

基于 SPC 的农产品冷链物流感知数据压缩方法^{*}

齐林¹ 田东² 张健³ 张小栓² 傅泽田¹

(1. 中国农业大学工学院, 北京 100083; 2. 中国农业大学信息与电气工程学院, 北京 100083;
3. 北京信息科技大学经济管理学院, 北京 100192)

【摘要】 针对无线传感器网络实时监控的农产品冷链物流环境中, 传感器节点的感知数据传输频率高、能耗高和监测时间短的问题, 提出基于统计过程控制(SPC)的感知数据压缩方法, 设计了适用于传感器节点数据处理的动态-单值-滑动极差($D-X-R_s$)算法。通过冷链物流环境试验仿真, 对比了 $D-X-R_s$ 算法、阈值算法和K-滑动均值算法的数据压缩性能。结果表明, $D-X-R_s$ 算法数据的每数据点压缩开销(CCP)在温度稳定和波动状态均与最优方法所得结果接近, 具有良好的平衡性和适应性。

关键词: 冷链物流 无线传感器网络 数据压缩 统计过程控制

中图分类号: S126; TS205.7 **文献标识码:** A **文章编号:** 1000-1298(2011)10-0129-06

Sensing Data Compression Method Based on SPC for Agri-food Cold-chain Logistics

Qi Lin¹ Tian Dong² Zhang Jian³ Zhang Xiaoshuan² Fu Zetian¹

(1. College of Engineering, China Agricultural University, Beijing 100083, China

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

3. School of Economics and Management, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract

In order to reduce the sensing data transfer frequency and node energy consumption in real-time cold-chain logistics monitoring system for agri-food based on wireless sensor networks, a sensing data compression method based on statistics process control was presented. The $D-X-R_s$ algorithm that fits for the sensor nodes' data processing capability was designed. The data compression performance of $D-X-R_s$ algorithm, threshold algorithm and K-means algorithm by cold-chain logistics environmental simulation in laboratory was compared. The result showed that the compression cost per point (CCP) of $D-X-R_s$ algorithm was close to the best algorithm both in temperature fluctuation and steady states. The algorithm had good balance and flexibility.

Key words Cold-chain logistics, Wireless sensor networks, Data compression, Statistics process control

引言

中国是农产品生产和消费大国, 但由于 80% 以上的生鲜农产品物流依靠常温实现, 使我国农产品供应链各环节问题频发。农产品冷链物流作为使生鲜农产品从产地采收后, 在产品加工、贮藏、运输、分

销、零售等环节始终处于适宜的低温控制环境下, 最大程度保证产品品质和质量安全、减少损耗、防止污染的特殊供应链系统, 是保障食品安全、降低经济损失、提高生鲜农产品市场竞争力的必然发展需求^[1]。

我国冷链物流建设处于起步阶段, 要实施现代

收稿日期: 2010-12-20 修回日期: 2010-12-27

^{*} 国家自然科学基金资助项目(71072153)、中央高校基本科研业务费专项资金资助项目(2010JS101)、上海市教委重点学科食品经济管理资助项目(J50703)和中国农业大学研究生科研创新专项资助项目(KYCX2011067)

作者简介: 齐林, 博士生, 主要从事农业系统工程研究, E-mail: kylin_xi@cau.edu.cn

通讯作者: 傅泽田, 教授, 博士生导师, 主要从事农业工程研究, E-mail: fzt@cau.edu.cn

冷链物流过程中的温度监测感知数据具有 5 个特点: 采样值是物理量; 单次采样结果为单值数据; 采样具有时间上的不可逆性; 单个传感器采样无数据分层; 需要尽可能早的发现温度状态变化。基于以上分析, 本文选用单值-滑动极差控制图作为数据压缩算法的核心。

控制图中的温度状态异常通常有两种表现形式, 即监测点越界和监测点的不随机排列。应用控制图对温度状态进行判异的过程, 以小概率事件原理为依据。GB/T 4091—2001 标准根据这一原理, 确定了 8 种判异模式: ① 1 个点落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 以外 (采样点超出控制界限)。② 连续 9 点落在中心线同一侧。③ 连续 6 点递增或递减。④ 连续 14 点中相邻两点上下交替。⑤ 连续 3 点中有 2 点落在中心线同一侧的 $(\mu - 2\sigma, \mu + 2\sigma)$ 以外。⑥ 连续 5 点中有 4 点落在中心线同一侧 $(\mu - 3\sigma, \mu + 3\sigma)$ 以外。⑦ 连续 15 点全部在中心线两侧 $(\mu - 2\sigma, \mu + 2\sigma)$ 区以内。⑧ 连续 8 点在中心线两侧, 但无 1 点在 $(\mu - 2\sigma, \mu + 2\sigma)$ 区内。

其中模式①为最重要的判异准则; 模式②、③为①的补充, 反映了均值 μ 可能变化; 模式④、⑧反映数据可能存在分层; 模式⑤、⑥可认为是①的推广; 模式⑦反映了方差 σ 的收敛。

