

面向无线传感器网络的流数据压缩技术^{*}

刘向宇 王雅哲 杨晓春 王斌 于戈
(东北大学信息科学与工程学院 沈阳 110004)

摘要 目前无线传感器网络广泛应用在各个领域来进行复杂的监控和计算任务。因其资源有限的特点(处理、传输带宽、能量),对无线传感器网络在数据处理和传输方面提出了新的挑战。本文提出一种基于曲线拟合技术的流数据的压缩传输方法 CODST,用于压缩每个传感器采集到的数据,并在基站进行数据还原。实验表明,提出的方法在保证压缩传输后的流数据的精度前提下可以大大减少数据传输量,节省网络带宽和能量。

关键词 无线传感器网络,数据压缩,数据流,曲线拟合

Study on Compression Technique for Streaming Data in Wireless Sensor Networks

LIU Xiang-Yu WANG Ya-Zhe YANG Xiao-Chun WANG Bin YU Ge
(Information Science and Engineering College, Northeastern University, Shenyang 110004)

Abstract Wireless sensor networks have been applied in several scenarios to perform complex monitoring and computing tasks. Such networks pose new challenges in data processing and dissemination because of the limited resources (processing, bandwidth, energy) that such devices possess. In this paper, a compression technique (named CODST) based on curve simulating is proposed to compress streaming data collected by each sensor. The compressed streaming data are recovered in the base station. Experimental results show that the proposed compression technique CODST preserves the precision of the compressed streaming data. CODST also reduces the quantity of transmitting data, which saves the bandwidth and energy.

Keywords Wireless sensor network, Data compressing, Data stream, Curve simulating

1 前言

随着低能源嵌入式通讯设备方面科技的进步,无线传感器网络已应用在越来越多的领域。大规模传感器网络需要实时的数据处理和传输。

为降低数据传输量,本文提出了一种受限环境(例如无线传感器网络)中的数据压缩传输技术。本文采用对采集数据进行分段曲线拟合的方法来挖掘数据中固有的模式,基于模式近似估计原始数据并进行传输从而实现流数据的压缩传输。通过实验证明,本方法减少了网络带宽负荷量并提高了数据传输实时性,数据接收端采集到的数据具有较高的精度。

当前很多文献^[1~5]对数据流上连续查询的通用规则进行了研究。文[6]研究了当存在重复查询时精度和执行之间的平衡问题,采用数据缓存进行解决。文[7]给出一种概率查询评估方法,对查询结果设定了置信度从而量化记录数据的不确定性。文[8]采用线性回归进行数据流的在线多维分析。文[9]采用基于线性回归和多维流数据之间的相关性对流数据进行近似压缩传输。而本文主要采用了线性回归和非线性回归中的多项式拟合从而对流数据进行分段曲线拟合,实现数据流的压缩传输。

本文第2节介绍如何将曲线分段拟合技术同流数据传输相结合从而实现流数据的压缩传输,并对提出的 CODST 方法进行详细介绍。第3节给出实验结果,并对算法性能进行分析。最后总结全文并介绍未来工作。

2 面向无线传感器网络的流数据压缩传输

本文主要将高次多项式的最小二乘法曲线拟合技术同流数据传输相结合,从而实现流数据的压缩传输。高次多项式的最小二乘法曲线拟合技术在本文不再赘述。下面给出面向无线传感器网络的流数据压缩传输算法。首先定义流数据压缩传输中的数据格式。

2.1 流数据压缩传输的数据格式

流数据压缩传输中,传感器结点 P 将要发送到基站的 m 个数据点 $\{(id, t_j, s_j); j=1, \dots, m\}$ 存入缓存拟合得到 $s(x) = a_0 + a_1x^1 + \dots + a_nx^n$ 。

