

普适计算中复合事件检测的研究与挑战*

周春姐⁺, 孟小峰

中国人民大学 信息学院, 北京 100872

The Researches and Challenges of Complex Event Detection in Pervasive Computing*

ZHOU Chunjie⁺, MENG Xiaofeng

School of Information, Renmin University of China, Beijing 100872, China

+ Corresponding author: E-mail: lucyzcj@ruc.edu.cn

ZHOU Chunjie, MENG Xiaofeng. The researches and challenges of complex event detection in pervasive computing. *Journal of Frontiers of Computer Science and Technology*, 2010, 4(12): 1057-1072.

Abstract: In pervasive computing environments, wide deployment of sensor devices has generated an unprecedented volume of atomic events. However, most applications such as healthcare, surveillance and facility management, as well as environmental monitoring require such events to be filtered and correlated for complex pattern detection. Therefore how to extract interesting, useful and complex events from low-level atomic events is becoming more and more important in daily life. At present, there are a lot of researches of complex event detection, and each has its own particular research points. Some pay attention to the time information, especially the importance of time interval; some research in the complex event detection in distributed data sources; recently some propose the probabilistic data management on complex event detection. Due to the increasingly importance of complex event detection, this paper analyzes the challenges in the research of complex event detection, and gives a survey of existing researches from three aspects including event types, time information, and precision of data. Finally, some open issues and future researches are given.

Key words: pervasive computing; sensor; complex event detection; time interval; probabilistic data

*The National Natural Science Foundation of China under Grant No.60833005, 60573091(国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2007AA01Z155 (国家高技术研究发展计划(863)); the Doctoral Fund of Ministry of Education of China under Grant No.200800020002 (国家教育部博士点基金).

Received 2010-04, Accepted 2010-06.

摘要: 普适计算环境中, 传感器设备的大规模使用产生了数量巨大、错综复杂的原子事件, 而现实中的许多应用却更注重复合事件的检测, 例如: 健康护理、监督设施管理、环境/安全监控等, 因此如何从这些底层的原子事件中抽取人们感兴趣的、有用的复合事件就变得越来越重要。目前, 针对复合事件检测有大量的研究, 其内容各有侧重。有的重视时间因素, 特别强调时间段的重要性; 有的研究分布式数据源中的复合事件检测; 近期有人提出了不确定性数据上的复合事件检测。由于复合事件检测日益重要, 对复合事件检测研究中存在的挑战性问题进行了分析, 从事件类型、时间因素和数据的精确程度 3 个方面归纳总结了复合事件检测现有的研究成果, 并指出了未来的发展方向。

关键词: 普适计算; 传感器; 复合事件检测; 时间段; 不确定性数据

文献标识码: A **中图分类号:** TP391

1 引言

随着计算机、通信、网络、微电子、集成电路等技术的发展, 信息技术的硬件环境和软件环境发生了巨大变化。这种变化使得通信和计算机构成的信息空间, 与人们生活和工作的物理空间正在逐渐融为一体。普适计算(pervasive computing)的思想就是在这种背景下产生的。普适计算环境的特点是以人为本, 为用户提供更高效精确的、无处不在的人性化服务, 即系统可以根据用户的爱好、需求对服务进行自由的裁剪和定制。为了达到上述目标, 在追踪和监控等实际应用中, 正在大规模地使用无线感知网络和无线射频技术(radio frequency identification, RFID)等传感器设备。这些设备的广泛部署产生了大量的、直接反映物理世界的原子事件(定义见第 2 章)。一个标准的传感器设备通常有成千上万条数据记录, 这使得操作人员通过观测每一条记录来发现异常事件变得非常困难; 此外, 在事后分析异常事件时, 需要操作人员及时找出相关记录。然而传统的检测方法缺乏智能分析, 数据无法被有效地检索, 只能根据大致的时间段来人工查找, 导致数据分析工作消耗了大量的工作时间和精力。

解决上述问题的一个有效方法是对事件进行自动智能分析, 对数据集中出现的用户感兴趣的事件进行实时提取和记录, 从而达到及时报警

并利用存储的事件信息来有效地检索数据。例如, 在零售管理系统中, 通过将一系列原子事件整合成复合事件(定义见第 2 章), 可以及时地发现偷窃行为。当发生这样一种场景时: 商品被从货架上取下→没有结账→带出商店, 系统就会自动发出预警信号。

复合事件检测在现实世界中有许多应用, 例如: 健康护理^[1-2]、监督设施管理^[3]、环境监控^[4]、供应链管理^[5]以及各种普适计算应用^[6-8], 都需要将大量原子事件过滤成相互关联的复合模式检测, 或者转化成可以直接服务于终端应用的、富含语义的新事件。其中, 室内环境下的应用如:

(1) 健康护理中需要系统实时地从大量的传感器数据中推断出被看护者的行为, 看护者可能想获知一系列的行为流^[1-2,9], 如: 被看护者是否按时吃药, 是否按时吃饭, 在睡觉之前是否刷牙, 以及状态是否正常等, 从而判断出被看护者是否已经被很好地照顾;

(2) 通过对大量原子事件的分析处理, 抽出用户感兴趣的、有用的复合事件, 长期的复合事件的存储管理, 可以总结分析出用户的行为模式, 从而可以在默认的情况下, 自动为用户提供各种服务;

(3) 安全系统可以通过分析不同日期的、相同时刻的、与用户相关的复合事件, 来决定是否

需要实施某些预警信息。室外环境下的应用如：在机场、车站、港口、建筑物周围，以及街道、小区等场所，可用于检测、分类、跟踪和记录过往行人、车辆及其他可疑物体，或用于判断是否有行人及车辆在禁区内发生长时间徘徊、停留、逆行、奔跑、打斗等异常行为。

由上可见，这些应用需求都需要根据用户的指定，或者根据某些规则，自动地从这些底层的、原子性的事件中抽取一些复合事件，从而将反映物理世界的原始数据实时地转化成可直接服务于终端应用的信息。因此，复合事件检测的研究意义重大，它将会成为未来普适计算环境中非常重要的一部分，并将渗入到未来生活的各个方面。

本文结构如下：第 2 章将用例子说明什么是复合事件检测，以及其研究的必要性；第 3 章介绍在事件检测研究中所面临的挑战；第 4 章归纳总结了事件检测的研究现状；第 5 章说明有待进一步解决的问题；最后是总结。

2 复合事件检测的基本概念

符合普适计算“以人为本”的特点，复合事件检测的目标就是把从物理空间得到的原子事件转化为用户感兴趣的，可以直接服务于用户的复合事件。本章将给出复合事件检测中的一些基本概念。

事件被定义为用户感兴趣的行为。例如：在房间中发现某个人、启动 CPU 定时器、在网络中拒绝某个攻击服务等都是不同应用领域中的事件的例子。所有的事件都表示特定的行为，然而他们的复杂性却存在很多差异。例如：启动定时器是一个瞬间的、简单的发现，而拒绝攻击服务则需要计算多个简单事件。因此，事件可被划分为原子事件和复合事件^[10]。下面分别给出原子事件和复合事件的详细定义。

定义 1(原子事件) 原子事件是在某个时间点

上瞬间发生的、原子性的事件。可表示为 $E_{atomic_i} = Action < o_i, p_i, t_i >$ ，其中 o_i 表示某个对象(object)，它可以是某个人，也可以是某个物体； p_i 表示地点或位置(place)，即对象 o_i 的当前位置； t_i 表示某时间点；Action 表示 t_i 时刻，对象 o_i 在 p_i 位置的行为，即是一个原子事件 E_{atomic_i} 。例如：Coffee('Mary', 'Room 301', 10:00 am)表示上午 10 点这一刻，Mary 在 301 房间喝咖啡。

