|  |  |
| --- | --- |
| **分类号—————————————————— 密级 —** | 校徽 |
| **UDC** |
|  |

**本 科 毕 业 论 文（设计）**

传感器网络与RFID技术的数据处理研究

**学生姓名 学号**

**指导教师**

**院、系、中心**

**专业年级**

**论文答辩日期 年 月 日**

**中 国 海 洋 大 学**

传感器网络与RFID技术的数据处理研究

完成日期：

指导教师签字：

答辩小组成员签字：

**传感器网络与RFID技术的数据处理研究**

**摘 要**

（题目、摘要、关键词由ChatGPT总结生成）本文研究了在传感器网络和RFID技术应用中面临的不确定数据管理、复杂事件处理和隐私保护问题。随着传感器和RFID技术的广泛应用，传感器网络和RFID系统产生了大量数据，其中包括不确定性数据。我们首先介绍了不确定数据的产生原因，如冲突读、漏读和粒度不匹配，并强调了不确定数据管理在各领域中的重要性。

在复杂事件处理方面，我们研究了如何将原子事件组合成有意义的复合事件，以挖掘潜在信息。我们以超市偷盗监控系统为例，说明了复合事件处理在实际应用中的重要性。此外，我们还提出了针对不确定数据的相关性概率流查询处理问题，解决了数据相关性问题，进一步提高了数据处理效率。

隐私保护方面，我们介绍了隐私问题的存在，并提出了基于事件类型级的隐私保护方法，同时提高了数据的效用增益值。最后，我们总结了本文的研究贡献，包括复合事件查询处理方法、相似性查询处理方法和隐私保护方法的提出。

**关键词：不确定数据；相关性概率流；隐私保护；传感器网络；RFID技术**

**Research on Sensor Networks and RFID Data Management**

**Abstract**

This research addresses the challenges of uncertain data management, complex event processing, and privacy protection in the context of sensor networks and RFID technology. With the widespread use of sensor networks and RFID systems, there is a growing need to manage and process large volumes of data, some of which are inherently uncertain.

The study begins by elucidating the origins of uncertain data, stemming from factors such as collision readings, missed readings, and granularity mismatches. The importance of effective uncertain data management across various domains is underscored.

In the realm of complex event processing, we delve into the amalgamation of atomic events into meaningful composite events, highlighting their pivotal role in unearthing latent insights. Through practical examples, such as theft detection in a supermarket surveillance system, we emphasize the significance of complex event processing in real-world applications.

Furthermore, this research introduces a novel approach to address the issue of data relatedness by proposing a correlation probability flow query processing method, thus enhancing data processing efficiency.

Regarding privacy protection, the paper discusses privacy concerns and presents a privacy protection strategy based on event type-level measures, concurrently maximizing data utility gain.

In conclusion, this study contributes innovative techniques and privacy protection solutions to enhance data processing and security in sensor networks and RFID technology.

**Keywords: Uncertain data; correlation probability flow; privacy protection; sensor networks; RFID technolog**

**目 录**

**1 引言**

1.1 研究背景 ………………………………………………………………………………… 1

1.1.1 不确定数据 …………………………………………………………………………… 1

1.1.2 复杂事件处理 ………………………………………………………………………… 3

1.2 问题提出 ………………………………………………………………………………… 4

1.3 本文贡献 ………………………………………………………………………………… 4

1 引言

本章首先从不确定数据、复杂事件处理和隐私保护三个方面介绍本文的研究背景，然后总结目前针对不确定数据管理、复杂事件处理和隐私保护方面的工作，指出不足，针对其不足和挑战，提出本文需要解决的问题，然后总结本文的贡献点，最后指出本文的结构安排。

1.1 研究背景

近年来，传感器网络和射频识别技术RFID广泛的应用在供应链管理系统、对象跟踪和监控系统等领域。以射频识别技术为例，它通过射频信号自动识别目标物体并获取目标物体的相关信息。RFID系统主要由标签、阅读器和应用程序三部分组成。阅读器在其有效的范围内获取到标签的数据信息，形成RFID数据。RFID数据观察值的基本形式为三元组：<标签ID，阅读器ID，时间戳T>，其中标签具有全球统一的ID，阅读器ID可以理解成标签ID所在的位置。其表示的实际意义为标签ID的物体在时间T出现在阅读器ID所在的位置。而由于RFID数据具有海量性、关联性以及时序性等特点，使针对该类数据的复杂事件处理技术得到广泛的研究。并且，由于数据能够表示某个个体ID的一些信息，通过某种查询，可以使个体的一些敏感信息泄露，所以隐私保护技术也得到了广泛的关注。

1.1.1 不确定数据

在传感器应用中，由于硬件设备限制，网络延迟、通讯质量等限制，会产生缺失的数据或不精确的数据。同样，在射频识别系统中，由于射频识别信号之间存在一定的干扰和阅读器读取数据的本身不可靠性等因素，造成其数据具有一定的不确定性。RFID数据的不确定性主要由三方面的原因造成：(1)冲突读，例如某个对象所携带的标签被两个阅读器读到，造成了其真实位置的不确定。(2)漏读，阅读器仅仅能探测其周围60%-70%的标签，所以有很大一部分标签不会被阅读器探测到。(3)粒度不匹配，例如查询中需要标签所在的位置，而实际上只能提供阅读器信息，而同一个阅读器可能覆盖多个位置。如图1.1所示，在该场景中，Joe携带RFID标签，在时刻6时，阅读器A探测到了Joe携带的标签，此时他在走廊，而在时刻7，有可能由于阅读器漏读等原因，不能确定Joe的位置，根据时刻6的位置，此时Joe可能沿着走廊继续向前走，或沿着楼梯下楼，也有可能他进入了办公室O2，无法确定其真实位置，处理该种情况的最简单的方法是认为前后事件相互独立，以均等的概率位于其可能出现的位置，即有33.3%的概率位于位置H2，有33.3%的概率位于位置O2，有33.3%的概率其正在下楼，产生了不确定数据[1]。在许多现实应用中，例如军事、海洋、物流、气象等领域，都会产生大量的不确定数据，不确定数据无处不在，所以针对不确定数据的研究成为热点。



