

监测型传感网的“预测-修正”数据管理方法

洪文兴^{1*}, 朱顺痣^{2,3}, 李茂青¹

(1. 厦门大学信息科学与技术学院, 福建 厦门 361005; 2. 厦门理工学院计算机科学与技术系, 福建 厦门 361024;

3. 福州大学空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002)

摘要: 将无线传感器网数据管理技术分为 4 类: 陈述性语句、信号处理、模型预测以及中间件, 并进行对比研究总结出如何既将传统的数据管理技术移植到传感网, 又考虑到传感网自身的能源和网络带宽约束. 针对监测型传感网, 提出了一种新的“预测-修正”传感网数据管理方法, 对传感节点的每个测量建立一个对应的虚拟表, 所有的虚拟表又组成一个虚拟数据库. 通过读取、预测、修正 3 个步骤, 获得较为精确的测量值. 最后用实验室温度监测实验验证了本方法.

关键词: 无线传感器网络; 数据管理; 预测; 监测

中图分类号: TP 393. 07

文献标识码: A

文章编号: 0438-0479(2010)04-0509-05

无线传感器网络(传感网)是当今比较前沿的研究课题, 广泛用于军事、工业和家庭应用等领域. 传感网的用途一般有监测、时间检测、跟踪和目标分类等 4 种^[1]. 比较知名的案例, 包括美国缅因州大鸭岛环境监测项目^[2]、德拉华大学的机动车跟踪项目^[3]和美国国家航空与航天局(NASA) JPL 实验室的 SensorWebs 计划等^[4]. 这些案例都集中反应了传感网的特有条件约束, 包括能量限制、通讯能力限制和数据冗余等.

针对这些约束, 在传感网数据管理方面, 研究人员除了将传统的数据管理技术应用于传感网外, 还提出很多新的方法. 文献[5]以美国 Cornell 大学研发的 Cougar 系统^[6]、UC Berkeley 研制的 TinyDB 系统^[7]和 UCLA 研制的 Dimensions 系统^[8]为例, 介绍了传感网的数据管理技术. 另一方面, 从技术发展维度而言, 传感网数据管理技术经历以下 4 个阶段. Govindan 等^[9]将传感网视为一个关系型数据库, 并提出两个重要的想法: 一是在网内实现主要的数据库查询工作; 二是对传感网数据允许使用近似值. 文献[10-11]定义了混合传感数据库, 将存储数据视为关系, 而传感数据作为时间系列. 文献[12-14]为传感网提出了声明式编程语言, 以降低数据操作的难度. 文献[15-16]则提出了各自的模型驱动的方法用来预测传感数据.

本文正是从技术发展的维度, 对这 4 种传感网数据管理技术进行对比研究, 并提出一种新的适用于监测型传感网的数据管理方法, 称为“预测-修正”方法.

1 传感网数据管理技术的发展维度

1.1 将传感网直接映射为传统数据库

根据传感网的部署特征, Govindan 等^[9]认为传感网是虚拟的关系数据库. 在传感网中, 传感节点可以视为感知数据流, 按需获取数据. 从这个角度而言, 每个传感器节点就是一个小型即时数据库. 此外, 研究人员也很容易搭建一个非结构化用户界面去获取和维护这些小型数据库. 至此, 传感网便可当成传统数据库使用, 但却包含一对矛盾:

如果传感网是传统的集中式数据库, 而它固有的通讯能力和能源约束却限制集中式数据库的实现. 反之, 如果传感网是分布式数据库, 而它却缺乏各个节点的全局信息和网络拓扑结构.

为解决这对矛盾, 此阶段的研究引入两个基本的策略, 一是利用网内通讯实现简单的查询; 二是允许传感网数据使用近似值. 这两个策略能奏效的根据是, 传感网各节点的计算能力能将复杂的查询命令转化为简单的查询结果. 为了降低通信量, 一旦收到查询指令, 节点就利用自身的计算能力提供候选数据集. 然后, 在头节点处对这些数据集进行聚合, 并最终反馈给终端用户. 此外, 这些节点也会运行数据恢复算法和信号处理算法, 以提供估计值. 在一定的置信区间内, 此技术下的数据精度可以满足传感网的一些基本应用.