定义 2(复合事件) 复合事件通常是由用户指定的、或者自动地从原子事件中抽取出来的，某个时间段上发生的事件。可表示为 $E_{complex_i} = < Q_i, E_i, T_i >$ ，其中 Q_i 表示某个查询(query)，它只在用户指定的情况下是有效的，用来表示用户的查询条件，而当自动抽取时其值为空； E_i 表示一系列原子事件的集合，即 $E_i = \{1 \leq i \leq n \mid E_{atomic_i}\}$ ，集合中的各个原子事件之间是相互关联的，它们之间存在某种集合运算关系(如正相关、负相关、并行、串行等)； T_i 表示某个时间段，在时间段 T_i 中，查询/抽取出一系列相互关联的原子事件就组成一个复合事件。例如：当检测到原子事件“Mary is in her office”，“Mary is in coffee room”，“Mary is in her office again”时，就可以从中抽取复合事件“Mary is getting coffee”。复合事件是由原子事件或其他复合事件通过一系列事件组合运算得到的。

复合事件检测的处理过程如图 1 所示。

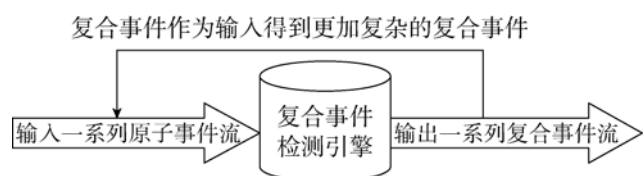


Fig.1 The principle of complex event detection

图 1 复合事件检测原理图

将一系列的原子事件流作为输入，经过复合事件检测引擎的分析处理，输出结果就是一系列复合事件。考虑到特定情况下用户可能需要更为复杂的复合事件，因此复合事件检测引擎应该支

持将输出的复合事件再次作为输入,从而得到更加复杂的复合事件。复合事件检测的方法有两种:(1)根据用户指定。即用户提出某个查询规则,复合事件检测引擎根据此查询规则,从输入的一系列原子事件中检测出满足条件的一系列复合事件;(2)自动检测。即复合事件检测引擎根据用户的生活习性和历史记录,自动地从一系列输入的原子事件中检测出用户感兴趣的,能够为用户提供直接服务的复合事件。

在普适计算环境中,原子事件的数量是巨大的,构成复合事件的不同原子事件之间的关系是复杂的,大量的事件分布在不同节点上,产生的事件具有分布、并发、异步、随机、不确定等特点。并且,事件源可能是传感器、无线设备、移动设备等,事件的到达因网络连接的不稳定以及移动设备的移动而发生时延、失序的情况。因此理想的复合事件检测方法应该能够高效地进行分布式检测、并发异步事件的检测、失序事件的检测以及断接检测、移动检测。在普适计算环境中,事件分布在地理位置分散的节点中,其中有些节点是移动节点。若用一个中心节点来探测基本事件,形成最后的复合事件表达式,那么这个节点将成为事件探测的瓶颈。因此,要高效地检测事件,就要根据普适计算环境的特点和系统要求选择合适的检测方法。目前存在的复合事件检测方法有:(1)基于事件树的复合事件检测^[11];(2)基于图的检测方法^[12];(3)基于自动机的复合事件检测^[13];(4)基于Petri网的复合事件检测^[14];(5)流水线操作的检测方法^[15]。上述每种复合事件检测方法都各有利弊:GEM^[11]考虑了事件发生与检测之间的延迟,并通过指定最大能容忍的延迟来处理事件检测的失序。但它假定了存在一个良好的全局同步时钟,这不适合没有集中管理以及存在时钟漂移的大规模松耦合的分布式系统。由于没有考虑不可预知的延迟,不能有效进行移动数据库中的断接检测和移动检测;Snoop^[12]只

提供简单的时间模型,把事件看作一个确定的时间点,原子事件根据定义来确定时间点,而复合事件的时间则根据其语义来定义时间点,这比较适合应用于集中式系统或局域网;ODE^[13]数据库中使用的有限状态自动机表示事件,能直观地表达现实中的事件,建立自动机并据此检测复合事件。但是纯粹的自动机不检测带参数的事件,不能表示事件的时序关系,不能检测并发事件,这不符合分布式系统的需求。

由于上述的各种复合事件检测方法都没有考虑不确定性的问题,而不确定性又是普适计算环境中的本质特征。因此对现有研究工作的归纳总结不是按照不同的检测方法来分类,而是从复合事件检测的3个特征(时间因素、复合事件表示方法、数据的精确程度)入手,分别说明和归纳(见第4章)。

3 复合事件检测研究面临的挑战

复合事件检测的应用领域广泛,研究意义重大,但是在普适计算环境下,数目巨大的、并发的、无序的原子事件,以及大量不确定性数据的产生给复合事件检测的研究工作带来了许多挑战。其中包括:

(1)大量的事件流。由于传感器等设备的大量部署,每秒钟都会有成千上万的事件产生。例如,在零售管理系统中,商品的入库和出库、上架和下架、购买、结账、带出商店等都会产生大量的事件流。复合事件检测必须能妥善处理这些规模巨大的事件流,准确分析各事件之间的相互关联和影响。

(2)时间窗口大小的选定。复合事件检测中通常使用一个滑动时间窗口来保存一系列感兴趣的事件。在许多应用场景中,这些窗口是很大的,并且与某个查询相关的事件在窗口中是分散分布的。不像原子事件检测那么简单,复合事件检测必须抽取所有相关的事件,返回满足某个查

询的所有可能的结果，这就使数据处理的复杂性很大。例如，图 2(a)表示时间窗口为过去 6 小时的事件分布情况；图 2(b)表示时间窗口为过去 12 小时的事件分布情况。与 2(b)相比，2(a)中包含的事件数目少，检测的复杂度也小，但是由于它选取的时间段短，因此会出现检测不全的现象，无法返回所有可能的结果。例如图中椭圆阴影的检测结果在时间窗口图 2(a)中就无法得到。图 2(b)中虽然能检测全面些，但是包含的事件多，从而检测的复杂度很大。可见，事件数目和检测全面性是两个矛盾体，如何权衡两者，选定合理的时间窗口大小是一个挑战性的问题。

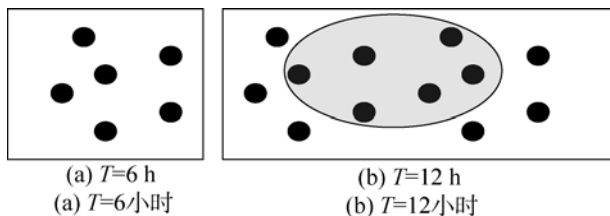


Fig.2 The choice of time window

图 2 时间窗口的选择

(3) 时间同步问题。事件发生的时刻及事件之间发生的先后关系，表明了大量的基本事件是如何构成相关的复合事件的，因此基于时序的事件检测要求各节点的时间同步。而普适计算环境中，没有集中统一的管理。节点之间松散耦合，

