传感器网络中的数据管理

纪德文[1]，王晓东[1]
（1.国防科学技术大学计算机学院 长沙市410073）

摘　要： 无线传感器网络集成了传感器技术、无线通信技术、嵌入式计算技术、分布式信息管理技术和数据库技术，成为当前学术界和工业界关注的热点。传感器网络以数据为中心的特点，使得感知数据管理成为网络应用的核心技术。介绍了数据管理的概念和特点，分析了数据管理的研究内容和存在问题，并着重对当前国内国外已有的研究进展状况进行了总结和归纳。

关键词： 无线传感器网络；数据管理；数据操作。
中图分类号： TP393　　**文献标识码：** A

Data Management in Sensor Networks

Dewen Ji[1], Xiaodong Wang[1]
(1. Collage of Computer Science, National University of Defense Technology, Changsha City 410073.)

Abstract: Wireless sensor networks are the integration of sensor techniques, wireless communication techniques, nested computation techniques, distributed computation techniques and database techniques, and have become popular in industry and academe. In a data-centric sensor networks, data management is the critical technique to the network application. This paper introduced the concepts and characters of data management, analyzed the problems existing in the data management, and mainly summarized the research progress.
Keywords: Wireless sensor networks, data management, data operation.

1　引言

作为新兴的多学科交叉研究领域，无线传感器网络综合了传感器技术、嵌入式计算技术、分布式信息处理技术和通信技术，可以使人们在任何时间、地点和任何环境条件下获取大量详实而可靠的信息。传感器网络以数据为中心的特点，使得感知数据的管理和处理成为传感器网络的核心技术。本文就传感器网络中数据管理相关内容的研究与进展进行了介绍。

2　传感器网络中的数据管理

2.1　以数据为中心的传感器网络

传感器网络中，各个分布的节点通过监测周围环境不断产生大量的感知数据。而传感器节点一般比较简单，存在通信能力低、计算存储能力低、能量受限等特点，无法像传统的分布式数据库那样管理数据。如何存储、传输和访问这些数据，成为制约传感网应用的关键。

对于用户来说，传感器网络的核心是感知数据，而不是网络硬件。用户感兴趣的是传感器产生的数据，而不是传感器本身。用户经常会提出如下的查询：“网络覆盖区域中哪些地区出现毒气”，“某个区域的温度是多少”，而不是：“如何建立从A节点到B节点的连接”，“第27号传感器的温度是多少”。

综上所述，传感器网络是一种以数据为中心的网络，不同于以传输数据为目的的通信网络。对数据的管理和操作，成为传感器网络的核心技术。

2.2　数据管理的概念

以数据为中心的传感器网络，其基本思想是把传感器视为感知数据流或感知数据源，把传感器网络视为感知数据空间或感知数据库，把数据管理和处理作为网络的应用目标。

数据管理主要包括对感知数据的获取、存储、查询、挖掘和操作，目的就是把传感器网络上数据的逻辑视图和网络的物理实现分离开来，使用户和应用程序只需关心查询的逻辑结构，而无需关心传感器网络的实现细节。

对数据的管理贯穿于传感器网络设计的各个层面，从传感器节点设计到网络层路由协议实现以及应用层数据处理，必须把数据管理技术和传感器网络技术结合起来，才能实现一个高效率的传感网，它不同于传统网络采用分而治之的策略。

3　数据管理研究的内容

在传感器网络中进行数据管理，有以下几个方面问题：

（1）感知数据如何真实反映物理世界；
（2）节点产生的大量感知数据如何存放；
（3）查询请求如何通过路由到达目标节点；
（4）查询结果存在大量冗余数据，如何进行数据融合；
（5）如何表示查询，并进行优化。