1.2 混合传感数据库

Cougar 系统^[6]是一种混合型传感网数据管理技术, 考虑一个工厂仓库的传感网应用系统, 根据查询需

收稿日期: 2010-02-20

基金项目: 空间数据挖掘与信息共享教育部重点实验室开放课题资助(2010); 厦门大学国家 211 三期项目

* 通讯作者: hwx@xmu.edu.cn

要的数据的不同,将数据查询分为如下 3 类:

1) 历史数据查询: 比如, 获取过去 10 min 各个楼层的平均温度. 历史数据查询需要的是过去的的数据, 其预期的结果是一些统计值, 比如平均值、最大值、最小值等;

2) 快照数据查询: 比如, 获取各个传感器节点捕捉到的异常温度. 快照数据查询表示一旦某个事件发生, 传感器必须立即反馈当前的读数数据;

3) 持续数据查询: 比如, 获取第 3 楼层各个传感器节点每分钟的温度读数. 持续数据查询是一种监测, 在线和实时记录了每个传感器节点读数.

这些查询涉及两种数据类型: 存储数据和传感数据. 对于一些经常被使用的数据, 将之存储在各节点有限的存储器中, 并用关系型数据库来表示. 比如, 历史数据查询的结果就可以认为是存储数据, 存放在各个节点中. 传感数据则是那些为了满足当前查询的需要, 从节点即时读出的数据. 当运行一个快照数据查询或持续数据查询, 系统就需要即时读数. 传感数据就是这些即时的时间系列值.

为了管理这两类数据, Cougar 提供了一种新的数据模型, 称为抽象数据类型 (Abstract data type, ADT). 以下是一个数据类的表示.

Sensors (location POINT, floor INT, s SENSORNODE)

其中, location 表示节点的位置, 属于存储数据; floor 表示节点所处的楼层, 属于存储数据; s 表示一个传感节点, 属于支持信号处理功能的 ADT.

这个 ADT 提供两种方法, getTmp() 和 detectAlarmTemp(threshold). 调用 s.getTmp() 方法用以获取当前节点的温度读数, 调用 s.detectAlarmTemp(阈值) 表示当前节点会在温度超过设定阈值时发出警报.

总之, 这种管理技术融合了传统数据库技术和信号处理技术, 传感节点被认为是设备数据库系统的一部分. 根据数据使用的频次和场景, 将数据划分为存储数据和传感数据两类.

1.3 用于传感网的声明式编程语言

TinyDB 系统使用新的工作机理, 建立了程序开发人员和传感网之间的桥梁^[7]. 该系统为传感网的数据管理提供了一种有力的声明式工具. 在此基础上, 同一研究小组又发布了声明式传感网 (Declarative sensor network, DSN), 并提出了 Snlog 语言用来操控和管理传感网的数据以及网络拓扑结构^[12-14].

众所周知, 由于 TinyOS 和 nesC 缺少成熟的编译环境, 传感网难于编程. 而 TinyDB 和 DSN 的目标就是要掩盖这种困难, 让编程人员忽略传感网底层实现

的细节. DSN 的执行机理易于理解: 借助于 Snlog 语言, 当一个数据管理任务产生时, 编程人员 (甚至最终用户) 都能快速的编写程序. 这些程序被提交给系统以后, 系统会将其优化、编译, 并转化为 nesC 代码. 而后, 现有的 nesC 编译器对这些代码进行编译, 从而生成可执行的双字节程序映像, 用于下载到各个传感节点中.

声明式语言提供了一种简单的编程查询方法, 即使对传感网底层运行机制毫不知情的最终用户也可以轻易的管理传感网的数据. 该系统同时也处理了网络带宽约束, 并会对查询结果自动进行优化. 最重要的一点是, 这种机理不仅考虑了关系型数据库也考虑信号处理功能, 不仅用于数据管理也可用于网络拓扑管理.