各节点的时钟偏频和漂移会造成节点间的时间不同步，节点移动与不稳定连接使事件发生到事件检测之间产生不可预测的延迟，这些情况会造成事件检测失序。

(4) 事件的不确定性。复合事件检测中存在两类不确定性，一种是事件的局部不确定性，另一种是事件的全局不确定性。当进行事件检测时只考虑元组/对象自身的不确定性，认为它们独立于其他的对象/元组，称之为事件的局部不确定性；例如第 2 章所提到的喝咖啡的例子中，Mary 想什么时间喝咖啡，到哪个咖啡屋喝咖啡，以及喝咖啡持续的时间等都是不确定的。如表 1(a)所示，Mary 喝咖啡的时间可能是 10:15 am 也可能是 9:55 am，可能在 1 号咖啡屋也可能在 2 号咖啡屋，持续时间可能是 15 min 也可能是 17 min。但是这些因素只由 Mary 自身的意愿决定，与其他人无关，因此称为事件的局部不确定性。当需要考虑事件之间的关联性和相互影响时，称之为事件的全局不确定性。仍以 Mary 喝咖啡为例，假设 Mary 喜欢在多数人喝咖啡的时候，和 Joe 一起去喝咖啡，这时 Mary 喝咖啡的时间、地点、持续时间等不仅仅由 Mary 自身的意愿决定，而且受周围其他人的情况以及 Joe 的意愿等不确定性因素共同决定。如表 1(b)所示，7 月份由于受到 Joe 的影响，

Table 1(a) The event local probability

表 1(a) 事件的局部不确定性

日期	姓名	开始喝咖啡的时间	咖啡屋的编号	喝咖啡持续时间/min
3月1日	Mary	10:15 am	1	15
3月2日	Mary	9:55 am	2	17
3月1日	Joe	8:55 am	1	10
3月2日	Joe	9:05 am	1	8

Table 1(b) The event global probability

表 1(b) 事件的全局不确定性

日期	姓名	开始喝咖啡的时间	咖啡屋的编号	喝咖啡持续时间/min
7月1日	Mary	9:35 am	1	12
7月2日	Mary	9:25 am	1	10
7月1日	Joe	9:35 am	1	12
7月2日	Joe	9:25 am	1	10

Mary 喝咖啡的时间提前了, 持续时间也比 3 月份缩短了, 但是他们不同日期的喝咖啡时间和持续时间等仍然是不确定的。在许多重要应用中, 不确定性已经成为本质特征, 因此在进行复合事件检测时必须全面考虑事件的不确定性。

4 复合事件检测的研究现状

如第 2 章所提到的, 已有的基于事件树的复合事件检测、基于图的检测、基于自动机的复合事件检测等都不能很好地满足普适计算环境的需求。因此, 针对普适计算环境的特点, 将主要从复合事件检测的 3 个特征入手对目前的研究情况进行归纳总结。目前对复合事件检测的研究一般包含以下几个方面: 从事件类型来考虑, 描述了复合事件的表示方法; 从时间角度来考虑, 描述了各种时序表示方法; 从数据的精确程度来考虑, 对不确定性数据进行分析处理。现有研究工作对这 3 个方面的考虑各有侧重, 总结如下。

4.1 现有研究的分类

根据复合事件检测的上述 3 个特性, 对现有的研究工作进行分类总结如图 3 所示。其中, 3 个坐标轴分别对应 3 个特性: 时间(时间点和时间段)、数据(精确的和不确定的)、事件(原子的和复合的)。3 个坐标轴将空间分为 8 个象限, 如图

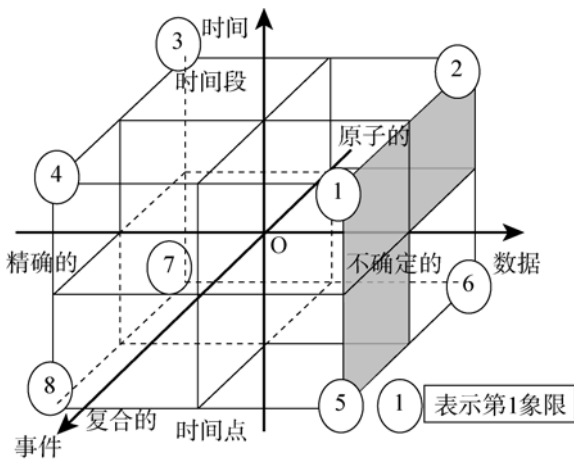


Fig.3 The existing researches of complex event detection
图 3 复合事件检测的现有研究总结

中的标注, 每个象限都对应 3 个特性的不同属性值。

如图 3, 第 7 象限的立方体区域表示在时间点上针对确定性数据的原子事件的研究, 目前这方面的研究工作已经有很多^[16]; 第 3、4、6、8 象限的立方体区域包括时间段上的精确数据的原子事件, 时间段上的精确数据的复合事件, 时间点上的不确定性数据的原子事件, 时间点上的精确数据的复合事件, 目前已经有一些这方面的研究工作^[10,17-22]; 第 1、2、5 象限的立方体区域主要是针对不确定数据方面的, 包括时间段上的不确定性数据的原子事件, 时间段上的不确定性数据的复合事件和时间点上的不确定性数据的复合等, 目前还没有相关的研究工作。其中, 如表 2 所示, 第 3、4、6、8 象限的立方体区域中有关于时间点、原子事件和确定性数据的^[16]; 有关于时间点、复合事件和确定性数据的^[10,17]; 有关于时间段、原子事件和确定性数据的^[18-20]; 有关于时间段、复合事件和确定性数据的^[21]; 有关于不确定性数据的^[22]。

Table 2 The comparison of existing researches
表 2 现有研究工作的分析比较

相关研究工作	时间段	复合事件	不确定性数据
文献[16]	No	No	No
文献[10, 17]	No	Yes	No
文献[18-20]	Yes	No	No
文献[21]	Yes	Yes	No
文献[22]	No	No	Yes

4.2 复合事件表示方法

如第 2 章所提到的, 事件可被划分为原子事件和复合事件, 复合事件是从一系列原子事件中抽取得到的。因此, 提出了一种最简单的复合事件表示方法^[22]。例如, “Mary 喝咖啡” 这个复合事件可被表示成 3 个连续的原子事件: (1) “Mary 在她的办公室里”; (2) “Mary 在咖啡屋”; (3) “Mary 又回到她的办公室里”。如果假设 Mary 的

办公室在220房间，咖啡屋位置不确定，则这个复合事件可被表示为：

$$q_{MaryCoffee} = At('Mary', '220') \\ (\sigma_{CRoom(l)} At('Mary', l)); At('Mary', '220')$$

这种方法存在一个强假设，就是事件是相互独立的，事件与事件之间不会相互影响。但这与实际情况是很不符合的，比如上面的例子中，如果还存在一个事件“咖啡屋里没有咖啡了”，这时按照上述方法(不考虑事件之间的相互关联和影响)就会得出错误的结论，因为实际上 Mary 是没有喝到咖啡的。

文献[17]提出了一种事件描述语言，它可以用来实现事件的过滤、关联和相互转换。下面以在第1章中所提到的零售管理为例，来详细介绍这种事件描述语言。在零售管理中，处理商品的存放错误通常会耗费大量的时间和人力。此事件描述语言与RFID技术相结合，就能提供一种方法来自动处理这一过程，从而减少大量的人力，并能加快货架的补充。处理商品误放的过程可以表示如下：

Event SEQ(shelf-reading x , shelf-reading y , !
(any(counter-reading, shelf-reading) z))

Where [id] $x.shelf_id$ $y.shelf_id$ $x.shelf_id = z.shelf_id$

Within 1 hour

其中，SEQ操作保证用户感兴趣的事件以特定的顺序发生。上述过程认为“在货架1上读到一个条款，紧接着在货架2上又读到相同的条款，并且此条款没有在结账口读到，也没有被重新放回到货架1上”是商品误放的情况，但是“在货架1上读到一个条款，在结账口又读到相同的条款”不属于商品误放的情况。where中用已经定义好的变量来比较不同事件之间的属性，上述过程比较的是满足SEQ的3个事件的 id 属性。“ $x.shelf_id \neq y.shelf_id$ ”保证了shelf-reading的前两项指向不同的货架，而“ $x.shelf_id = z.shelf_id$ ”保证了如果any操作返回一个shelf-reading，那么所读