根据流数据特点,如果数据点的采集时间不均匀,为了在基站端对数据进行还原,则传输时需要将采集数据的时间序列进行传输。此时传输数据为 $Data_{Trans} = (id, t_{begin}, t_{end}, indexVec, coefVec)$ 。其中 $indexVec = (x_1, x_2, \dots, x_m)$ 表示拟合时数据点的 X 坐标,其中 $x_j = \frac{t_j - t_{begin}}{t_{end} - t_{begin}}$ 。 $coefVec = (a_0, a_1, \dots, a_n)$ 表示多项式的系数向量,若 $coefVec = (0)$ 表示采用上次的拟合结果。

采集时间均匀时, $indexVec = (x_1, x_2, \dots, x_m)$ 中 X 坐标 $x_j = m \left(\frac{t_j - t_{begin}}{t_{end} - t_{begin}} \right) = j$, 即 $indexVec = (1, 2, \dots, m)$, 这是时间均匀的性质决定的。传输 $Data_{Trans} = (id, t_{begin}, t_{end}, num, coefVec)$ 即可表示传输数据的所有信息,其中 num 表示数据点个数,即 m 。采用 num 即表示了拟合数据时的 X 坐标向量

^{*} 基金项目:国家自然科学基金(60503036,60473073)和霍英东青年基金优选课题资助(104027)。刘向宇 硕士研究生,主要研究领域为无线传感器查询处理数据安全,王雅哲 硕士研究生,主要研究领域为数据流,杨晓春 博士,副教授,主要研究领域为无线传感器网络,数据隐私保护,王斌 博士研究生,主要研究领域为 P2P, Web 查询处理,于戈 教授,博士生导师,主要研究领域为分布式数据库,Web 服务,数据流。

$indexVec = (x_1, x_2, \dots, x_m)$ 。其他的数据设置与时间非均匀情况时相同。

2.2 流数据压缩传输的压缩比

基于 2.1 节中定义的流数据的压缩传输格式,现定义流数据压缩传输中的压缩比为:

$$Rate = \frac{sizeof(Data_{Trans})}{sizeof(\{(id, t_j, s_j); j=1, \dots, m\})} \quad (1)$$

其中 $sizeof()$ 表示计算数据所占用的字节数量的操作。从式(1)中可以看出, $sizeof(Data_{Trans})$ 的大小取决于 $coefVec$ 所占字节数量,即多项式系数数目,取决于多项式最高次数 n 。进行压缩传输时需要对压缩比设定上限数值 $Compress_{Limit}$,从而符合用户对流数据压缩传输效果的要求,而满足 $Rate < Compress_{Limit}$ 的最大 n 定义为 $maxN$ 。

2.3 流数据压缩传输中的效率提高

流数据压缩传输中,很多情况下某一段时间内数据的变化符合某种模式,如果对每个时间段进行拟合传输,显然不符合数据内在性质和变化趋势并且导致数据处理效率不高。因此在流数据的拟合压缩传输中,基于拟合验证的方法来减少拟合计算次数,从而提高效率。对于一个传感器结点,如果传输前一时间段数据采用了多项式 $s(x)$,对于即将传输的当前时间段的数据采用多项式 $s(x)$ 进行拟合的误差 err 小于误差限制 $errLimit$,则不必再对当前时间段数据进行处理,直接采用 $s(x)$ 即可,从而减少拟合的计算次数,提高数据处理效率。