1.4 模型驱动的数据管理技术

到现在为止, 我们可以方便的获取和操作传感网数据. 但对大部分的传感网应用而言, 数据的准确性也是很重要的. 在文献[15-16], 研究人员用信号处理、模型修正和预测方法来确保数据的准确性. 此技术基于以下两个合理的假设:

1) 数据伪表示: 对某个查询而言, 我们预计能获取所有的相关数据. 但是, 由于物理现象的连续性, 我们不得不对数据进行采样并转化为时间系列, 这势必为数据带来噪声. 因此, 所获得的数据并不是客观世界的真实反应, 只能是“伪表示”.

2) 非预期的近似查询: 在一个采样数据库中, 由于采样位置和采样时间点的选择, 导致所有的数据都是真实数据的近似值. 而数据查询则是在这些近似值上产生一定的结果, 显然这些结果也并不可信.

以上假设说明传感网测量的最终数据并不准确. 为解决这个问题, 我们引入信号处理方法, 并利用交叉属性模型来预测数据. 图 1 是该方法的一个例子. 在该例子中, 预期的数据查询结果是来自于 8 个传感节点的温度读数. 通过对历史数据的分析发现: 1) 温度读数和节点电压密切相关; 2) 获取电压的代价远小于获取温度读数的代价; 3) 各个节点温度读数是有关联的. 基于这 3 条知识, 我们只需要获得节点 4 的温度读数、节点 1 和节点 2 的电压读数, 就可以预测出其他 7 个节点的温度读数. 此外, 用类似的方法, 我们还可以验证节点数据的正确与否, 或者判断节点是否失效.

2 “预测-修正”数据管理方法

2.1 基本思想

无线传感器网络中的数据管理应该根据应用的不

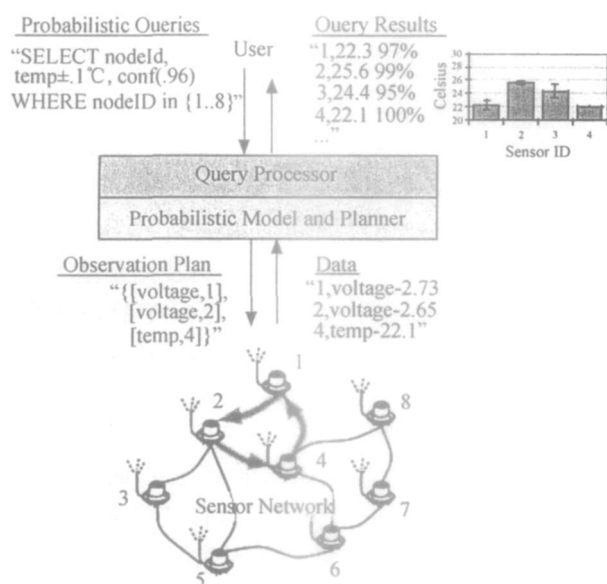


图 1 使用交叉属性模型预测数据

Fig. 1 Use cross attributes model to predict
(Source: TinyDB Group, UC Berkeley)