因而，传感器网络中的数据管理研究内容主要包括数据获取

技术、存储技术、查询处理技术、分析挖掘技术以及数据管理系统的研究[1]。

数据获取技术主要涉及传感器网络和感知数据模型、元数据管理技术、传感器数据处理策略、面向应用的感知数据管理技术。

数据存储技术主要涉及数据存储策略、存取方法和索引技术。

数据查询技术主要包括查询语言、数据融合方法、查询优化技术和数据查询分布式处理技术。

数据分析挖掘技术主要包括OLAP分析处理技术、统计分析技术、相关规则等传统类型知识挖掘、与感知数据相关的新知识模型及其挖掘技术、数据分布式挖掘技术。

数据管理系统主要包括数据管理系统的体系结构和数据管理系统的实现技术。

从2002年开始，国际上关于传感器网络中数据管理的研究，就有研究结果发表，主要集中在把数据库技术，尤其是分布式数据库技术和传感器网络技术相结合，实现数据管理。

4 基于感知数据模型的数据获取技术

4.1 问题描述

在传感器网络中对数据进行建模，主要用于解决以下四个问题：

第一，感知数据具有不确定性。节点产生的测量值由于存在误差并不能真实反映物理世界，而是分布在真值附近的某个范围内，这种分布可用连续概率分布函数来描述。传统文献中讨论的数据模型技术大多采用离散概率分布函数，并不能很好地适用于传感器网络。

第二，利用感知数据的空间相关性进行数据融合，减少冗余数据的发送，从而延长网络生命周期。同时，当节点损坏或数据丢失时，可以利用周围邻居节点的数据相关性特点，在一定概率范围内正确发送查询结果。

第三，节点能量受限，必须提高能量利用效率。根据建立的数据模型，可以调节传感器节点工作模式，降低节点采样频率和通信量，达到延长网络生命周期的目的。

第四，方便查询和数据分布管理。

4.2 研究进展

早期对数据模型的研究主要是对传统关系模型、对象关系模型或时间序列模型的扩展。美国加州大学伯克利分校开发的TinyDB系统[2,3,4]对传统关系数据模型进行了简单扩展，把传感器网络定义为一个单一的无限长的虚拟关系表，节点的每个数据对应关系表中的一行。美国康乃尔大学Cougar系统[5]的数据模型中，传感器网络元数据（节点ID等）用传统的关系来表示，而感知数据用时间序列来表示，数据模型包括关系代数操作和时间序列操作，操作尽可能在网内分布式处理以减少通信资源的消耗。

这样的数据模型与传统分布式数据库结构比较相似，有利于查询的实现，但是并不能解决数据不确定性问题。在此基础上提出了结合概率的数据模型，主要包括基于概率模型的数据缓冲策略，客户端-服务器模式的概率传输策略以及对数据流处理的概率模型。

针对节点损坏、数据丢失、数据不确定性等问题，文献[6]提出了一种基于概率的多维数据模型，通过观察确定数据某一属性的概率分布情况，根据概率模型计算感知数据，从而提高查询效率，保障结果的正确性，并支持某些特殊查询。

文献[7]针对“select *”查询能量效率问题，采用动态概率复制模型实现了Ken机制，在基站和网络分别维持一个感知数据模型，当节点感知数据与网络内模型预测值不符时才传输数据到基站。这种策略尤其适用于事件驱动的应用，但是结果精度和模型相关。

国内，哈尔滨工业大学的研究人员[8]提出了一种基于感知数据概率模型的分布式采样和通信动态调度算法，使传感器节点根据概率模型确定自己的采样和通信时机，最小化采样频率和通信量，减少传感器节点的能量消耗，延长传感器网络的生命期。

5 数据存储与索引技术

数据存储策略按数据存储的分布情况可分为以下三类。

（1）集中式存储。节点产生的感知数据都发送到基站节点，在基站处进行集中存储和处理。这种策略获得的数据比较详细完整，可以进行复杂的查询和处理，但是节点通信开销大，只适合于节点数目比较小的应用场合。加州大学伯克利分校在大鸭岛上建立的海鸟监测试验平台就是采用这种策略。

（2）分布式存储和索引。感知数据按数据名分布存储在传感器网络中，通过提取数据索引进行高效查询，相应存储机制有DIMENSIONS、DIFS、DIM等。

DIMENSIONS[9]采用小波编码技术处理大规模数据集上的近似查询，有效的以分布式方式计算和存储感知数据的小波系数，但是存在单一树根的通信瓶颈问题。为了解决这一问题而提出的DIFS[10]，其中使用感知数据的键属性，采用散列函数和空间分解技术构造多根层次结构树，同时数据沿结构树向上传播，防止了不必要的树遍历。DIFS是一维分布式索引，而DIM（Distributed Index for Multidimensional data）[11]则是多维查询处理的分布式索引结构，使用地理散列函数实现数据存储的局域性，把属性值相近的感知数据存储在邻近节点上，减少计算开销，提高查询效率。文献[12]在DIM基础上对保持局域性的散列函数提出了一种分布式算法，有效地解决了DIM存在的区域热点问题。