2.4 面向压缩的流数据传输方法

基于流数据压缩传输中的数据格式和压缩比的定义,给出面向压缩的流数据的传输(CODST)方法。

在拟合过程中采用的误差为均方差。通过输入 X 坐标向量 $indexVec$,数值向量 $valueVec$,所允许的最大误差限制 $errLimit$ 和所允许的最大拟合阶数 $maxN$,得到拟合多项式。图 1 给出了流数据压缩传输的 CODST 算法。其中 $polyfit$ 函数对输入的数据进行拟合。通过 CODST 算法即可得到符合误差限制 $errLimit$ 且阶数小于 $maxN$ 的多项式 $s(x)$,其中 $s(x)$ 具有如下特点:(1) $s(x)$ 是拟合误差小于 $errLimit$ 且阶数小于 $maxN$ 的多项式中阶数最小的多项式;(2)如果在阶数小于 $maxN$ 的多项式中不存在拟合误差小于 $errLimit$ 的多项式,则 $s(x)$ 是拟合误差最小的多项式。

```
Function: CODST
Input: indexVec /* X 轴向量 */, valueVec /* 值向量 */,
      errLimit /* 误差限制 */, maxN /* 最大拟合阶数 */
Output: coefVec /* 系数向量 */, error /* 误差 */,
       rank /* 拟合阶数 */
1. i=1; // 做一阶拟合
2. (coefVec, error) = polyfit(indexVec, valueVec, i);
   // 保存当前误差最小的结果
3. minCoefVec = coefVec; minErr = error; minN = i;
   // 误差大于限制或没有得到拟合结果则循环
4. while(error > errLimit || isNull(coefVec)){
   4.1 i++;
   // 超过最大拟合系数仍没找到满足误差限制的结果,
   // 将误差最小的结果作为拟合结果 */
   4.2 if(i > maxN){
       return(minCoefVec, minErr, minN);
   }
   4.3 (coefVec, error) = polyfit(indexVec, valueVec, i);
   // 保存当前误差最小的结果
   4.4 if(error < minErr){
       minCoefVec = coefVec; minErr = error; minN = i;
   } // end of while
5. return(coefVec, error, i);
```

图 1 流数据压缩传输 CODST 算法

在流数据压缩传输过程中,当传感器结点采集到一定数量的数据才能够进行拟合传输。因此传感器先将采集到的数

据存入缓存中,当缓存中的数据数目符合要求时再进行处理。定义传感器所能容纳的数据数目为传感器的容量(cache size)。

3 实验分析

实现 CODST 算法并对其性能在不同环境下进行测试,基于实验结果进行性能分析。

3.1 实验设置

实验测试数据为气象数据,包括空气温度(Air temperature)、压力(Pressure)等观测属性,数据点采集时间为 2006.3.15—2006.4.12(http://www.k12.atmos.washington.edu/k12/grayskies)。实验环境为:CPU: Intel Pentium4 2.8GHz,系统主存 512MB,操作系统为 WindowsXP。

3.2 实验测试及分析

3.2.1 流数据压缩传输精度分析

为了对压缩传输后的流数据质量进行量化分析,现定义相对误差如式 2 所示。

$$Error_{Relative} = \frac{|value_{origin} - value_{simulate}|}{value_{origin}} \quad (2)$$

图 2 中给出了在不同压缩传输参数条件下的传输结果。其中 $value_{origin}$ 表示原始的流数据, $value_{simulate}$ 表示压缩传输之后的流数据。图 2(a) 给出了空气温度的原始流数据以及在缓存容量为 90、压缩比极限为 0.3 时压缩传输后的流数据。图 2(b) 给出了在上述条件下空气温度流数据的相对误差。从图 2(a) 可以看出,压缩传输后的流数据能够非常接近原始流数据;并且从图 2(b) 可以看出对于空气温度,相对误差基本不会超过 0.05。因此 CODST 方法能够很好地保证压缩传输后的数据质量,处理后的数据具有较高的实用性。

3.2.2 流数据压缩比和传感器容量的关系

图 3(a) 和 (b) 分别给出了空气温度(Air Temperature)和压力(Pressure)流数据的压缩比和传感器容量的关系。从图中可以看出,当传感器容量相同时,用户设置的压缩比率的极限值越小,则压缩传输后流数据的压缩比越小。当压缩比率极限值相同时,对于空气温度属性,压缩比随着传感器容量的增大而减小;而对于压力属性,在传感器容量由 30 增长到 180 时,压缩比逐渐减小,但是当传感器容量大于 180 时,压缩比增大。因此需要根据具体情况采用合适的传感器容量。