取的条款不是来自货架1的。within指定一个时间段，例如上述过程中的1小时，用户感兴趣的事件必须在此时间段内发生。但是这种方法也存在几个限制条件，他假设所有的事件都是完全有序的，不考虑事件的并发性；并且，他只考虑了直接由原子事件组合得到的复合事件，而没有考虑由复合事件整合成的更为复杂的事件。

由以上分析可见，已有的处理复合事件的方法都存在一定的缺陷，假设条件太强，考虑不太全面，因此一种新的、全面的、高效的复合事件检测方法亟待提出。

4.3 时序关系表示方法

每个事件都对应一个时间段，用来表示其发生周期。对原子事件而言，其时间段就是一个点，开始和结束时间重合；对复合事件而言，其对应的时段包含所有子事件的时间段^[23]。现有的研究工作大多只考虑了时间点，都假设事件是没有持续时间的^[24-27]。这种假设通常将事件简化成一个有序序列，例如：“头痛->胃痛->呕吐”。然而现实世界中的很多事件都是有持续时间的，并且这些事件之间的时序关系也是很复杂的^[18-20]，所以上述这种时序模式不足以表达复杂的时序关系。在医疗、多媒体、气象学和财政学等领域，事件的持续时间都起了很重要的作用。例如：许多糖尿病患者的症状都是血糖的增高和尿糖的缺失同时并存，只有准确地表示这两者的重叠性才能很好地诊断；又如，骨热病的一般症状是在发热后的第三天血小板就开始减少，只有能够很好地表示这种间隔性和时序性，才能准确地把握治疗时间。

为了有效地抽取基于时间段的复合事件，就需要一种独特的、无损耗的表示方法来获取事件之间的时态关系。传统的方法是用Allen的区间代数^[28]来表示基于时间段的两个事件之间的时态关系，如表3所示。然而，当要获取3个或更多事件之间的时态关系时，这种方法就失效了。

Table 3 The temporal relationship between E_i and E_j
表 3 事件 E_i 和 E_j 之间的时态关系

Relation	Interval algebra	Dual relation
E_i Before E_j	$(E_i.end < E_j.start)$	After
E_i Meet E_j	$(E_i.end = E_j.start)$	Met-by
E_i Overlap E_j	$(E_i.end > E_j.start) \wedge (E_i.end < E_j.end) \wedge (E_i.start < E_j.start)$	Overlapped-by
E_i Start E_j	$(E_i.start = E_j.start) \wedge (E_i.end < E_j.end)$	Started-by
E_i Finished-by E_j	$(E_i.end = E_j.end) \wedge (E_i.start < E_j.start)$	Finish
E_i Contain E_j	$(E_i.start < E_j.start) \wedge (E_i.end > E_j.end)$	During
E_i Equal E_j	$(E_i.start = E_j.start) \wedge (E_i.end = E_j.end)$	Equal

许多研究者试图运用分层的方法来表示事件之间的时态关系^[29-30]，但是这种表示法是有损耗的，因为它不能保持事件潜在的瞬时结构。任何有损耗的表示法都会导致许多伪造的模式，例如非频繁模式可能会成为频繁模式。

现有的基于时间段的很多算法或者是有损耗的表示法^[18]，或者没有很好的扩展性^[19-20]。如表 3 中对事件的时态表示就存在歧义性，给定表示法 $(A \text{ Overlap } B) \text{ Overlap } C$ ，无法推断 C 仅仅与 B 重叠，或者 C 同时与 A 、 B 重叠。图 4 给出了这种时态模式的不同解释，这些不同的解释就会导致对具有确切关系的事件的错误推断。为了解决这个问题，提出了^[21]一种无损耗的表示法，它是在分层表示法的基础上引入了一些附加信息。即用 5 个变量(包含数目 c 、结束数目 f 、相交数目 m 、重叠数目 o 、开始数目 s)来区分所有可能的情况。因此，复合事件 E 可表示为：

$$E = (E_1 R_1 [c, f, m, o, s] E_2) R_2 [c, f, m, o, s] E_3) \dots R_{n-1} [c, f, m, o, s] E_n)$$

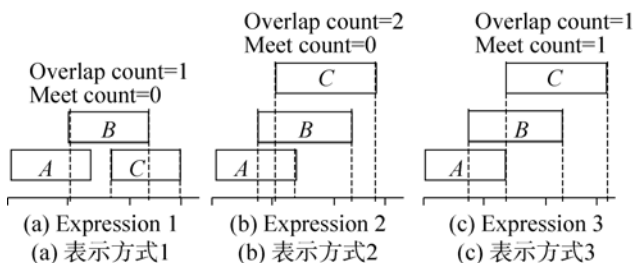


Fig.4 Different expressions of pattern $(A \text{ Overlap } B) \text{ Overlap } C$

图 4 模式 $(A \text{ Overlap } B) \text{ Overlap } C$ 的不同解释

图 4 中的时态模式可分别表示如下：

$$(A \text{ Overlap}[0,0,0,1,0] B) \text{ Overlap}[0,0,0,1,0] C$$

$$(A \text{ Overlap}[0,0,0,1,0] B) \text{ Overlap}[0,0,0,2,0] C$$

$$(A \text{ Overlap}[0,0,0,1,0] B) \text{ Overlap}[0,0,1,1,0] C$$

目前对复合事件检测的研究中大多引入时间窗口的概念。时间窗口表示事件存储在事件序列中的时间，超过这个时间，此事件的记录则会被抛弃。它表示一个时间范围，有 3 个参数，即窗口的起始时间 TWB、终止时间 TWE 和窗口的大小 TWL，其中， $TWE = TWB + TWL$ 。时间窗口的大小可由系统默认或程序员来指定。前者是静态不变的，不能随着网络传输情况的变化而改变；后者可以通过一定的算法动态设置窗口。窗口设置过大，会在事件队列中存储过多失效事件；设置过小，会使许多复合事件检测不到。因此，复合事件检测中时间方面的研究还需要更加的深入细致。

4.4 数据的精确程度

复合事件检测中的一个重要挑战就是数据的不确定性(例如：RFID 数据)。产生数据不确定性的原因很多，例如：(1) 数据错误或丢失，这主要是由电源波动和噪声等因素造成的。研究表明，在现实应用中，RFID 的读取比率仅为 60% ~ 70%，也就是说至少有 30% 的数据被丢失^[31-32]；(2) 数据矛盾，例如两个 Sensor 读取的 Mary 的位置不同，该如何确定 Mary 的真实位置^[31]；(3) 粒度不匹配，例如选取的时间粒度是 ms，而某个事件查

询是“Mary 在 2008 年 5 月的前 3 天在做什么？”这就会导致 $(3 \times 24 \times 60 \times 60 \times 1\,000)$ 不确定的可能性。然而，现有的复合事件检测系统都假设数据是确定的，例如：Cayuga^[3]、SASE^[15]，和 SnoopIB^[33]，这些系统都无法检测不确定性事件流。

处理不确定性数据的常用方法是建立一个数据模型，并且将原始数据作为模型的输入数据。其中一个标准的方法就是用时态图模型，而最简单的时态图模型就是隐式马尔可夫模型(hidden Markov model, HMM)。通常，HMM 通过一系列观察值来推断隐含的状态信息。例如，基于传感器数据来推断一个人的位置。在实时应用中，通常运用一种更加复杂的技术——滤除(smoothing)^[34]。滤除技术不仅可以推断一个人的更加精确的位置，而且可以提供不同时刻此人的位置之间的关系。例如，如果知道 Mary 在 $t=7$ 时刻进了办公室，那么很有可能在 $t=8$ 时刻她仍然在办公室。也就是说，Mary 在 t 时刻的位置和 $t+1$ 时刻的位置是相关的。

