

REvent: 一种面向 RFID 应用的复杂事件处理系统

谷 峪^{1,2} 胡小龙¹ 李传文^{1,2} 于 戈^{1,2}

¹(东北大学信息科学与工程学院 沈阳 110004)

²(医学影像计算教育部重点实验室(东北大学) 沈阳 110004)
(guyu@ise.neu.edu.cn)

REvent: An RFID-Application-Oriented Complex Event Processing System

Gu Yu^{1,2}, Hu Xiaolong¹, Li Chuanwen^{1,2}, and Yu Ge^{1,2}

¹(College of Information Science and Engineering, Northeastern University, Shenyang 110004)

²(Key Laboratory of Medical Image Computing(Northeastern University), Ministry of Education, Shenyang 110004)

Abstract RFID technology is expected to be applied in a lot of application scenarios. In this study, an RFID application oriented complex event processing system, called REvent, is implemented. A simulation environment is built to test the data preprocessing and event detection functions of the REvent system. The system architecture, key technologies and major demonstration process are illustrated.

Key words RFID; system demonstration; complex event processing; data preprocessing; event detection

摘 要 无线射频识别技术 RFID 有着重要的应用背景. 设计并实现了一个面向无线射频识别应用的复杂事件处理系统 REvent. 构建了一个仿真环境并在此基础上测试 REvent 系统的数据预处理与事件检测等功能. 对系统框架、关键技术和主要演示过程进行了说明.

关键词 无线射频识别; 系统演示; 复杂事件处理; 数据预处理; 事件检测

中图法分类号 TP311

无线射频识别(radio frequency identification, RFID)技术, 是利用无线射频通信实现的非接触式自动识别技术, 通过无线射频方式在阅读器和标签之间进行非接触双向数据传输, 达到目标识别和数据交换的目的. 与传统的条形码或磁条识别技术相比, 无线射频识别技术具有非接触、作用距离远、精度高、适用于恶劣环境及可识别运动目标等一系列的优点, 因而在工业自动化、商品控制、交通运输控制管理等众多领域得到广泛的发展^[1].

RFID 技术的广泛应用带来了数据处理和数据管理的挑战. 例如 RFID 采集到的数据包包含了很多隐含的信息, 需要进行转换和集成, 转变成具有语义的数据, 然后用相应的数据模型表示出来. 而且由于 RFID 数据是瞬时的、流动的、大量的, 所以

需要被高速地处理. Siemens 公司的 RFID 中间件系统^[2]采用传统的数据库技术, 对 RFID 历史数据进行建模; SASE 和 Cayuga 等通用的复杂事件处理系统基于数据源对事件进行建模和检测, 可以用于简单的 RFID 场景^[3-4]. 同时, RFID 数据还存在漏读和乱序的问题, 需要进行数据填充和乱序处理^[5-6]等等. 因此, 一个成熟的 RFID 数据处理系统必须解决上述问题, 将 RFID 系统接收到的数据转换为与商务应用相对应的形式.

本文设计实现了一个复杂事件处理系统 REvent, 构建了一个仿真环境并在此环境基础上进行数据预处理与事件检测. 系统的总体框架包括 5 大功能模块: 场景仿真、查询接口、数据预处理、事件检测、结果显示. 在充分考虑 RFID 数据和应用特性的基础

上,REvent 集成了数据清洗系统、以数据为中心的检测系统和以事件为中心的检测系统的各种功能,对查询处理过程进行了优化,实现了完整且高效的 RFID 事件检测流程。

1 系统框架

1.1 系统结构

本节介绍 RFID 复杂事件处理系统的系统框架。图 1 给出了 REvent 的系统框架图,REvent 系统从功能上可以分为 5 大模块:场景仿真、查询接口、事件预处理、事件检测、结果显示。