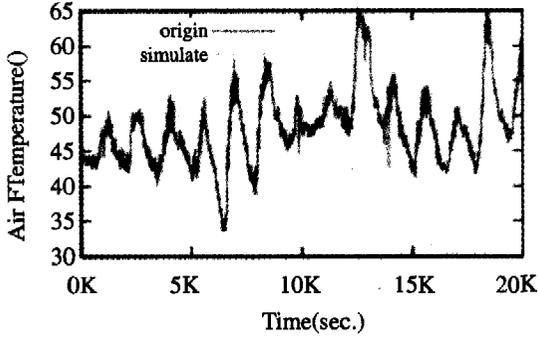
结论和未来工作 本文提出一种基于曲线拟合技术来对流数据进行压缩传输的方法—CODST 方法,通过采用多项式函数对传输的多个数据点进行近似拟合并传输,从而达到减少数据传输量和节省网络带宽的目的。大量实验结果显示, CODST 能够较高精度地对流数据进行压缩传输,处理后的数据具有较高的实用性。

参考文献

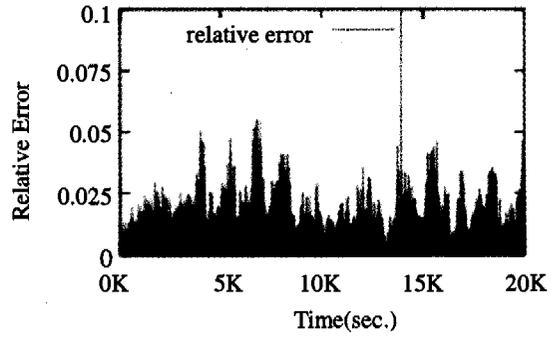
- 1 Chen J, Dewitt D J, Tian F, et al. NiagaraCQ: A Scalable Continuous Query System for Internet Databases. In: Proceedings of ACM SIGMOD Conference, 2000
- 2 Hellerstein J M, Franklin M J, Chandrasekaran S, et al. Adaptive Query Processing: Technology in Evolution. IEEE DE Bulletin, 2000, 23(2)
- 3 Motwani R, Widom J, Arasu A, et al. Query Processing, Resource Management, and Approximation in a Data Stream Management System. In: Proceedings of CIDR, 2003
- 4 Viglas S D, Naughton J F. Rate-based Query Optimization for Streaming Information Sources. In: Proceedings of ACM SIGMOD Conference, 2002
- 5 Zdonik S B, Stonebraker M, Cherniack M, et al. The Aurora and Medusa Projects. IEEE DE Bulletin, 2003
- 6 Olston C, Widom J. Offering a Precision-Performance Tradeoff for Aggregation Queries over Replicated Data. In: Proceedings of VLDB, 2000
- 7 Cheng R, Kalashnikov D V, Prabhakar S. Evaluating Probabilistic

8 Queries over Imprecise Data. In: Proceedings of ACM SIGMOD Conference, 2003
 Chen Y, Dong G, Han J, et al. Multi-Dimensional Regression Analysis of Time-Series Data Streams. In: Proceedings of VLDB,

2002
 9 Deligiannakis A, Kotidis Y, Roussopoulos N. Compressing Historical Information in Sensor Networks. In: Proceedings of ACM SIGMOD Conference, 2004

