

基于移动最小二乘法无线传感器网络数据 压缩算法研究

杨建翠¹ 李 英²

(1. 江苏医药职业学院 图书信息中心 江苏 盐城 224005;
2. 广东科学技术职业学院 广东 珠海 519090)

摘 要: 针对分布密集的数目庞大的节点使得传感器网络内部产生了大量数据,本文深入分析了 ODLRST 压缩算法存在的不足,主要是分析其在数据处理方面产生的误差问题,并针对这一问题,对算法进行了改进,充分考虑节点检测数据的各种波动情况,把节点的观察数据归类为可去间断点、连续间断点、跳跃间断点;而且阈值的取值区间更宽泛,从采集数据彼此的时间相关度出发,完成数据的线性拟合;对符合质量标准的精确算子与准确算子进行定义,利用质量性能的认识,确保在实施数据压缩时,一方面数据的质量符合需要,另一方面还要兼顾节能的需要。仿真实验结果表明本文提出的数据压缩算法对无线传感器网络传输的数据压缩效果显著。

关键词: 最小二乘法; 数据压缩; 时间序列; 断点

中图分类号: TP301.6

文献标识码: A

文章编号: 1001-7119(2018)01-0176-04

DOI: 10.13774/j.cnki.kjtb.2018.01.036

Research on Data Compression Algorithm for Wireless Sensor Networks Based on Moving Least Square Method

Yang Jiancui, Li Ying

(1. Book Information Center, Jiangsu Vocational College of Medicine, Jiangsu Yancheng 224005, China;
2. Guangdong Institute of Science and Technology, Guangdong Zhuhai 519090, China)

Abstract: According to the number of nodes densely distributed because of the huge sensor network generates large amounts of data, this paper analyzes the deficiency of existing ODLRST compression algorithm, error analysis is mainly the problems produced in the aspect of data processing, and to solve this problem, an improved algorithm, fully considering the fluctuation of detection data node the observation data, the node can be classified as to break point and break point, continuous jump discontinuity; and the broader range of threshold, from the time of collecting data related to each other, according to the number of complete linear fitting; precise operators comply with the quality standards of definition and accurate recognition of quality performance by operator. To ensure that, in the implementation of data compression, data quality on the one hand to meet the needs, on the other hand, but also take into account the need for energy saving. Simulation results show that the data compression algorithm proposed in this paper has a significant effect on the data compression of wireless sensor networks.

收稿日期: 2017-02-20

基金项目: 珠海市金湾区“十三五”社科规划(201632)。

作者简介: 杨建翠(1982-)女,汉族,江苏盐城人,讲师,硕士,从事计算机教育教学研究、计算机网络与通信技术研究。



Keywords: least square method; data compression; time series; break point

0 引言

在应用环境中,人们通常利用无线传感器网络来获得环境中传感器收集的环境数据信息^[1],普通节点按照一定的采集周期将所需的数据信息传输存入 Sink 节点,数据在进行传输时需要进行压缩处理,为了保证数据的时效性与准确性,确保基站获取更为可靠的数据信息,常用的有损压缩算法无法满足上述需要,应选用无损压缩算法以满足数据完整传输给基站的要求^[2]。

目前数据压缩算法是众多的学者专家研究的重点^[3]。Deligiannakis A 等提出了一种基于分段线性回归模型的数据压缩算法^[4]。该算法通过对节点接收到的数据进行分段处理,从处理后的数据中选出基础信号数据,通过线性回归推导出回归方程系数,将基础信号数据与回归系数方程相结合大大减小了无线传感器网络数据的传输。文献[5-7]利用基于数据时间相关的线性回归模型提出了不同的数据压缩算法,有效缓解了无线传感器网络数据传输的压力。杨圣洪等^[8]考虑到节点收集到的数据前后具有时间相关性以及空间相关性,提出了一种基于环模型的分布式时空小波数据压缩算法,大大扩展了该数据压缩算法的应用场景。李杨^[9]等提出了一种基于小波分段常值压缩的数据压缩算法,该算法首先在节点处对采集到的数据进行预压缩处理,然后通过分段化方法来减小网络数据传输量。文献[10]提出了一种基于小波压缩层次的无线传感器网络数据压缩算法,该算法考虑到数据时空相关性以及节点层数对数据时间性的影响,通过双重数据处理来减小数据传输量。侯鑫提出了基于神经网络的数据压缩算法^[11]。杨永健等提出了基于卡尔曼滤波模型的无线传感器网络数据算法^[12],两种算法都大大缓解了无线传感器网络数据传输的压力。