综合考虑感知数据特点、传感器节点的计算能力、各判异准则在程序中实现的时空复杂度和相应事件发生概率, 本文选取模式①、②和③作为温度状态监测的判异准则, 并用于动态-单值-滑动极差 (Dynamic-X-R_s, 简称 D-X-R_s) 压缩算法的构造。

1.4 动态-单值-滑动极差压缩算法

算法的应用须符合:

条件 1: 存在 n 个按时序采集的某一传感器温度监测数据 $t_1 \sim t_n$ 。

条件 2: 存在规则 R_1, R_2 和 R_3 , 分别对应 GB/T 4091—2001 的判异准则①、②和③。

条件 3: 存在判异算子 m_1, m_2 和 m_3 , 类型均为逻辑变量, 初始值均为 false (即认为温度处于稳定状态)。

条件 4: 存在逻辑变量 $s, s = m_1 \vee m_2 \vee m_3$ 。

算法使用 nesC 语言编写, 关键步骤如下:

- (1) INPUT($t_1 \sim t_n$);
- (2) CONSTRUCT (R_s-Chart),
CONSTRUCT (X-Chart);
- (3) FORM $i = 1$ TO $n - 1$ SET $t_i = t_{i+1}$,
INPUT (t_n), SET
 $m_1 = m_2 = m_3 = \text{false}$;

(4) ON R_s-Chart:

$$\{t_i; t_n\} \xrightarrow{R_1, R_2, R_3} \{m_1, m_2, m_3\}, \text{SET}$$

$$s = m_1 \vee m_2 \vee m_3;$$

(5) IF $s = \text{true}$ {TRANSMIT (t_n), GOTO (3)};

(6) ON X-Chart:

$$\{t_i; t_n\} \xrightarrow{R_1, R_2, R_3} \{m_1, m_2, m_3\}, \text{SET}$$

$$s = m_1 \vee m_2 \vee m_3;$$

(7) IF $s = \text{true}$ {TRANSMIT (t_n), GOTO (8)}
ELSE GOTO (3);

(8) IF $m_2 = \text{true}$ GOTO (2) ELSE GOTO (3);

对算法的解释: 算法开始时首先接收 n (基于判异准则②的统计数据数量, 本文取 $n = 9$) 个温度数据 $t_1 \sim t_n$, 并用这些数据计算统计量、控制限, 构造 X 控制图和 R_s 控制图; 构造完成后抛弃最旧的一个温度数据, 开始一次采样, 重置所有判异算子; 对当前的 n 个温度数据进行 R_s 控制图判异, 判异成功应向基站发送该温度数据并开始下一次采样, 判异失败则继续下一步; 对当前的 n 个温度数据进行 X 控制图判异, 判异失败则开始下一次采样, 判异成功则需发送该数据并分析, 如因准则②导致判异成功, 说明温度均值 \bar{t} 发生偏移, 温度状态迁移至另一稳态, 应重置控制图。其余情况导致的判异成功表示温度状态正在发生迁移, 尚未形成稳态, 不应急于重构控制图。

该算法在 ATmega 1281 芯片上运行, 由于 WSNs 节点程序为单任务无限循环程序, 因此算法仅在芯片掉电后才终止运行。该算法流程图如图 3 所示。

2 试验过程与结果

2.1 试验设备及数据采集

冷链温度环境使用天津苏瑞科技有限公司 TEM1880 型高低温交变试验箱 (以下称变温箱) 模

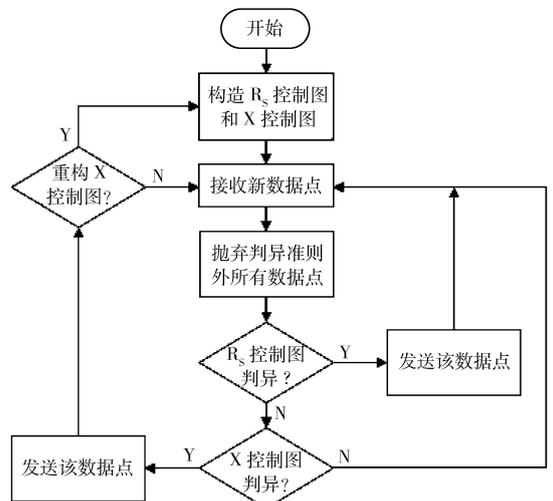


图 3 D-X-R_s 算法流程图

Fig. 3 Flow chart of D-X-R_s algorithm

拟。箱体内部温度数据使用美国 Crossbow 公司 MDA300 型数据采集板采集,板载 SHT15 型数字温湿度传感器在(-40℃~123.8℃)温度区间内分辨率为0.1℃,精确度为±0.3℃。感知数据无线传输使用 Crossbow 公司的 IRIS(XM2110)型节点,采用2.4 GHz 的 ISM 频段进行数据传输。

试验在中国农业大学信息与电气工程学院进行。试验时设置变温箱为空载运行,在60℃运行60 min以干燥箱体内部。安排2组共计1000 min 的温度数据采集,即设定变温箱在-4℃运行400 min,在-4~10℃区间内交替运行600 min,以分别模拟冷链物流过程中的稳定温度状态和波动温度状态。