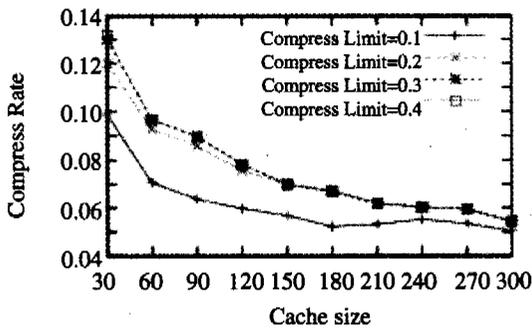


(a) Cache Size=90, Compress Limit=0.3

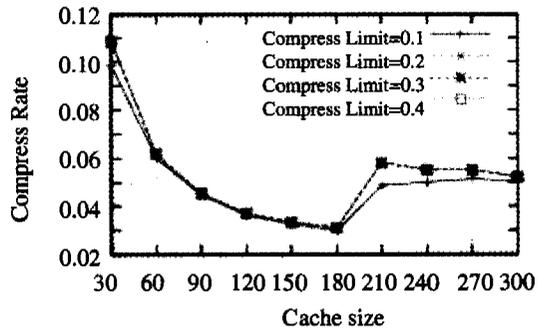


(b) Air Temperature

图2 空气温度(Air Temperature)在不同设置下的压缩传输情况及其相对误差



(a) Air Temperature



(b) Pressure

图3 空气温度(Air Temperature)和压力(Pressure)流数据的压缩比和传感器容量的关系

(上接第 63 页)

所有无线网卡传给此模块的帧必须经过帧格式的转换。同时在桥接模块中不区分是无线网卡来的帧还是有线网卡来的帧,经过桥转发的帧也不区分是转发给有线网卡的还是无线网卡的。

4.2 帧格式转换处理模块

802.11b 帧的帧格式与有线帧的帧格式是不同的,因此网关必须具有将无线与有线这两种不同的帧格式进行互相转换的能力,以实现无线网络与有线网络之间的通信。帧格式转换处理模块就完成了这一功能。

图 4 给出了从有线帧格式到 IEEE 802.11b 帧格式转换的流程图。

结论 本文采用模块设计,网关协议转换等办法解决数据在 LAN 之间发送、接收的问题。本文研究及设计的无线网络网关有以下特点:

对小区内移动终端的管理,例如登录、认证;

完成 IEEE 802.11b 中入口“Portal”的功能,实现从 802.11b 到 802.x 帧格式和从 802.X 到 802.11 帧格式的转换;

完成数据帧从无线网络到有线网络的桥接过程,实现地址过滤以及地址的学习功能。

参考文献

1 王宇. 无线网络基本原理及前沿应用. 中国数据通信, 2003(8)
 2 吴盘龙. 无线局域网技术及组网方式. 世界电信, 2001(12)
 3 萧文龙. 最新 TCP/IP 实用教程. 中国铁道出版社, 2001
 4 Patil B, 等著. 无线网络中的 IP. 张传福, 彭灿译. 人民邮电出版社, 2002
 5 Stevens W R 著. TCP/IP Illustrated Volume: 1-3. 胡谷雨, 谢希

仁, 等译. 机械工业出版社, 2002

6 金纯, 陈林星编著. IEEE802.11 无线局域网. 电子工业出版社, 2004

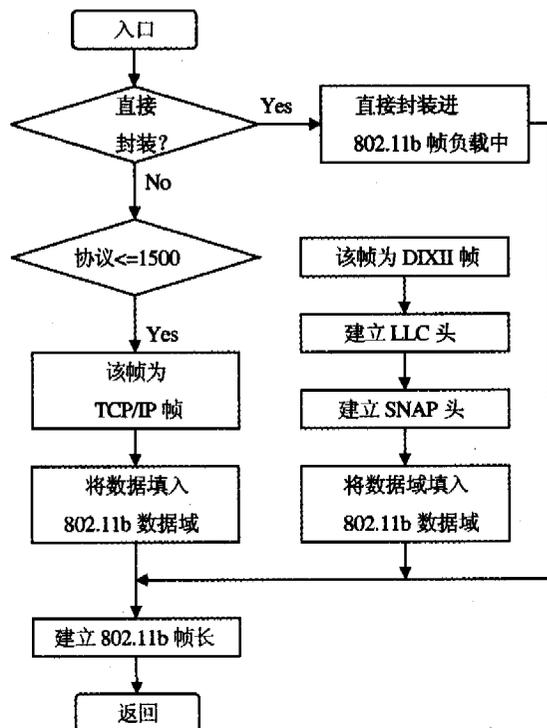


图4 从有线帧格式到 802.11b 帧格式的转换