综上所述,本文给出一种新的无线传感器数据聚集压缩算法,这种算法极为接近无损压缩,并且对采样环境的要求更为宽泛,能够适用于嘈杂的监测环境,无需再定义其阈值区间,能够实现对各种采样条件下相异波动采样数据的压缩处理,对无线传感器网络数据传输的压缩效果显著。

1 拟合模型

对于无线传感器网络而言,其极易受到外界环境、噪声等干扰,这使得环境中的真实数据出现不稳定的波动,同时对于上述网络而言,其自身内存容量并不大,在执行预测以及保存的操作时,系统的采样数据必然会遭受影响,为了有效降低最小二乘法拟合带来的偏差,确保拟合曲线更具光滑性、紧致性,选择移动最小二乘法实施线性拟合,尽可能的降低预测结果与实际数据的偏差,确保其拟合精准度、可靠性满足系统需要。

1.1 拟合原理

将采样序列定义为 $(D, T) = \{(t_1, d_1), (t_2, d_2), \dots, (t_n, d_n)\}$ 用 $f(t)$ 代表拟合函数,那么其计算公式为:

$$f(t) = \sum_{j=1}^n a_j(t) p_j(t) = p^T(t) a(t) \quad (1)$$

式中: $a(t) = [a_1(t), a_2(t), \dots, a_m(t)]^T$ 待求系数, $p^T(t) = [1, t, t^2, \dots, L]$ 基函数

为了保证预测值与实际值的偏差尽可能的缩小,获得更为可靠的近似值,要满足两者偏差的二次方符合带权最小的条件,即:

$$J = \sum_{j=1}^n \omega(t - t_j) [f(t_j) - d_j]^2 = \sum_{i=1}^n \omega(t - t_i) [p^T(t) a(t) - d_i]^2 \quad (2)$$

上式中: $t = \|T\|_2 = (\sum_{i=1}^n t_i^2)^{1/2}$, 即为时间向量 $T = [t_1, t_2, \dots, t_n]$ 的 L_2 范式。

$\omega(t - t_i)$ 为时刻 t_i 的权函数,满足光滑性与紧致性的要求^[2],定义

$$s = t - t_i, \eta = s/s_{\max}$$

那么 $\omega(t - t_i)$ 符合以下条件:

$$\omega(\bar{s}) = \begin{cases} 2/3 - 4\eta^2 + 4\eta^3 & (\eta \leq 1/2) \\ 4/3 - 4\eta + 4\eta^2 - 4/3\eta^3 & (1/2 < \eta \leq 1) \\ 0 & (\eta > 1) \end{cases} \quad (3)$$

1.2 模型求解

计算公式

$$J = \sum_{i=1}^n \omega(t - t_i) [f(t_i) - d_i]^2 =$$



$$\sum_{i=1}^n \omega(t-t_i) [p^T(t) a(t) - d_i]^2 \quad (4)$$

当 $a(t) = [a_1, a_2, L, a_m(t)]^T$ 时, J 的取值最小。

对 J 依次计算 $a_i(t)$ 的偏导数, 即:

$$\begin{cases} \frac{\partial J}{\partial a} = A(t) a(t) - B(t) y = 0 \\ a(t) = A^{-1}(t) B(t) y \end{cases} \quad (5)$$

上式中: $A(t) = \sum_{i=1}^n \omega(t-t_i) p(t_i) p^T(t_i)$, $B(t) = (\omega(t-t_i) p(t_i))_{i=1}^n$ 。