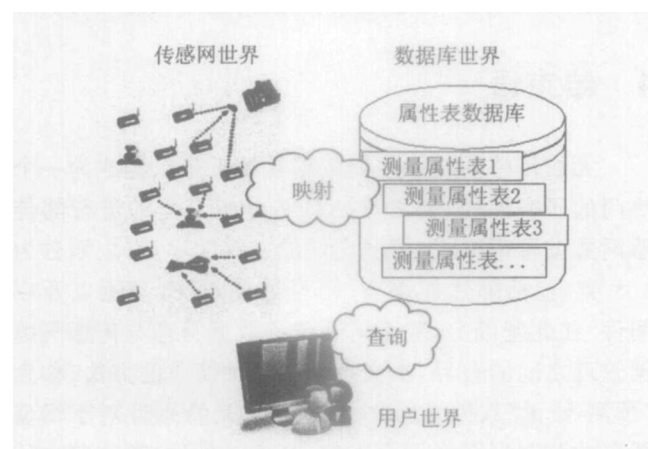


图 2 “预测 修正”数据管理方法基本思想

Fig. 2 Basic idea of "Predict-Revise" data management method

同(比如跟踪或者监测)而有所不同。在跟踪应用中,即时数据是必需的,应该从传感节点中直接获得。与之对比的,在监测应用中,通常设定一个阈值以判断是否需要从节点中直接读取数据。另一个影响数据管理模式的因素是节点的部署方式。如果节点的部署是稠密的,则必须考虑路由算法;反之,如果节点的部署是稀疏的,则考虑读取所有节点的数据。本方法是针对稠密部署的监测型无线传感器网络提出的。

传感网数据管理中间件位于应用层和网络层之间,通过提供类似于 SQL 语句的操作语言,用于直接操作数据和传感节点,从而对终端用户掩盖了传感网工作的具体细节。我们的研究就是在此之上新建一个层,以将传感网的数据管理转化为传统数据库或数据

仓库管理。进而,再直接应用现有的数据管理技术。

图 2 展示了本算法的基本思想。算法为现实的传感网世界建立了对应的虚拟数据世界。这样,所有的数据操作就发生在如下两种场合:传感器节点与虚拟数据库之间,以及虚拟数据库和最终用户之间。

这个结构与中间件或传统的数据库视图相类似。但关键在于,它为传感器节点的每一个测量属性都提供了一个虚拟表。举例而言,如果一个节点有 3 个测量属性,温度、亮度和加速度,那么,方法将为这 3 个属性分别创建独立的数据表。这 3 个数据表又组成虚拟数据库。更进一步,为了运行历史数据查询,我们还可以将这个数据库转入数据仓库。这样,所有传统数据管理理论就可以无缝的移植到传感网中。

在每个虚拟表中,字段都由两部分组成,分别是自然属性字段和测量值字段。自然属性字段包括节点 ID、位置信息和时间戳等。测量值字段则是当前测量属性的取值。在本算法中,这个取值包括即时读数、基于当前属性模型的预测读数、基于属性交叉模型的预测读数。这 3 个读数互相验证,交替工作,以提高测量值的准确度。

2.2 方法描述

一个完整的“预测 修正”传感网数据管理算法包括以下 3 步:

第 1 步:初始步。本算法是使用循环数据管理模式。循环的初始化有两个条件:一个固定的时间窗完成或者一个事件被触发。当以上任何一个条件满足时,一次旧的数据读取结束而另一次新的数据读取开始。给定一个固定的时间周期(比如 10 s),称为时间窗。在时间窗的开始(循环的开始),算法直接从各传感节点中读取数据。此时,假定这次单个数据的读取是准确的,没有任何噪声。

第 2 步:预测步。在这 10 s 中周期内,如果系统从未达到规定的阈值,也就是所监测区域没有任何异常,称为静默期。在此期间,使用模型(比如 EKF 滤波模型)来进行数据预测,获得此期间各秒的读数。

第 3 步:修正步。当一个时间窗结束或者一个事件被触发,也即达到循环条件中的任何一个,系统将再次直接读取节点数据,并用于验证和修正预测步中产生的数据。此外,算法使用交叉属性模型对节点读数进行修正。交叉属性模型是指,该数据模型的变量包含除当前测量属性外的其他一个或多个属性。在本算法中,具体体现为多个虚拟表。注意到,每次修正步的执行都表示一个算法循环的结束,并开始一个新的初始步,启动下一个循环。

2.3 一个例子

温度监测实验是本文作者在新加坡南洋理工大学留学期间,在传感网实验室所进行的,用于说明“预测-修正”方法.实验的目标是实时监测实验室内的温度变化.实验室长 8 m,宽 6 m,使用中央空调调控温度.实验使用 6 个 Crossbow 的传感器节点(型号 SN24040CA),用来监控室内温度.节点位置使用二维坐标表示,分别与实验室的长宽对应.数据读取使用 Crossbow 附带的软件.读数和算法预测值片段见表 1(限于篇幅只取一个周期一个节点的数据).