图1 不确定数据产生实例

Fig. 1 The example of uncertain data generation

1.1.2 复杂事件处理

传感器网络和RFID应用中，产生了大量的数据，每一条RFID记录数据可以看作一个原子事件。RFID原子事件的语义十分有限。通常，人们对原子事件并不感兴趣，其并不能够提供有用的信息。实际应用中，通过对原子事件进行组合，包括序列，非，或等操作，形成复合事件，使其更有意义，对其进行深入处理和分析，可以挖掘出一些隐含信息，有效地支持事件监控、事件预警等先进应用。例如，物体ID在货架被读到，在收银台被读到，在出口被读到，其中的每一个都为原子事件，但并不能从中获取有用的信息，但是通过将这三个原子事件进行操作，包括序列，非，或等，就会获得有价值的信息，如图1.2所示的超市偷盗监控系统中，若物体ID首先在货架被读到，然后没有经过收银台，而直接到达出口，则说明该物体是被偷窃[2]。并且随着复杂事件处理技术的发展，其也广泛的应用在web事件流分析以及股票数据分析等领域中。



1.2 问题提出

当前，大多数关于不确定数据管理技术都是假设不确定数据之间是独立的，即任意两个随机变量X和Y，满足一定约束。然而，这种假设在很多实际应用中是不合理的。例如在图1.1中，如果我们确定在时间7时，Joe是在办公室O2，那么在时刻8他更有可能呆在O2，Joe在t+1时刻的位置和在t时刻的位置有关[1]。由于相关性概率流更符合实际应用需求，基于相关性概率流的查询也越来越多。下面列举出几种针对相关性概率的查询处理问题：

在医院的监控系统中个，在固定地点包括病房、医生办公室、洗手间等门口安置阅读器，每一个医生及病人携带RFID标签，在这个过程中，由于阅读器的冲突读、漏读、粒度不匹配等原因，产生了大量的不确定的RFID数据，而不同时间的不确定数据之间具有一定的相关性，上一时刻所在的位置，对下一时刻的位置有很大的影响。

表3.7 状态转移向量的更新算法

Table 3.7 The updating of STV Algorithm

|  |
| --- |
| 算法3.7 update\_buf(Buffer[i],ei) |
| 输入：Ei对应的缓冲区Buffer[i]，输入事件ei  输出：更新缓冲区中的STV变量 |
| 1. For each STV in buffer[i] Do 2. For each (EventType,p) in STV.Eventvector Do 3. For each entry (d,d',cp) in ei.CT Do 4. IF (d==EventType) Then 5. IF(d' not exist in ev) Then 6. insert (d',p\*cp) into ev\_map; 7. ELSE 8. update ev[d']+= p\*cp; 9. STV.Eventvector=ev\_map; |

(1)在监控系统中，有时需要判断哪些医务人员属于同一部门，那么可以根据该部门职务的特点，指定查询。例如想获取职位为外科主治医师的人员，因为外科主治医师需要给病人做手术，那么只需查询先后去过病人病房，然后进入消毒间，接着进入外科手术室的人员，就可以判断其为外科主治医师。

(2)同样在医院的监控系统中，在体检中心，想知道哪些病人是一起来的，进而找出其生活环境的共性，判断引发疾病的原因。我们可以根据一个病人的轨迹数据，对其他数据进行相似性搜索，找到和他在一定时间范围内先后进入相同科室的病人，可以根据结果推测出这些病人是伙伴。

在上面列举了一些在不确定数据上需要进行的一些查询，在我们生活中还有很多类似的查询。这些查询可以对应归结为面向Markov链复合事件检测问题，面向Markov链复合事件相似性检测问题和面向Markov链的隐私保护问题。

因此，面向Markov链的复合事件处理及隐私保护问题是亟待解决的问题。

1.3 本文贡献

在本文中，在分析和总结复杂事件处理、相似性查询、隐私保护的已有工作的基础上，针对Markov相关性事件数据的查询处理进行深入的研究。具体贡献点如下：

(1)针对Markov链的复合事件查询，提出了序列优先查询处理和相关性查询处理方法。利用事件类型、概率阈值对数据进行剪枝，减少处理数据的代价。

(2)针对Markov链的复合事件相似性查询，从概率阈值，相似性距离阈值、事件类型三方面对数据进行剪枝，提出了序列优先的相似性查询和相关性优先的相似性查询处理方法。

(3)针对Markov链的复合事件隐私保护，提出了效用增益的计算方法，利用整数线性规划方法，从事件类型级进行隐私保护的处理，为了提高其准确性，进一步，提出了基于事件实例的舍弃策略，综合考虑过去及将来到达的事件，在保护隐私信息的基础上，提高了效用增益值。

图1.3为本文研究的路线图，从常规复合事件检测、相似性复合事件检测以及隐私保护三个方面分别阐述了本文的研究方案。



图2 本文研究的路线图

Fig.2 Research roadmap of our thesis