那么, 回归方程表达式为:

$$f(t) = \sum_{i=1}^n \Phi_i^k(t) d_i = \Phi^k(t) d \quad (6)$$

上式中, $\Phi(t) = [\Phi_1^k, \Phi_2^k, L, \Phi_n^k] = p^T A^{-1}(t) B(t)$ 。

2 算法仿真与分析

数据压缩算法在对曲线实施拟合处理时, 选择了移动最小二乘法, 将预期趋势计算在内, 包括异常状况的预测趋势, 在确保预测数据的真实性方面, 充分考虑准度与精度要求, 确保预测数据与真实数据的误差尽可能的小, 有效保证了数据聚集压缩算法对其他监测环境的适用性问题, 文献^[8]就 ODLRST 压缩算法与 DWT 算法、NDA 算法的计算性能进行了对比, 证明了这一算法的优势所在, 鉴于此, 本文只就数据聚集压缩算法与上述算法进行计算性能的对比。

对比最小二乘法的曲线拟合算法, 许多文献详细分析了移动最小二乘法的曲线拟合的优势所在, 不再累述, 本文仅就压缩性能方面做对比, 参阅下表:

表 1 非波动数据压缩性能结果

Table 1 Performance results of non volatility data compression

压缩方式	压缩因子 CF	均方差 MSE	峰值信噪比 PSNR
聚集压缩算法	CF = 16%	MSE = 0.4307	PSNR = 81.1965
ODLRST 算法	CF = 4%	MSE = 2.8282	PSNR = 62.3766

参阅上表, 两种算法的压缩结果显示, ODLRST 算法在压缩比上占据优势, 但是预测结果的误差也相对放大许多, 监测效果不好, 准确度的误差会让监测的原始需求失真, 数据可靠性较差。很明显, 尽管聚集压缩算法的压缩比稍低, 前者是后者的 4 倍, 但是预测的结果却比前者精确的多, 特别是均方差, 后者时前者的 6 倍之多, 有效保证了监测数据的真实

可靠性, 将 ODLRST 算法的误差问题有效解决。

利用系统仿真, 两种算法的压缩性能对比结果参阅表 2。

表 2 波动数据压缩性能结果

Table 2 Performance results of volatility data compression

压缩方式	压缩因子 CF	均方差 MSE	峰值信噪比 PSNR
聚集压缩算法	CF = 25%	MSE = 1.1684	PSNR = 57.5282
ODLRST 算法	CF = 10%	MSE = 3.0418	PSNR = 47.7643

从表 2 结果可以得出, 数据聚集算法作为改进的算法仅在压缩性能方面表现稍差, 但是在均方差、峰值信号比方面均表现出色。无线传感器网络监测环境的初始目的就是要获得更为真实、可靠的数据, 较大的误差使得数据不具备采样价值。通过仿真, 可以明显的看出数据聚集压缩算法能够较为出色的完成数据还原的任务, 在数据的精确度、准确度方面相比简单的使用最小二乘法进行拟合的 ODLRST 压缩算法, 数据结果更符合需求, 特别是其完全能够应对数据的异常波动的情况, 在有效保证一定压缩比的情况下, 使得预测数据的精确度、准确度大幅提高, 应用效果极佳, 再一次证明了数据聚集压缩算法的先进性。

3 结语

在数据处理方面, 本文深入分析了 ODLRST 压缩算法存在的不足, 主要是分析其在数据处理方面产生的误差问题, 并针对这一问题, 对算法进行了改进, 提出一种新的数据聚集压缩算法。数据聚集压缩法不但可以完成预期趋势变化的数据处理任务, 对那些处于嘈杂环境、数据伴有异常波动的情况也能够适应, 可以认为是一种典型的无损数据压缩算法。通过系统仿真, 可以明显的看出数据聚集压缩算法能够较为出色的完成数据还原的任务, 在数据的精确度、准确度方面相比 ODLRST 压缩算法更满足系统需求, 特别是其完全能够应对数据的异常波动的情况, 在有效保证一定压缩比的情况下, 使得预测数据的精确度、准确度大幅提高, 完全规避了 ODLRST 压缩算法存在较大误差的问题。

参考文献:

- [1] 黄鹏, 李占山, 张永刚, 等. 基于列存储数据库的压缩态数据访问算法[J]. 吉林大学学报: 理学版, 2009, 47(5): 1013-1019.
- [2] Steinbach E, Hirche S, Kammerl J, et al. Haptic Data