表格第 1~3 行表示,在初始步阶段,系统直接读取各个节点的读数.将这些读数记录在“Reading”(读数)字段.这是一个循环的开始,当前时间戳是 14:00:50.此后系统进入静默期,由于时间窗未到(10 s),也没有任何异常事件发生(温度变化超过 1℃),系统每隔 2 s 自行预测一次,将对各个节点的预测值记录在“预测读数”字段中.在本例中,可以看到每隔 2 s,系统对节点 1 的预测基本保持在 24.1℃,其他节点大抵类似.表格第 8 行中,注意到节点 1 再次读取数据,这是因为运行已满 1 个时间周期(10 s),此时一个循环结束,下一个循环开始.我们利用系统模型对前一阶段的预测读数进行修正,并记录在“修正读数”字段中.这样,本方法在 10 s 内仅两次直接读取数据,便得到系统每隔 2 s 的具体读数.数据直接读取次数的减少,降低了节点的能量损耗和网络负载.

在实际应用场景中,监测周期可能很长,时间窗也可以随之延长,能量节省和网络负载减轻将更加明显.

3 结果和讨论

“预测-修正”传感网数据管理模型已在实验室环境下得以验证.但尚存在如下挑战和改进空间:

- 1) 选取更有效的预测模型和交叉属性模型来预测和修正读数.在本文中,这些模型的使用是基于历史数据和机器学习技术得来的.但对于一个最新部署的应用而言,并没有历史数据可以使用,机器学习技术的选择也因此受到限制.此情况下,如何保证模型预测的精度,将成为一个问题.
- 2) 本算法的动机和考虑的出发点是,如何在将现有成熟的数据管理模式从传统领域移植到传感网领域的同时,考虑传感网的能量约束和网络带宽约束.但对节点运算能力和存储能力的利用并不突出.
- 3) 关于阈值的设定,在实验室例子中,我们设定为“温度变化超过 1℃”.但在实际应用中,这个阈值应该是动态的.引入动态阈值,将使温度曲线更加精确.

4 结束语

无线传感器网络中的数据管理近来已经成为一个热门的研究课题.本文首先总结分析了当前流行的传感网数据管理模式.根据它们的发展顺序,将之划分为 4 大类,包括陈述性语言、信号处理、模型预测以及中间件.在此基础上,基于填补传统数据管理与传感网数据管理之间的差异,本文提出了一种简单的方法,称为“预测-修正”数据管理方法.该方法是特别针对于稀疏部署的监测型网络而提出的.文章的最后,在实验室环

表 1 单周期内节点# 1 读数
Tab. 1 Readings of Node # 1 in one time window

行	节点 ID	位置信息	时间戳	读数/℃	预测读数/℃	修正读数/℃	所属步骤
1	1	(1, 1)	20081201 14:00:50	24.0			初始步
2	2	(1, 3)	20081201 14:00:50	23.8			初始步
3	3	(4, 1)	20081201 14:00:50	24.2			初始步
...	...	...	...				
4	1	(1, 1)	20081201 14:00:52		24.1		预测步
5	1	(1, 1)	20081201 14:00:54		24.2		预测步
6	1	(1, 1)	20081201 14:00:56		24.1		预测步
7	1	(1, 1)	20081201 14:00:58		24.1		预测步
...	...						
8	1	(1, 1)	20081201 14:01:00	24.1		24.1	修正步
...							