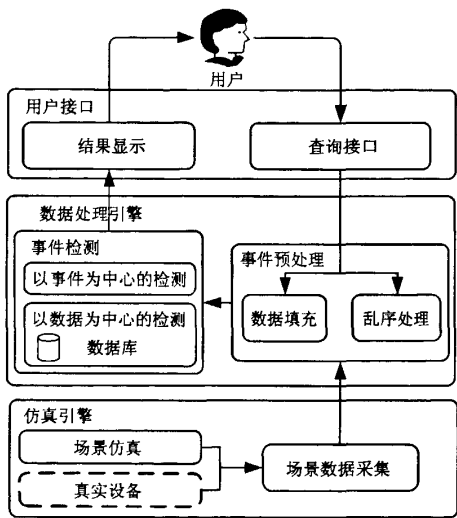


图 1 系统结构图

1. 场景仿真模块

本系统以公园背景为例进行仿真。假设公园中有 9 个景区,每个景区之间都有一条路径,在景区与围墙之间也有一条道路,两条道路相交的地方就是一个交叉路口,游客从最左上角进入到公园,然后向右或者向下沿着道路行走,等走到最右下角时,游客就走出公园,用户可以在公园的任何地方都可以设置阅读器。仿真场景中存在 2 个实体:游客实体及阅读器实体。特别的,针对真实设备的数据和应用接口也已经在开发和集成当中。

2. 查询接口

查询接口模块主要为用户提供查询事件的功能。对于面向数据库的查询,我们提供了对历史事件检测的 SQL 接口。特别地,对于在线的事件检测,本系统提供 2 种查询事件输入方式:1)采用类 SQL 语言方式进行查询,例如 $SEQ(E_1, E_2)$ 表示查询先经

过 E_1 再经过 E_2 的事件;2)用户通过图形操作建立自动机模型,并进行查询。用户通过图形操作输入的查询在系统内部被自动转换为类 SQL 语言查询。

3. 事件预处理模块

事件预处理模块主要负责解决数据采集及传输过程中产生的数据丢失和事件乱序问题。本系统采用多角度的数据填补技术解决数据丢失问题,主要通过数据缓存方式解决事件乱序问题。

4. 事件检测模块

事件检测模块实现了 2 种数据事件处理的方法,一种是以数据为中心即基于数据库技术的方法,采用基于中国剩余定理的技术将数据存储在数据库中进行高效的 adhoc 查询;另一种是以事件为中心的检测方法,采用扩展自动机模型对复杂事件进行连续的在线检测。通过某个开启事件来启动自动机模型,自动机会根据到达的基本事件进行状态跃迁,当自动机进入接受状态时,则说明复合事件匹配成功。

5. 结果显示模块

结果显示模块显示事件检测结果以及各阅读器的详细探测信息。事件结果的显示中主要是对事件检测部分发生的事件进行显示,其中包括物体的 ID,事件发生的时间以及物体的整个移动轨迹。阅读器的详细探测信息中包括阅读器物体的 ID,探测到 ID 的时间等。

1.2 系统执行流程

图 2 示例了系统的工作流程。首先由启动仿真过程,可以设置阅读器的节点,包括添加和删除阅读器节点,设置阅读器重的属性,然后再设置系统产生数据的运行速度,开始仿真运行。产生的数据有 2 个选择,一是存入数据库,以备后续查询;二是设置正确的复杂事件,对数据进行复杂事件检测。对于存入数据库中的数据可以用 SQL 语言进行编辑查询。

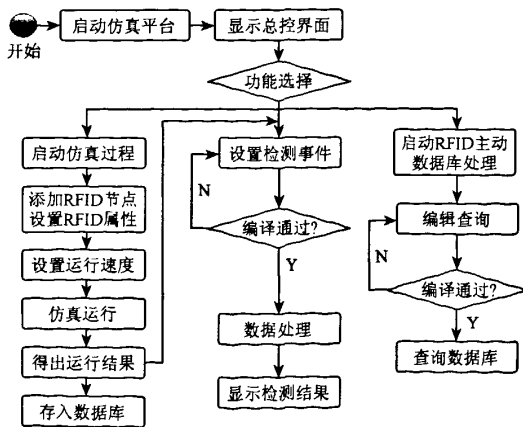


图 2 系统运行流程图

2 关键技术

2.1 数据预处理技术

RFID 事件检测中要求要检测的数据流必须满足 2 个条件:1. 完整性,即数据不能缺少;2. 有序性,即先读到的数据应该先到达. 如果由于某些原因使到达主机的数据中有丢失的数据或者无序到达,那么 RFID 事件检测就无法有效地检测出事件,针对这 2 个问题,本节介绍了 REvent 采用的处理方法.

1. 数据丢失

数据丢失指的是采集到的数据没有到达主机. 本系统采用多角度的数据填补技术对漏读的数据进行判断和填充. 假设再通过阅读器 A 后的下一个阅读器是 B,我们就称 A 和 B 之间具有直达性. 用有向无环图来存储这种直达性,如图 3 所示. 图中分两部分,一部分是节点模型,代表阅读器中所属的事件类型,标识的概率表示统计出的漏读率;另一部分是箭头连接,代表 2 个阅读器之间具有可达性,标识的概率表示选择相应路径的统计概率.