关于不确定性的、相关联的隐式马尔可夫模型已经有很多的研究^[35]，但是这些研究都没有考虑不同时间之间的关联性。为了更好地满足现实应用，提出了在相关联的、不确定性数据流上的复合事件查询方法^[22]。

假设 A_1, A_2, \dots, A_k 是 k 个属性值，其中 A_i 在区域 D_i 范围内取值。令 $\bar{D} = D_1 \times D_2 \times \dots \times D_k$ ，并且 $\bar{D}_\perp = \bar{D} \cup \{\perp\}$ 。 A_1, A_2, \dots, A_k 上的部分随机变量是一个函数 $P: D_\perp \rightarrow [0,1]$ ，并且 $\sum_{d \in D_\perp} p(d) = 1$ 。例如，

假设 e 为事件“Mary 上午 10 点在 326 房间喝咖啡的可能性是 0.2，在房间 327 喝咖啡的可能性是 0.7。”那么该属性值的随机变量可表示为：

$$P[e = \text{'Room326'}] = 0.2$$

$$P[e = \text{'Room327'}] = 0.7$$

同时，可以注意到，事件“Mary 上午 9 点 50 在 326 房间喝咖啡”与事件“Mary 上午 9 点 55 在 326 房间喝咖啡”是正相关的，而与事件“Mary

上午 9 点 50 在 327 房间喝咖啡”是负相关的。

假设 $\bar{e} = (e^{(1)}, e^{(2)}, \dots, e^{(t)}, \dots)$ 是一系列不确定性事件流，那么如果满足 $P[e^{(t+1)} | e^{(1)}, e^{(2)}, \dots, e^{(t)}] = P[e^{(t+1)} | e^{(t)}]$ ，则 \bar{e} 就是马尔可夫链。从而序列 $\bar{d} = (d^{(1)}, d^{(2)}, \dots, d^{(t)})$ 的不确定性值就可以用贝叶斯规则定义为：

$$\mu(\bar{d}) \stackrel{def}{=} P[e^{(1)} = d^{(1)}] \cdot \prod_{i=2, \dots, t} P[e^{(i)} = d^{(i)} | e^{(i-1)} = d^{(i-1)}]$$

目前在复合事件检测中考虑不确定性的研究还很少，然而在许多重要的应用中都涉及到数据的不确定性，因此对不确定的复合事件检测进行研究分析已势在必行。

4.5 复合事件检测方法

基于以上的介绍，对普适计算中复合事件检测的 3 个特征已经有了比较深入的理解。下面针对普适计算中复合事件检测的两类主流方法(见第 2 章，分别介绍一个典型的研究工作。

4.5.1 根据用户的指定实现复合事件检测

已有的事件检测实现模型都是基于某种特定的数据结构，如基于事件树的、基于有向图的、基于自动机的、基于 Petri 网的等。在这些模型中，查询执行必须严格符合该特定数据结构的内在模式，而不能使用其他类似的方法来实现查询。并且，这些模型的扩展性也很差，它们无法支持更为丰富的查询语言，从而无法满足很多重要应用的需求。为了克服以上不足，提出了一种复合事件检测方法^[17]，其关键的数据结构就是一个带有查询的事件序列，而没有严格的限制，其扩展性也有所改善。以下面的查询为例来详细介绍。

```
Event SEQ (A x1, B x2, ! (C x3), D x4)
```

```
Where [attr1, attr2] \w x1.attr3 = '1' \w x1.attr4 < x4.attr4
```

```
Within T
```

在这个查询中，A、B、C、D 表示 4 种不同的事件类型，SEQ 操作保证这些事件以特定的顺序

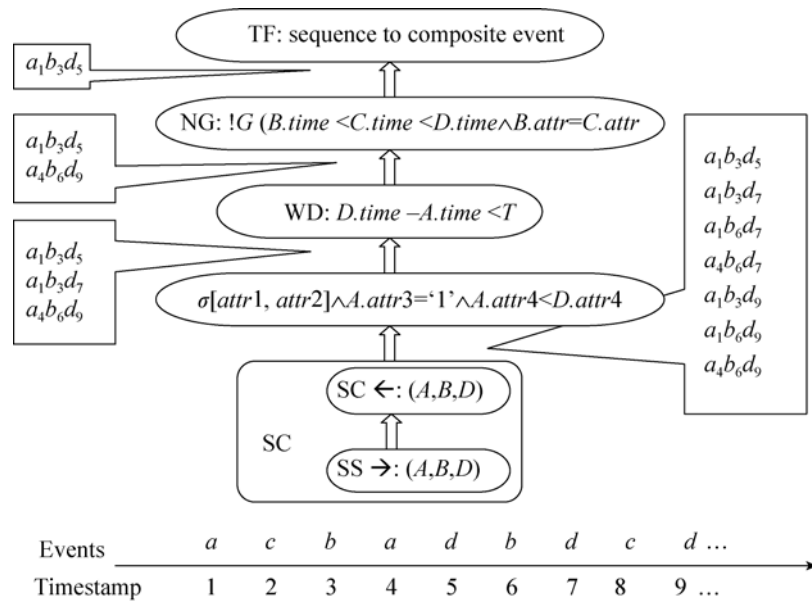


Fig.5 The process of complex event detection with query

图5 带查询的复合事件检测过程图

发生, 即A先发生, 其次B发生, 最后D发生, 并且B与D之间不允许有C发生; attr1, attr2是A、B、C、D的公共属性, attr3是事件A的属性, 事件A在属性attr4上的值小于事件D在属性attr4上的值; T表示一个特定的时间窗口大小。这种带查询的复合事件检测的实现过程如图5所示。

图5中最下面的事件流, 其小写字母(如a)表示的事件类型就是对应的大写字母值(如A)。每个事件下面的数字表示该事件发生的时间。

SSC(sequence scan and construction, 序列扫描和构造)是由两个操作组成的: (1) SS(序列扫描), 即扫描事件流, 从而发现与之匹配的子序列类型。以上述查询为例, SSC就将“!C”从SEQ(A, B,!C,D)移除, 从而得到子序列类型(A,B,D); (2) SC(序列构造), 即反向查询, 从而构造所有的事件序列, 将事件流转换成事件序列流。其中每个事件序列都唯一对应一种子序列类型。如图5, SSC的输出就是从底层事件流中产生的7个事件序列, 每个事件序列都对应子序列类型(A,B,D)的3个元组, 其中小写字母表示事件类型, 下标表示发生的时间。

σ (selection, 选择)根据查询条件对每个事件序列进行过滤, 除掉不满足条件的所有事件序列, 如图5, 7个事件序列中只有3个通过了选择。

WD(window, 时间窗口)检查每个事件序列中第一个事件和最后一个事件之间的时间差是否小于所设定的时间窗口大小T。图5中, T被设定为6小时, 因此, 第二个事件序列又被过滤掉。

NG(negation, 否认)处理被SSC忽略的SEQ中的否定元组。如图5, 对每个输入的事件序列检查在事件b与d之间是否有c发生。如果存在这样的事件c, 那么该事件序列就被删除, 图5中NG的第二个输入序列又被过滤掉了。

TF(transformation, 转换)将所得到的事件序列转换成复合事件。

尽管这种带查询的复合事件检测方法存在诸多优点, 但它本身还存在一些强假设条件, 有待进一步完善。如: 假设所输入的事件流是流水线型的, 有严格的时间序列, 而不考虑事件的并发等情况; 另外, 没有考虑更加复杂的复合事件类型, 即没有考虑将得到的复合事件再次作为输入, 从而得到更加复杂的复合事件的情况。

