In this paper, we present a probabilistic interval-based model where the *CRP* model is incorporated to capture the inherit structural varieties of complex activities. It is more efficient and flexible than existing methods including ITBN for complex activity recognition. As for future work, we will further investigate the difference between our two models on more datasets, and we will consider relaxing the assumption that a network is either chain-based or fully connected and will instead learn the network structures.

Acknowledgments

This research was supported in part by grants R-252-000-473-133 and R-252-000-473-750 from the National University of Singapore, and A*STAR JCO grants 15302FG149 and 1431AFG120.

The supplemental material can be downloaded from the link https://drive.google.com/folderview?id=0B20-pV67EQA7Q3NHR3hWVG85MFE.

References

Allen, J. 1983. Maintaining knowledge about temporal intervals. *ACM Communications* 26(11):832–843.

Brendel, W.; Fern, A.; and Todorovic, S. 2011. Probabilistic event logic for interval-based event recognition. In *CVPR*, 3329–3336.

Bui, H. H.; Phung, D. Q.; Venkatesh, S.; and Phan, H. 2008. The hidden permutation model and location-based activity recognition. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 1345–1350. AAAI Press.

Bulling, A.; Blanke, U.; and Schiele, B. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys* 46(3):33.

Frank, J.; Mannor, S.; and Precup, D. 2010. Activity and gait recognition with time-delay embeddings. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 1581–1586. AAAI Press.

Gupta, S., and Mooney, R. J. 2010. Using closed captions as supervision for video activity recognition. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 1083–1088. AAAI Press.

Helaoui, R.; Riboni, D.; and Stuckenschmidt, H. 2013. A probabilistic ontological framework for the recognition of multilevel human activities. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 345–354.

Hospedales, T. M.; Li, J.; Gong, S.; and Xiang, T. 2011. Identifying rare and subtle behaviors: A weakly supervised joint topic model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(12):2451–2464.

Hu, D. H., and Yang, Q. 2008. Cigar: Concurrent and interleaving goal and activity recognition. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 1363–1368. AAAI Press.

Kim, E.; Helal, S.; Nugent, C.; and Beattie, M. 2015. Analyzing activity recognition uncertainties in smart home environments. *ACM Transaction on Intelligent Systems and Technology* 6(4):52.

Lara, O. D., and Labrador, M. A. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15(3):1192–1209.

Lillo, I.; Soto, A.; and Niebles, J. C. 2014. Discriminative hierarchical modeling of spatio-temporally composable human activities. In *CVPR*, 812–819.

Liu, Y.; Nie, L.; Liu, L.; and Rosenblum, D. S. 2015. From action to activity: Sensor-based activity recognition. *Neurocomputing*.

Minka, T. 2000. Estimating a dirichlet distribution. Technical report, MIT.

Morariu, V. I., and Davis, L. S. 2011. Multi-agent event recognition in structured scenarios. In *CVPR*, 3289–3296.

Oliver, N., and Horvitz, E. 2005. A comparison of hmms and dynamic bayesian networks for recognizing office activities. In *User Modeling*. 199–209.

Padoy, N.; Blum, T.; Feussner, H.; Berger, M.-O.; and Navab, N. 2008. On-line recognition of surgical activity for monitoring in the operating room. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 1718–1724. AAAI Press.

Pinhanez, C. S. 1999. *Representation and recognition of action in interactive spaces*. Ph.D. Dissertation, MIT.

Pitman, J. 2002. Combinatorial stochastic processes. Technical report, UC Berkeley.

Raftery, A. E., and Lewis, S. 1992. How many iterations in the gibbs sampler. *Bayesian statistics* 4(2):763–773.

Roggen, D.; Calatroni, A.; Rossi, M.; Holleczek, T.; Forster, K.; Troster, G.; Lukowicz, P.; Bannach, D.; Pirkl, G.; Ferscha, A.; et al. 2010. Collecting complex activity datasets in highly rich networked sensor environments. In *International Conference on Networked Sensing Systems*, 233–240.

Ryoo, M. S., and Aggarwal, J. K. 2009. Semantic representation and recognition of continued and recursive human activities. *International Journal of Computer Vision* 82(1):1–24.

Saguna, S.; Zaslavsky, A.; and Chakraborty, D. 2013. Complex activity recognition using context-driven activity theory and activity signatures. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20(6):32.

Yürüten, O.; Zhang, J.; and Pu, P. 2014. Decomposing activities of daily living to discover routine clusters. In *Proceedings of the 28rd AAAI Conference on Artificial Intelligence*, 1348–1354. AAAI Press.

Zhang, Y.; Zhang, Y.; Swears, E.; Larios, N.; Wang, Z.; and Ji, Q. 2013. Modeling temporal interactions with interval temporal bayesian networks for complex activity recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(10):2468–2483.