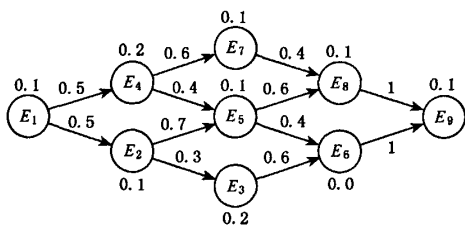


图 3 数据填补图

采用有向无环图进行数据填补的基本原理是: 每一个移动的物体都有一个有向无环图用来做数据填补,当移动物体到达某个阅读器的时候,就查询上一个阅读器是否到达,如果上一个或者上几个阅读器没有标识的话,那么就根据节点漏读率和路径选择概率计算最可能漏读的节点并进行填补. 此外我们还结合单个物体的时序数据特征和多个物体构成的小组空间相关性,来提高填补的准确率.

2. 数据乱序

实际应用中,阅读器会被分布在不同的地方. 由于网络传输延迟或者机器故障,阅读器采集的数据无法按时到达,就会导致事件发生的顺序与到达的顺序不一致,即产生了乱序现象^[6].

本系统将无序的事件按时间戳顺序排序. 首先开辟一段缓存用来存储事件并按时间戳排序,待时间窗口内的所有事件都达到有序的时候再进行输出. 为了确定缓存的时间以及缓存区的大小,我们假

设乱序事件至多延时 T 秒后到达,即把 T 值作为乱序的上限. 在实际应用中如果某些无序事件的延迟高于上限 T ,则事件将被丢弃. T 值的设定会影响到事件检测的准确率和检测性能. 如果 T 值设定比较小,丢掉的数据会增多,则检测的准确率将会降低,而 T 值太大造成事件缓存的时间太长,则会降低查询的性能. 此外,我们还采用了基于后缀自动机的乱序匹配技术对乱序事件进行及时的处理.

2.2 事件检测技术

本节主要介绍以事件为中心的检测方法. 对事件进行即时处理可以实时检测到事件的发生,并且不需要过多硬盘空间来存储数据. REvent 采用自定义的扩展自动机模型来进行复杂事件检测.

组成复杂事件的基本事件中,某个基本事件的发生标志着复杂事件的发生的开始,我们称这样的事件为开启事件. 很显然,一个复杂事件可以有很多个开启事件. 比如:复杂事件 $OR(E_1, E_2)$, 其中开启事件就是 E_1 和 E_2 . 组成复杂事件的基本事件中,某个基本事件的发生标志着复杂事件的结束,我们称这样的事件为结束事件.

本系统中自动机模型的主要原理就是通过某个开启事件来启动查询自动机,当某个参与复合事件的基本事件到达时,自动机就会从一个状态跃迁到下一个状态,当自动机进入某个可接受的状态(结束事件)时,则说明复合事件发生了.

以图 4 为例介绍本系统中的自动机模型,该自动机对应查询 $SEQ(E_1, AND(OR(E_2, E_3), E_4), NOT(SEQ(E_5, E_6)), OR(E_7, E_8))$. 图 4 中上部通过箭头连接 SEQ 操作的一系列节点(E_1, E_2, \dots, E_8 等)以及相关操作(OR, AND 等). 图 4 中 OR 操作和 AND 操作分别连接到查询语句中的每个操作的对应节点. 由于 NOT 操作的特殊性, NOT 操作及其对应节点需要新建一个自动机并进行独立检测,我们对此进行了特殊的标识和处理. 具体地,对于每个操作,REvent 都定义了高效的处理方法,并针对各个操作之间的关联性和多查询的共享性进行了优化.

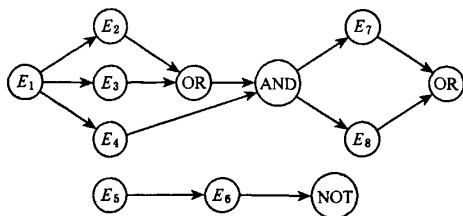


图 4 扩展自动机模型

3 系统演示

图 5 示例了查询接口模块的图形输入界面, Palette 中各个事件的节点以及连接都可以拖到的编辑界面中,在仿真运行之前在编辑界面中进行建立要查询的复杂事件,生成有限自动机,进行事件查询.在图 5 中我们要查询的事件为 $SEQ(E_1, OR(E_2, E_4), E_5)$, 即查询经过 E_1, E_2, E_5 的事件或者 E_1, E_4, E_5 的事件.