Table 4 The generation of 4-pattern
表 4 4-模式的产生

频繁 3-模式	频繁 2-模式
(A Overlap[0,0,0,1,0] B) Before[0,0,0,1,0] <u>D</u>	C Contain[1,0,0,0,0] D
(A Before[0,0,0,0,0] F) Before[0,0,0,0,0] <u>G</u>	A Before[0,0,0,0,0] D
(A Overlap[0,0,0,1,0] B) Overlap[0,0,0,2,0] <u>C</u>	B Before[0,0,0,0,0] D
(A Overlap[0,0,0,1,0] <u>C</u>) Contain[1,0,0,0,0] D	A Before[0,0,0,0,0] F
(A Before[0,0,0,0,0] D) Before[0,0,0,0,0] <u>E</u>	A Overlap[0,0,0,1,0] C
(A Overlap[0,0,0,1,0] B) Before[0,0,0,0,0] <u>E</u>	B Overlap[0,0,0,1,0] C
(B Overlap[0,0,0,1,0] <u>C</u>) Contain[1,0,0,0,0] D	D Before[0,0,0,0,0] F
(B Before[0,0,0,0,0] D) Before[0,0,0,0,0] <u>E</u>	A Overlap[0,0,0,1,0] B
	B Before[0,0,0,0,0] F

4.5.2 自动检测

提出了一种基于时态模式的复合事件检测方法^[21]。以往的研究都是从两个 $(k-1)$ 时态模式中产生 k 级模式，但是这种方法会导致产生大量无用的模式，因此文献[21]引入模式中“控制事件” (dominant event)的概念。如果一个事件在模式 P 中发生，并且此事件的结束时间是 P 的所有事件中最晚的，那么就称该事件是模式 P 的控制事件。因此，如果 $(k-1)$ -模式中的控制事件正好是一个 2-模式的第一个事件，则这两个模式就可以做连接，组成一个 k -模式的复合事件。如表 4 所示，3-模式的控制事件有下划线，2-模式的第一个事件被加粗。可以通过连接有共享控制事件的 3-模式和 2-模式来得到 4-模式。例如，连接表 4 中第 1 列的第 3 个模式和第 2 列的第 1 个模式，可以得到 4-模式 $((A \text{ Overlap}[0,0,0,1,0] B) \text{ Overlap}[0,0,0,2,0] C) \text{ Contain}[1,0,0,0,0] D$ 。

这种复合事件检测方法的关键是维护频繁 2-模式的实时更新，将不满足条件的时态模式及时地从频繁 2-模式的列表中删除。该条件 1 定义为：如果一个 $(k+1)$ 模式是由一个频繁 k 模式和一个 2-模式产生的，并且该 2-模式至少在 $(k-1)$ 个频繁 k 模式中发生过，那么该 $(k+1)$ 模式就可以作为一个候选模式(详细证明见文献[21])。基于上面的定义，从一系列频繁 k 模式中产生一个 2-模式列表的同时，会为列表中的每个实体记数，这个数

字表示包含该实体的频繁模式的数目。当一个实体的记数小于 $(k-1)$ 时，就将该实体从 2-模式列表中删除，因为它已经不能用于产生 $(k+1)$ 模式。例如表 4 中的 2-模式“ $F \text{ Before}[0,0,0,0,0] G$ ”只在频繁 3-模式的第 2 行中出现。如果用该 2-模式来扩充频繁 3-模式“(A Before[0,0,0,0,0] D) Before[0,0,0,0,0] F”，就会产生候选模式“(A Before[0,0,0,0,0] D) Before[0,0,0,0,0] F) Before[0,0,0,0,0] G”。由条件 1，它当中的每个子模式必须是频繁的，但它的子模式“(D Before[0,0,0,0,0]F) Before[0,0,0,0,0] G”不是频繁的。因此，即使保留该记数小于 $(k-1)$ 的 2-模式，也无法由它产生有效的候选时态模式。

这种基于时态模式的复合事件检测方法的算法表示如图 6。首先扫描数据库得到所有频繁原子事件(第 1 行)，这些事件被放入频繁集合 FrequentSet 中(第 2 行)。然后调用函数 GetNextCandidateSet^[21]得到一个初始化的 2 级候选集合 CandidateSet(第 3~4 行)，本算法的目标就是从 CandidateSet 中得到频繁时态模式。对事件列表集合 EventListSet 中的每个 EL 都调用 CountSupport^[21]来获取 CandidateSet 中每个时态模式的支持数目(第 6~8 行)。当 EventListSet 中的每个 EL 都被检测过之后，就可以得到频繁模式了(第 9 行)。函数 GetNextCandidateSet 返回下一级的候选时态模式(第 10~11 行)。当 CandidateSet 为空时，算法终止(第 12 行)。

```

1. Scan database and obtain all single frequent events
2. FrequentSet  $\leftarrow$  all single frequent events
3. CandidateSet  $\leftarrow$  GetNextCandidateSet (FrequentSet)
4. Level  $\leftarrow$  2
5. Repeat
6.   For all (EL $\in$ EventListSet) do
7.     CountSupport (Level, EL, CandidateSet)
8.   End for
9.   FrequentSet  $\leftarrow$  obtain frequent patterns
10. CandidateSet  $\leftarrow$  GetNextCandidateSet (FrequentSet)
11. Level  $\leftarrow$  Level +1
12. Until (CandidateSet =  $\emptyset$ )

```

Fig.6 Algorithm of complex event detection based on temporal pattern

图 6 基于时态模式的复合事件检测算法

5 研究展望

随着复合事件检测应用的日益广泛, 在复合事件检测的研究与分析中涌现出了许多有趣的问题。如: 在“事件”方面, 除了要考虑原子事件和复合事件之外, 还要考虑更加复杂的事件, 即考虑事件类型的等级性, 另外, 在复合事件检测的过程中, 还要充分考虑各事件之间的关联性; 在“时间”方面, 需要考虑事件无序性的情况, 即在实际应用中事件并不都是完全有序的; 在“数据”方面, 需要考虑普适计算环境下不可避免的数据不确定性问题, 分析不确定性事件的检测(包括局部不确定性和全局不确定性)。具体说明如下。

5.1 事件的不确定性

现有的复合事件检测研究通常都假设事件是精确的, 然而在许多现实应用中事件都是不确定的。不确定性问题是普适计算环境下的一个本质问题, 也是复合事件检测中的一个主要问题。例如在“智能家居”应用中, 各传感器数据是不确定的, 家居中人的行为模式和习性也是不确定的, 如何根据这些不确定的数据推导出对用户有用的、确定的信息是一个挑战性问题; 在路网应用中, 各车辆和信息载体都是运动的, 它们的位置是不确定的, 获取的信息也可能是

如何有效地处理这些不确定性数据也是很重要的。虽然近几年不确定性数据研究成为一个热点问题, 但在不确定性复合事件检测中还有很多问题有待深入探讨。例如: 由于各种原因产生的传感器数据的不确定性、事件的局部不确定性、事件之间相互关联的全局不确定性等(见第3章的具体分析)。目前的研究工作主要集中在不确定性数据的表示模型以及不确定性数据的查询处理等方面, 但是这些模型和算法在计算代价、查询效率等方面还存在诸多缺陷。今后的研究除了对表示模型和查询处理继续优化之外, 还要在不确定性数据的存储与索引技术, 位置相关的不确定性数据服务, 不确定性数据的分析与挖掘技术等方面进行深入探讨。总之, 随着大量不确定数据的产生, 对不确定的复合事件检测的研究分析将变得越来越重要。不确定性复合事件检测是一个亟待解决的问题。