文献[13]在地理路由协议GPSR基础上，提出了一种以数据为中心的存储方法GHT，采用周界更新协议保证节点失效时的数据可靠性，同时提出了结构复制技术解决热点问题。

（3）本地化存储。数据完全保存在本地节点，数据存储的通

信开销最小，但是查询效率低下，一般采用泛洪式查询，当查询频繁时，网络的通信开销极大，并且存在热点问题。

文献[14]面向数据需要大量存储在节点的科学应用，提出了一种协作存储机制，利用相邻节点数据的空间相关性显著地减少了需要存储的数据量，并且能够提供负载均衡。

6 数据查询处理

传感器网络中的数据查询主要分为快照查询和连续查询。快照查询是对传感器网络某一时间点状况的查询，连续查询则主要关注某段时间间隔内网络数据的变化情况。查询处理与路由策略、感知数据模型和数据存储策略紧密相关，不可分割。当前的研究方向主要集中在以下几个方面：

(1) 查询语言研究。这方面的研究目前比较少，主要是基于SQL语言的扩展和改进。TinyDB系统的查询语言是基于SQL的，康乃尔大学的Cougar系统[5]提供了一种类似于SQL的查询语言，但是其信息交换采用XML格式。

(2) 连续查询技术。传感器网络中，用户的查询对象是大量的无限实时数据流，连续查询被分解为一系列子查询提交到局部节点进行执行。子查询也是连续查询，需要扫描、过滤、综合数据流，产生部分的查询结果流，经过全局综合处理后返回给用户。局部查询是连续查询技术的关键，由于节点数据和环境情况动态变化，局部查询必须具有自适应性。

文献[15]介绍Cougar系统的查询处理方法，并提出了网内(in-network)查询处理思想，给出了相应的查询处理器结构，同时讨论了查询语言、多查询优化、catalog管理等问题。文献[16]认为树型的查询处理过程中，把查询分解为子查询，交给局部操作子处理，实际上就是一种任务分配调度过程。他们提出了一种自适应的分布式算法，根据邻居节点精简局部子操作，减少网络负载。

文献[17]提出了一种在无限实时感知数据流上处理连续查询的自适应技术CACQ。对于传感器节点上的单个连续查询，CACQ把查询分解为操作序列，针对进入系统的数据来调度执行序列中的每个操作，产生查询结果。当节点同时执行N个连续查询时，CACQ轮流地把数据传递到N个查询操作序列，完成处理，而不用每次都复制数据。CACQ存在的问题是：只能处理单个传感器节点查询，而不能处理整个网络查询。

(3) 近似查询技术。感知数据本身存在不确定性，用户对查询的结果的要求也是在一定精度范围内的。采用基于概率的近似查询技术，充分利用已有信息和模型信息，在满足用户查询精度要求下减少不必要的数据采集和数据传输，将会提高查询效率，减少数据传输开销。

文献[18]证明了基于概要的近似技术可以应用到传感器网络中。

文献[19]提出了一种信息驱动的传感器查询处理方法IDSQ，使用概率模型估计目标跟踪应用中目标的位置。IDSQ按照最小化目标位置不确定性原理为每个传感器节点分派任务。

文献[20]提出了一种模型驱动的查询处理方法。这种方法根据已经存储和正在产生的感知数据，建立一个感知数据的数学模型，然后基于这个模型来回答用户的查询，减少了数据传输开销。该方法的问题是：当传感器节点数量较大时，数学模型的计算复杂性非常高。

(4) 多查询优化技术。在传感器网络中一段时间间隔内可能进行着多个连续查询，多查询优化就是对各个查询结果进行判别，减少重叠部分的传输次数以减少数据传输量。

文献[21]针对多询问的优化，提出了一种完全分布式的技术。把每一时间元(epoch)分为两阶段：查询准备阶段和结果传播阶段；提出了结果共享技术以有效处理多询问融合，结果编码技术以备传感器数据的随机更新。

7 数据融合方法

数据融合是将多个数据流或多份数据进行综合和处理，组合出更精简有效且能满足用户需求的数据过程。一般来说，传感器网络中数据冗余度比较高，而大量冗余数据的存储和发送会严重浪费网络的有限资源。数据融合方法是解决这一问题的重要手段。