境下用温度监测的实验验证了这个方法, 并对这个数据管理方法的不足进行了讨论, 并提出下一步的研究方向。

参考文献:

- [1] Kuorilehto M, Hannikainen M, Hamalainen T D. A survey of application distribution in wireless sensor networks [J]. EURASIP Journal on Wireless Communications and Networking, 2005, 5(5): 774-788.
- [2] Mainwaring A, Culler D, Polastre J, et al. Wireless sensor networks for habitat monitoring[C]//Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications. Atlanta, Georgia, USA: ACM, 2002: 88-97.
- [3] Shen C C, Srisathapornphat C, Jaikaeo C. Sensor information networking architecture and applications[J]. Personal Communications, IEEE, 2001, 8(4): 52-59.
- [4] Jet Propulsion Laboratory N. Volcano Sensorweb[EB/OL]. [2009-01-01]. <http://sensorwebs.jpl.nasa.gov/>.
- [5] 刘琳, 于海斌, 曾鹏. 无线传感器网络数据管理技术[J]. 计算机工程, 2008, 34(2): 62-65.
- [6] Yao Y, Gehrke J. The cougar approach to in-network query processing in sensor networks [J]. SIGMOD Record, 2002, 31(3): 9-18.
- [7] Madden S R, Franklin M J, Hellerstein J M. TinyDB: an acquisitional query processing system for sensor networks [J]. ACM Transactions on Database Systems (TODS), 2005, 30(1): 122-173.
- [8] Ganesan D, Estrin D, Heidemann J. Dimensions: why do we need a new data handling architecture for Sensor Networks? [J]. ACM SIGCOMM Computer Communications

Review, 2003, 33(1): 143-148.

- [9] Govindan R, Hellerstein J, Hong W, et al. The sensor network as a database[R]. Los Angeles, CA, USA: University of Southern California, 2002: 1-8.
- [10] Bonnet P, Gehrke J, Seshadri P. Querying the physical world[J]. Personal Communications, IEEE, 2000, 7(5): 10-15.
- [11] Bonnet P, Gehrke J, Seshadri P. Towards sensor database systems[J]. Lecture Notes In Computer Science, 2001(01): 3-14.
- [12] Madden S R. The design and evaluation of a query processing architecture for sensor networks[D]. Berkeley: University of California, 2003.
- [13] Franklin M J, Hellerstein J M, Madden S. Thinking big about tiny databases[J]. Data Engineering, 2007, 30(3): 37-48.
- [14] Chu D, Popa L, Tavakoli A, et al. The design and implementation of a declarative sensor network system[C]//Proceedings of the 5th International Conference on Embedded Networked Sensor Systems. Sydney, Australia: ACM, 2007: 175-188.
- [15] Amol Deshpande, Carlos Guestrin, Samuel R. Madden, model driven data acquisition in sensor networks[C]//Proceedings of the Thirtieth International Conference on Very Large Data Bases. Toronto, Canada: VLDB Endowment, 2004, 30: 588-599.
- [16] Deshpande A, Guestrin C, Hong W. Exploiting correlated attributes in acquisitional query processing[C]//21st Intl conference on Data Engineering (ICDE 2005). Tokyo: IEEE Computer Society, 2005: 143-154.

"Predict-Revise" Data Management Method for Wireless Sensor Networks of Monitoring

HONG Wenxing^{1*}, ZHU Shurui^{2,3}, LI Maoping¹

(1. School of Information Science and Technology, Xiamen University, Xiamen 361005, China;

2. Department of Computer Science and Technology, Xiamen University of Technology, Xiamen 361024, China;

3. Key Lab of Spatial Data Mining and Information Sharing, Ministry of Education, Fuzhou University, Fuzhou 350002, China)

Abstract: The data management techniques of wireless sensor networks (WSN) are summarized into four categories, i.e., declarative language, signal processing, model prediction, and middleware. This paper describes the key of moving the traditional data management techniques to WSN, and considers the constraints of power consumption and bandwidth limit at the same time. For the WSN of monitoring, a new "Predict-Revise" data management method is proposed. It creates one virtual table for each measurement variable of sensor nodes, and combines all the tables into a virtual database. After approaching reading, prediction, revise, the system will get the result with higher accuracy. An experiment for monitoring the temperature of lab is presented to verify this new method.

Key words: wireless sensor networks; data management; prediction; monitoring