5.2 事件之间的关联性

现有的复合事件检测研究通常都存在一个强假设, 就是事件流之间是相互独立的, 不同事件之间不会相互关联, 也不会互相影响。但是在现实应用中, 事件之间存在着千丝万缕的联系。例如: 第3章中 Mary 和 Joe 喝咖啡的时间、地点以及持续时间都是相关的; 4.2 节中商品是否误放与商品是否结账是相关的; 4.3 节中患者是否有糖尿病与血糖的增高和尿糖的缺失是否同时并存是相关的, 等等。如果不考虑这些事件之间的关联性, 就会得出如 4.2 节所描述的错误结论。因此在进行复合事件检测时, 必须全面地考虑同一个体不同时刻之间的关系, 以及不同个体之间的相互作用和相互影响, 另外可能还需要考虑此个体的身份职位等因素的影响和关联。在此, 以文献[22]中的一个例子来说明。如图 7(a)和 7(b)分别表示 Joe 在 $T=7$ 和 $T=8$ 时刻的位置, 由图可知其位置是不确定的; 文献[22]中以某位置所含粒子的个数占全部个数的比例来简单地表示对

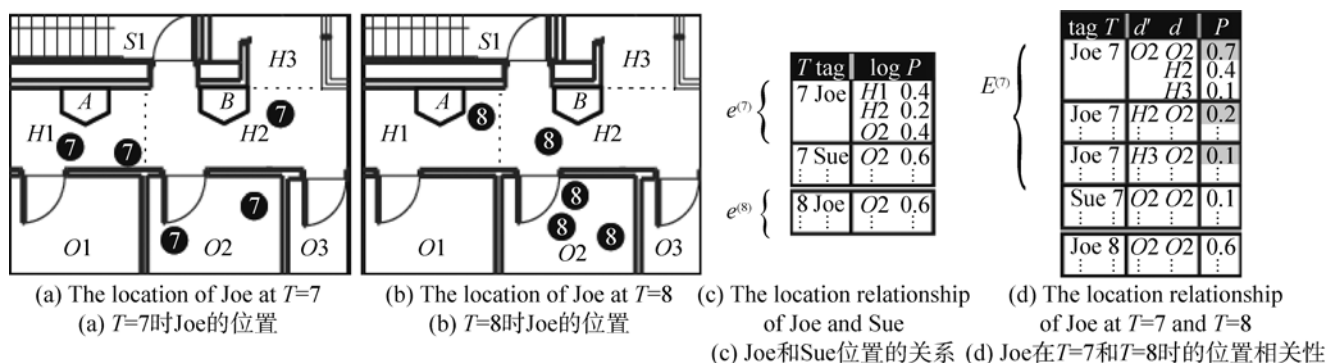


Fig.7 The relationship of events

图 7 事件的关联性

象在该位置的概率；由 7(c)看到，Joe 和 Sue 的位置也是相互关联的，即根据 Joe 的位置的不确定度，可以大致推断出 Sue 的位置。7(c)中 $T=7$ 时刻 Joe 在 $H1$ 和 $O2$ 的概率都是 0.4，但是如果知道 Sue 是 Joe 的秘书，即 Sue 和 Joe 一般是在一起的，那么根据 $T=7$ 时刻 Sue 在 $O2$ 的概率为 0.6，可以推断出 $T=7$ 时刻 Joe 很可能也在 $O2$ 位置。然而包括文献[22]在内的目前的研究工作都没有考虑这些因素的影响和关联；又由 7(d)看到，Joe 在 $T=7$ 和 $T=8$ 时刻的位置是相互关联的，即根据 Joe 在 $T=7$ 时刻的位置的不确定度，可以大致推断出 Joe 在 $T=8$ 时刻的位置。如果 Joe $T=7$ 时刻在 $O2$ 位置，那么 $T=8$ 时刻 Joe 很有可能还在 $O2$ 位置，所以 $\langle O2, O2 \rangle$ 的概率为 0.7，记为 $P\langle O2, O2 \rangle = 0.7$ ，比 $P\langle H2, O2 \rangle = 0.2$ 和 $P\langle H3, O2 \rangle = 0.1$ 大的多。

5.3 复合事件类型的等级性

目前的很多研究工作都是将事件从原子类型转化成复合类型，很少有研究工作是将复合类型的事件进一步转化成更为复杂的复合事件类型。后者的研究是要将前者的输出结果作为输入，因此前者的研究也是进行更为复杂的事件复合过程的重要一步。4.5.1 小节中^[17]提出的复合事件检测方法就无法处理更为复杂的复合事件检测过程，因为其输入事件流仅限于含有时间戳的原子事件，也就是说这种检测方法的输出结果无法作为输入。然而随着现实世界应用的日益广泛，更为复杂的复合事件检测必将越来越重要。例如

第 1 章中提到的健康护理，要想知道被照顾人是否已经被很好地照顾，需要知道被照顾人的一系列行为流，如：“他是否按时吃药了？是否按时吃饭了？在睡觉之前是否刷牙了？体温血压是否正常？”等”。在这个例子中，可以把健康护理的整个流程看作一个更为复杂的复合事件，其中牵涉到的一系列行为既可能是原子事件，也可能是复合事件。如检查“他是否按时吃药”就是由以下原子事件构成的：“倒了一杯水”、“拿起了药瓶”、“喝水”等，因此检查“他是否按时吃药”就是一个复合事件，而检查“体温、血压的示数是否正常”就是原子事件。用图 8 来直观地表示复合事件类型的等级性。今后的研究工作需要考虑这种等级性，根据不同的需求检测出合理的复合事件类型。



Fig.8 The gradability of complex events

图 8 复合事件类型的等级性

5.4 事件的无序性

目前的很多研究都是在进入事件处理系统

之前,为每个事件设定一个时间戳。这些时间戳是离散的、有序的,能够反映这些事件的实际发生顺序,他们通常都假设这些事件是完全有序的。如 4.5.1 小节中^[17]提出的复合事件检测方法就假设输入的原子事件是完全有序的,也就是说不考虑事件的并发性和重叠性。而实际上这种假设并不是在所有场景下都成立,在很多情况下,事件可能是同时发生的。例如:一个复合事件通常从它的原子事件中获取时间戳,当这些由很多原子事件组成的复合事件用来检测更为复杂的复合事件时,事件之间全序性的假设就不成立了。仍以第 1 章提到的健康护理为例,在进行“健康护理”这个更为复杂的复合事件检测的过程中,刷牙、量体温等事件很有可能是同时进行的,并且由于不同人的生活习性不同,此过程中各原子事件或复合事件之间很有可能是完全无序的,因此传统的检测方法就无效了,未来的复合事件检测研究中必须考虑事件的无序性特征。

5.5 事件的分布性

普适计算中的大量设备如相机、汽车、手机、家电等均将具有一定的存储能力,用于收集和存储相应的各种数字信息。其中每个移动设备均是一个数据源,每个数据源能力有限,且不存在集中固定的服务器为各移动数据源提供支持,也就是说许多现实应用(例如监控环境)中包含大量的分布事件源(如硬件传感器和软件接收器等)。但是目前的研究工作通常都假设所有相关的事件是集成的,很少考虑分布式环境中的复合事件检测机制。如果将传统的检测方法应用到分布式环境中,效率会很低,因为它需要检测所有原子事件和处理单元的交互,而实际上组成复合事件的往往只是所有原子事件中的一小部分。例如路网交通安全监控中所获得的原始数据/信息,并不都是用户感兴趣的,有些甚至是完全无用的。因此,如何从众多的分布式事件源中选取有价值的信息,进行有效的查询、监控和分发是普适计算