- Compression and Communication [J]. IEEE Signal Processing Magazine, 2015, 54(38): 11209–11213.
- [3] 费贤举, 李晓芳. 基于压缩感知理论的无线传感器网络数据融合算法[J]. 吉林大学学报: 理学版, 2016, 54(3).
- [4] Deligiannakis A, Kotidis Y, Roussopoulos N. Compressing historical information in sensor networks [J]. Sigmod, 2004: 527–538.
- [5] 王玲, 石为人, 石欣, 等. 基于时间相关性的无线传感器网络数据压缩与优化算法[J]. 计算机应用, 2013, 33(12): 3453–3456.
- [6] 王举, 房鼎益, 陈晓江, 等. 文物监测中无线传感器网络数据压缩算法[J]. 西安电子科技大学学报(自然科学版), 2012, 39(1): 157–162.
- [7] 宋欣, 王翠荣. 基于线性回归的无线传感器网络分布式数据采集优化策略[J]. 计算机学报, 2012, 35(3): 568–580.
- [8] 杨圣洪, 贾焰, 周四望. 无线传感器网络基于虚拟节点的小波压缩方法[J]. 软件学报, 2013(3): 557–563.
- [9] 李杨, 郭龙江, 李金宝, 等. 传感器网络基于小波分段常值压缩的数据收集研究[J]. 仪器仪表学报, 2013, 34(1): 119–127.
- [10] Ganesan D, Estrin D, Heidemann J. DIMENSIONS: Why do we need a new Data Handling architecture for Sensor Networks? [J]. in Proceedings of the First Workshop on Hot Topics In Networks (HotNets-I, 2002, 33(1): 143–148.
- [11] 侯鑫, 张东文, 钟鸣. 基于事件驱动和神经网络的无线传感器网络数据融合算法研究[J]. 传感技术学报, 2014(1): 142–148.
- [12] 杨永健, 刘帅. 基于 SOFM 神经网络的无线传感器网络数据融合算法[J]. 传感技术学报, 2013(12): 1757–1763.

(上接第 175 页)

4 结论

本文提出一种基于节点团连通度和虚亏性的混合网络不确定入侵目标检测方法, 将混合网络中的一个网段以及连接到该网段上的工作站看成一个节点团, 基于混合网络中全部节点团间可能通信路径的平均数, 获取混合网络连通度, 基于节点团的虚亏性过滤薄弱节点团, 依据混合网络节点团的连通度和虚亏性指标, 得到混合网络的安全度指标, 实现混合网络安全性的有效评估. 仿真结果说明, 该方法检测混网网络中不确定入侵目标是有效的, 提高网络系统的安全性, 并且具有较高的优越性.

参考文献:

- [1] 邹洪, 龙震岳, 陈力. 混合无线网络区域的入侵主动防御模型仿真[J]. 计算机仿真, 2016, 33(1): 280–283.
- [2] 傅明丽. 网络安全中混合型入侵检测系统设计[J]. 通讯世界, 2016, 32(1): 226–227.
- [3] 王江峰. 混合网络入侵特征的高效分类模型研究与仿真[J]. 科教导刊, 2015, 25(18): 50–51.
- [4] 赵夫群. 基于混合核函数的 LSSVM 网络入侵检测方法[J]. 现代电子技术, 2015, 38(21): 96–99.
- [5] 李岳云, 许悦雷, 马时平, 等. 深度卷积神经网络的显著性检测[J]. 中国图象图形学报, 2016, 21(1): 53–59.
- [6] 韩德强, 程博, 杨艺. 一种基于证据推理的视频运动目标融合检测算法[J]. 控制与决策, 2016, 31(8): 1345–1353.
- [7] 张宏, 李杰. 复杂网络的微信网络信息传播研究[J]. 科学技术与工程, 2014, 14(17): 244–247.
- [8] 单外平, 曾雪琼. 基于深度信念网络的信号重构与轴承故障识别[J]. 电子设计工程, 2016, 24(4): 67–71.
- [9] 杨雷, 李贵鹏, 张萍. 改进的 Wolf 一步预测的网络异常流量检测[J]. 科技通报, 2014, 23(2): 47–49.
- [10] 李强懿, 马冬前, 张聚伟. 基于感知概率的无线传感器网络节点部署算法[J]. 计算机测量与控制, 2014, 22(2): 643–645.