数据融合技术可以在传感器网络协议栈的各个层次中实现，但主要在网络层和应用层。

(1) 应用层的数据融合。在应用层，由于能够直接理解应用数据的语义，从而根据应用需求可以最大限度的压缩数据。近年来研究比较多，提出了一些采用树型和DAG拓扑结构的聚集算法。

文献[22，23]中提出了基于树的数据融合算法。算法以sink节点为根把传感器网络组建为树状结构，从叶节点开始，自下而上进行聚集计算。算法的问题是：当节点故障或者连接故障发生时，将丢失一棵子树上所有节点的数据，会引起很大的结果误差。文献[24]提出了一种基于DAG的聚集计算方法，分布式的网内处理，采用类SQL语言带来的端到端技术部分克服了基于树算法的问题，但是仍然存在结果误差的问题。为了进一步解决上述问题，文献[25]提出了动态维护聚集树的算法，讨论了人工因素对系统性能的影响，并证明通过精心选择路径能够很好地消除节点故障或者连接故障对聚集结果的影响。这个算法的缺点是每个节点都要存储维护链路质量的统计信息，存储资源消耗大。文献[26]结合duplicate-insensitive-sketches方法和多路径路由技术，提出了一种避免节点或链接故障影响的聚集算法，能够给出比较精确的聚集结果。

如何构造最优的路由树是上述算法的关键。在文献[27]中证明了对于一个节点任意放置的传感器网络，构造每个数据传输次数都最少的路由树问题是NP难度问题。

(2) 独立的数据融合层。文献[28]提出了在网络层和MAC层之间的独立的数据融合层，不关心数据内容而直接根据下一跳地址进行数据单元合并，主要减少数据封装开销和降低MAC层的

发送冲突率。

(3) 网络层的数据融合。在网络层，数据融合技术都是与以数据为中心的路由协议相结合的。在数据转发中，路由节点根据数据内容对接收到的数据进行综合和处理，然后转发。

定向扩散(directed diffusion)路由协议[29]能够实现在传输数据的同时进行数据融合。该协议中的融合操作包括路径建立阶段的查询请求聚集和数据发送阶段的数据聚集。查询请求聚集得益于数据基于属性的命名方式。数据名字相同、监测区域完全覆盖的查询请求在某些情况下可以聚集成一个请求。在数据聚集阶段，通过缓存转发过的数据，抑制数据的重复转发。

LEACH[30]和 TEENL[31]路由协议是层次式路由协议，采用分簇的方法进行数据融合。每个簇头在收到本簇成员的数据后进行聚集处理，并将结果发送给 sink 节点。LEACH 算法仅强调了数据融合的重要性，并未给出具体的融合方法。TEEN 与定向扩散路由一样，通过缓存机制抑制不需要转发的数据。同时提出硬软门限值机制，对与前一次监测结果差值较小的数据传输也进行了抑制。

8 结束语

无线传感器网络是一种综合了传感器技术、嵌入式计算技术、分布式信息处理技术、数据库技术等多种技术的新兴网络，当前在国际上备受关注，被认为是将对21世纪产生巨大影响力的技术之一。本文对数据管理技术的概念、研究内容和已有的发展及研究状况进行了归纳和总结，着重介绍了当前国际上研究热点领域。

参考文献

[1] 李建中，李金宝，石胜飞. 传感器网络及其数据管理的概念、问题与进展. 软件学报，2003，14(10):1717～1727。

[2] TinyDB, http://telegraph.cs.berkeley.edu/tinydb/

[3] Sam Madden, Joe Hellerstein, and Wei Hong, TinyDB: In-Network Query Processing in TinyOS, Intel Research,? IRB-TR-02-014,Oct. 1, 2002.

[4] Samuel Madden. The Design and Evaluation of a Query Processing Architecture for Sensor Networks, Ph.D. Thesis. UC Berkeley. Fall, 2003。

[5] Gerhke J. COUGARdesign and implementation http://www.cs.cornell.edu/database/cougar/。

[6] Amol Deshpande,Carlos Guestrin,Samuel R. Madden. Using Probabilistic Models for Data Management in Acquisitional Environments.Proceedings of the 2005 CIDR Conference.

[7] David Chu,Amol Deshpande,Joseph M. Hellerstein,Wei Hong. Approximate Data Collection in Sensor Networks using Probabilistic Models.