环境下的基本问题。分布式环境中复合事件检测的目标就是检测尽可能少的原子事件,同时又不遗漏任何用户感兴趣的复合事件,其中要全面考虑原子事件的并发性、异步性、不确定性等因素,因此还需要做大量的研究工作。

6 结论

随着传感器和无线设备的大规模使用,产生了数量巨大的原子事件,因此如何从这些底层的原子事件中抽取人们感兴趣的、有用的复合事件就变得越来越重要。目前研究人员围绕复合事件检测,从时间因素、复合事件表示方法、数据的精确度等方面做了很多研究。从上述 3 个方面,对近几年国际上在该领域的主要研究成果进行了回顾和总结,并提出了仍然存在的问题和今后的研究方向。总的来说,复合事件检测的日益重要和普适计算环境下大量不确定性数据的产生,使得这一研究领域中的关键问题需要进一步深入探索。

References:

- [1] Liao L, Patterson D J, Fox D, et al. Learning and inferring transportation routines[J]. *Artif Intell*, 2007, 171(5/6): 311–331.
- [2] Philipose M, Fishkin K P, Perkowitz M, et al. Inferring activities from interactions with objects[J]. *IEEE Pervasive Computing*, 2004, 3(4): 50–57.
- [3] Demers A J, Gehrke J, Hong M, et al. Towards expressive publish/subscribe system[C]//*Proceedings of Advances in Database Technology Conference*, 2006: 627–644.
- [4] Jobst D, Preissler G. Mapping clouds of SOA-and business-related events for an enterprise cockpit in a Java-based environment[C]//*Proceedings of the 4th ACM International Conference on Principles and Practices of Programming In Java(PPPJ)*, 2006: 230–236.
- [5] Wang F, Liu P. Temporal management of RFID data[C]//*Proceedings of the 31st VLDB Conference*, September

- 2005.
- [6] Das S K. The role of prediction algorithms in the MavHome smart home architecture[J]. IEEE Wireless Communications, 2002, 9(6):77–84.
- [7] Lamming M, Bohm D. SPECS: Another approach to human context and activity sensing research, using tiny peer-to-peer wireless computers[C]//Proceedings of the 5th International Conference on Ubiquitous Computing, 2003, 2864: 192–199.
- [8] McCarthy J F, Anagnost T D. Event manager: Support for the peripheral awareness of events[D]. Hebrew Union College, 2000: 227–235.
- [9] Virone G, Wood A, Selavo L, et al. An assisted living oriented information system based on a residential wireless sensor network[C]//Proceedings of the 1st Distributed Diagnosis and Home Healthcare (D2H2) Conference, April 2006.
- [10] Mert A, Ugur G, Nesime T. Plan-based complex event detection across distributed sources[J]. The VLDB Endowment, 2008, 1(1): 66–77.
- [11] Samani M, Sloman M, Gen M. A generalized event monitoring language for distributed systems[J]. IEEE/IOP/BCS Distributed Systems Engineering, 1997, 4(2): 96–108.
- [12] Chakravarthy S, Rasad V K, Anwar E. Anatomy of a composite event detector, Technical Report UF–CIS–TR–93–039[R]. Gainesville: University of Florida, 1993.
- [13] Gehani N H, Jagadish H V, Shmueli O. Event specification in an active object-oriented database[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1992: 81–90.
- [14] Gatziau, S, Dittrich K R. Events in an active object-oriented database system[C]//Proceeding of the 1st Int'l Conference on Rules in Database Systems, 1993: 23–39.
- [15] Wu E, Diao Y, Rizvi S. High-performance complex event processing over streams[C]//Proceedings of the 2006 ACM SIGMOD, New York, USA, 2006: 407–418.
- [16] Jagrati A, Diao Y L, Daniel G, et al. Efficient pattern matching over event streams[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, Canada, 2008: 147–160.
- [17] Eugene W, Diao Y L, Shariq R. High-performance complex event processing over streams[C]//Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, Chicago, IL, USA, 2006: 407–418.
- [18] Kam P S, Ada W F. Discovering temporal patterns for interval-based events[C]//Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery, 2000.
- [19] Panagiotis P, George K, Stan S, et al. Discovering frequent arrangements of temporal intervals[C]//Proceedings of the 5th IEEE International Conference on Data Mining, 2005: 354–361
- [20] Wu S Y, Chen Y L. Mining nonambiguous temporal patterns for interval-based events[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(6): 742–758.
- [21] Dhaval P, Wynne H, Mong L L. Mining relationships among interval-based events for classification[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, Canada, 2008: 393–404.
- [22] Christopher R, Julie L, Magdalena B, et al. Event queries on correlated probabilistic streams[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, Canada, 2008: 715–728.
- [23] Zimmer D, Unland R. On the semantics of complex events in active database management systems[C]//Proceedings of the 15th International Conference on Data Engineering (ICDE), 23-26 March, 1999.
- [24] Antunes C, Oliveira A L. Generalization of pattern-growth methods for sequential pattern mining with gap constraint[C]//Proceedings of the Int'l Conference Machine Learning and Data Mining (MLDM'03), 2003.
- [25] Pei J, Han J, Mortazavi A B, et al. Prefixspan: Mining sequential patterns ϵ efficiently by prefix-projected pattern growth[C]//Proceedings of the 17th International Conference on Data Engineering (ICDE), April 2001.
- [26] Mannila H, Toivonen H, Verkamo I. Discovery of frequent episodes in event sequences[C]//Proceedings of the

- Annual ACM SIGKDD Conference, 1995.
- [27] Agrawal R, Srikant R. Mining sequential patterns[C]// Proceedings of the IEEE International Conference on Data Engineering (ICDE), 1995.
- [28] Allen J F. Maintaining knowledge about temporal intervals[J]. Communications of the ACM, 1983, 26(11).
- [29] Hakeem A, Sheikh Y, Shah M. A hierarchical event representation for the analysis of videos[C]//Proceedings of the 19th National Conference on Artificial Intelligence (AAAI 2004), San Jose, California, July 25-26, 2004.
- [30] Zhao T, Nevatia R, Hongeng S. Hierarchical language-based representation of events in video streams[C]//Proceedings of the IEEE Workshop on Event Mining, 2003.
- [31] Jeffery S. Adaptive cleaning for RFID data streams[C]// Proceedings of the 32nd VLDB Conference, September 2006.
- [32] Floerkemeier C, Lampe M. Issues with RFID usage in ubiquitous computing applications[C]//Proceedings of the 2nd Pervasive Conference, April 2004.
- [33] Adaikkalavan R, Chakravarthy S. Snooipib: Interval-based event specification and detection for active databases[J]. Data Knowledge Engineering, 2006, 59(1): 139–165.
- [34] Levinson S E, Rabiner L R, Sondhi M M. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition[J]. Journal of Bell System Technology, 1983, 62(4).
- [35] Kanagal B, Deshpande A. Online filtering, smoothing and probabilistic modeling of streaming data, Technical Report CS-TR-4867[R]. Maryland: University of Maryland, 2007-05.



ZHOU Chunjie was born in 1981. She is a Ph.D. candidate at Renmin University of China. Her research interests include pervasive computing, mobile data management and workflow management, etc.

周春姐(1981–), 女, 山东烟台人, 中国人民大学博士研究生, 主要研究领域为普适计算, 移动数据管理和 workflow 管理等。



MENG Xiaofeng was born in 1964. He is a professor and doctoral supervisor at Renmin University of China, and the member of CCF. His research interests include Web data management, XML database and mobile data management, etc.

孟小峰(1964–), 男, 河北邯郸人, 中国人民大学教授、博士生导师, CCF 会员, 主要研究领域为 Web 数据管理, XML 数据库, 移动数据管理等。