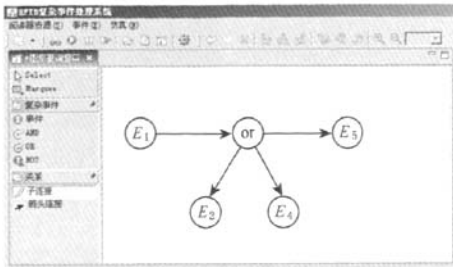


图 5 事件查询接口

图 6 展示了系统在仿真运行时的界面元素以及上图查询中的事件结果. 其中:

1. Palette 中的阅读器可以拖动到当前环境中, 可以添加或者删除 RFID 节点.
2. 时间控制视图里面可以控制仿真时间快慢, 时间间隔表示仿真环境一个单位时间相当于真实环境中的毫秒数. 图 6 中一个时间单位等于 50 ms.
3. 属性视图显示当前选中阅读器的各种属性, 包括阅读器的 ID 号、探测半径、漏读率、周期以及位置坐标. 我们选中图中的一个节点, 其 ID 号为 1, 探测半径为 50, 漏读率为 0.0, 探测周期为 1 个时间单位, 位置坐标为 (45, 45).
4. 场景仿真视图主要仿真游客在公园移动的

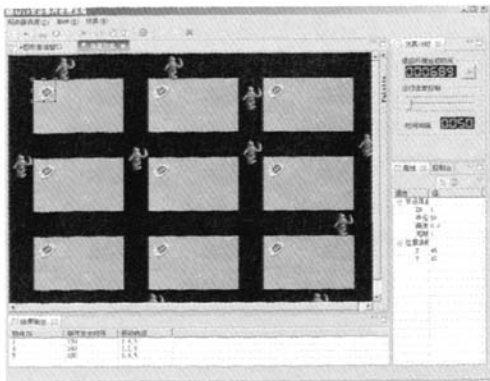


图 6 REvent 系统运行界面

过程,负责产生数据.

5. 结果输出视图主要是对复杂事件检测的结果进行显示. 图中显示了已经发生的事件, 物体 ID 号分别为 1, 4 和 5, 它们的移动路径满足我们要查询的事件.

4 结 论

本文介绍了我们设计和实现的一个复杂事件处理系统 REvent, 该系统构建了一个仿真环境并在此环境基础上进行数据预处理与事件检测. 系统的总体框架包括 5 大功能模块: 场景仿真、查询接口、数据预处理、事件检测和结果显示. 目前, 我们已经开发了读取多种真实 RFID 阅读器的可视接口, 下一步将会把 REvent 扩展到真实的 RFID 应用场景中.

参 考 文 献

- [1] 谷 峪, 于戈, 张天成. RFID 复杂事件处理技术. 计算机科学与探索, 2007, 1(3): 255-267
- [2] Wang F S, Liu P Y. Temporal management of RFID data // Proc of the 31st Int Conf on Very Large Data Bases. New York: ACM, 2005: 1128-1139
- [3] Wu E, Diao Y, Rizvi S. High-performance complex event processing over streams // Proc of 2006 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2006: 407-418
- [4] Brenna L, Demers A, Gehrke J, et al. Cayuga: A high-performance event processing engine // Proc of 2007 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2007: 1100-1102
- [5] Jeffery S R, Garofalakis M, Franklin M J. Adaptive cleaning for RFID data streams // Proc of the 32nd Int Conf on Very Large Data Bases. New York: ACM, 2006: 163-174
- [6] Liu M, Li M, Golovnya D, et al. Sequence pattern query processing over out-of-order event streams // Proc of the 25th Int Conf on Data Engineering. Los Alamitos: IEEE Computer Society, 2009: 784-795

谷 峪 男, 1981 年生, 副教授, 主要研究方向为 RFID 数据管理、时空数据管理、数据流.

胡小龙 男, 1985 年生, 硕士, 主要研究方向为 RFID 数据管理.

李传文 男, 1982 年生, 博士研究生, 主要研究方向为时空数据管理.

于 戈 男, 1962 年生, 教授, 博士生导师, 主要研究方向为数据库理论与技术.