[8] 李建中,石胜飞,王朝坤。基于感知数据概率模型的无线传感器网络采样和通信调度算法。计算机应用，2005.9。

[9] Deepak Ganesan, Deborah Estrin, John Heidemann, Dimensions: why do we need a new data handling architecture for sensor networks, ACM SIGCOMM Computer Communication Review,? Volume 33 Issue 1, January 2003.

[10] C.S.Raghavendra, Krishna M.Sivalingam, Taieb Zhati(Eds.), Wireless Sensor Networks. Kluwer Academic Publishers, 2004. P185-252.

[11] XinLi Young, Jin Kim, Ramesh Govindan, Wei Hong, Multi-dimensional Range Queries in Sensor Networks.SenSys03.

[12] Xin Li,Ramesh Govindan,Wei Hong,Fang Bian. Rebalancing Distributed Data Storage in Sensor Networks.

[13] Sylvia Ratnasamy, Brad Karp, Scott Shenker, Deborah Estrin, Ramesh Govindan, Li Yin and Fang Yu。Data-Centric Storage in Sensornets with GHT, A Geographic Hash Table.appear in Mobile Networks and Applications (MONET), Special Issue on Wireless Sensor Networks, Kluwer, mid-2003.

[14] Sameer Tilak, Nael B. Abu-Ghazaleh and Wendi Heinzelman.Collaborative Storage Management In Sensor Networks.arXiv: cs. NI/0408020 v1 6 Aug 2004.

[15] B. Bonfils and P. Bonnet, "Adaptive and Decentralized Operator Placement for In-Network Query Processing,"Information Processing in Sensor Networks) (ISPN), April 2003.

[16] Samuel Madden, Mehul Shah, Joseph M. Hellerstein, Vijayshankar Raman, Continuously Adaptive Continuous Queries over Streams, in Proc. of ACM SIGMOD, USA, 2002..

[17] Y. Yao and J. Gehrke, HYPERLINK 。The cougar Approach to In-Network Query Processingin Sensor Networks, SIGMOD, March 2002

[18] G. Kollios, J. Considine, F. Li, and J. Byers. Approximate aggregation techniques for sensor databases. In Proc. of ICDE, 2004.

[19] M. Chu, H. Haussecker, and F. Zhao. Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks. Journal of High Performance Computing Applications, 2002.

[20] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, Wei Hong, Model-Driven Data Acquisition in Sensor Networks, in Proc. of the 30th VLDB Conference, Canada, 2004.

[21] Niki Trigoni,Yong Yao,Alan Demers,Johannes Gehrke,and Rajmohan Rajaraman. Multi-query Optimization for Sensor Networks。

[22] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. The design of an acquisitional query processor for sensor networks. In ACM SIGMOD, 2003.

[23] Y. Yao and J. Gehrke. Query processing in sensor networks. In CIDR, 2003.

[24] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. Tag: A tiny aggregation service for ad hoc sensor networks. In USENIX OSDI, 2002.

[25] J. Zhao, R. Govindan, and D. Estrin. Computing aggregates for monitoring wireless sensor networks. In IEEE SPNA, 2003.

[26] J. Considine, F. Li, G. Kollios, and J. Byers. Approximate aggregation techniques for sensor databases. In IEEE ICDE, 2004.

[27] Krishnamachari B, Estrin D, Wicker S. Modelling data-centric routing in wireless sensor networks, in Proc. of IEEE INFOCOM, 2002.

[28] He T, B M,Stankovic J A, Abdelzather T F.AIDI: Adaptive application independent data aggreation in wireless sensor networks.ACM Transactions on Embedded Computing ststenm2004,3(2):426-457.

[29] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. MobiCOM 2000, Boston, Massachusetts, August 2000.

[30] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In the Proceedings of the Hawaii International Conference on System Sciences, Maui, Hawaii, January 4-7, 2000.

[31] Manjeshwar and D.P. Agrawal. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. Parallel and Distributed Processing Symposium.,Proceedings 15th International , 23-27 April 2001.

作者简介：

纪德文(1983-)，男，山东德州人，国防科技大学硕士，主要研究方向为无线传感器网络。

王晓东(1974-)，男，山东人，博士，国防科技大学副教授，主要研究方向为无线通信，移动自组网等。