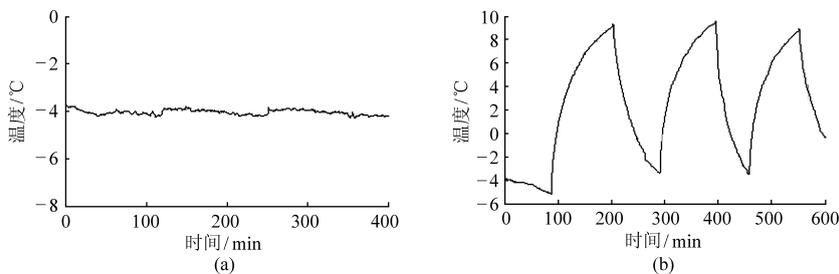


图4 采样温度状态分布

Fig.4 Sampled temperature states distribution

(a) 稳定温度状态 (b) 波动温度状态

2.2 数据压缩算法评价模型构建

对于同一数据分布序列,数据压缩比高的算法能够更大程度地减少数据冗余,但高的压缩比更易造成有用信息丢失,导致较大的数据拟合误差,从而对波动温度下的生鲜农产品货架期预测精度产生不利影响。考虑到运行在 WSNs 节点上,算法的时间复杂度 $T(o)$ 、空间复杂度 $S(o)$ 、算法的能耗 C 、数据压缩比 R 和数据拟合结果的残差平方和 S_e 等指标都应该予以考虑。

由于时间复杂度 $T(o)$ 、空间复杂度 $S(o)$ 是算法的固有属性,不可量化,本文采用静态分析方法,对阈值算法、K-滑动均值算法、D-X-R_s 算法的 $T(o)$ 和 $S(o)$ 给出对比。对于能耗、数据压缩比和拟合精度,采取动态仿真分析的方法,引入试验数据,分别求解各算法的每数据点压缩开销 (compression cost per point, 简称 CCP),计算方法为

$$t_c = f(C, R, S_e, N) = \frac{CS_e}{RN} \quad (3)$$

式中 N ——序列中数据点个数

本文认为,一个性能好的算法应该同时具有低的计算时空复杂度和小的 t_c 值。

2.3 算法性能分析

2.3.1 算法静态分析

在无数据压缩的情况下,节点采样后压缩运算

采样间隔均为10 s。温度状态设定如表1所示,采样温度的实际分布如图4所示。

表1 温度状态设定

Tab.1 Temperature status settings

采样时段/min	目标温度状态/℃
0~400	-4
400~500	10
500~600	-4
600~700	10
700~800	-4
800~900	10
900~1000	-4

开销的时间复杂度和空间复杂度均是零,但同时数据冗余最大,传输能耗开销最大;采用阈值算法,采样数据需要与触发数据传输的温度波动阈值进行比较,阈值需要存储,算法的时间、空间复杂度为常数阶,即 $T(o) = O(1)$, $S(o) = S(1)$;对于 K-滑动均值算法,需要存储滑动窗口长度 k 个变量,每 k 次采样求一次均值并传输,算法的时间、空间复杂度亦为常数阶;D-X-R_s 算法每个采样点至多进行两轮判异,存储 n (本文 $n=9$) 个历史数据和2个控制图的 L_c 、 L_u 和 L_l ,算法的时间、空间复杂度均为常数阶。各算法的时间、空间复杂度对比如表2所示。

表2 3种算法的时间复杂度和空间复杂度

Tab.2 Time complexity and space complexity of three algorithms

算法	$T(o)$	$S(o)$
阈值算法	$O(1)$	$S(1)$
K-滑动均值算法	$O(1)$	$S(1)$
D-X-R _s 算法	$O(1)$	$S(1)$

2.3.2 算法动态分析

为估算算法能量消耗,使用一阶无线模型 (first order radio model)^[16] 进行 Matlab 仿真。该模型指出射频芯片发送 k 位数据到距离 d 所消耗电能为

$$E_i(k, d) = E_e k + \varepsilon_a k d^2 \quad (4)$$

式中 E_e ——发送器能耗 ε_a ——信道能耗

已知 $E_e = 50 \text{ nJ/bit}$, $\varepsilon_a = 100 \text{ pJ} \cdot \text{m}^2/\text{bit}$, ATmega 系列处理器的指令能耗为 5 nJ , 使用以上量化结果, 通过统计在一段监控时间中, 算法指令执行次数及采样值发送次数, 可估算节点运行压缩算法的能耗。

算法的仿真评估分别在稳态数据和波动数据的条件下进行。由于 D-X-R_s 算法中的控制限由 3σ 准则确定, 而算法的压缩比仅由控制限和数据分布决定, 因此在确定的温度数据分布下, D-X-R_s 算法具有确定的压缩比。为使算法之间的动态性能具有可比性, 本文通过调整阈值算法的触发数据传输阈值和 K-滑动均值算法的滑动窗口大小, 使阈值算法和 K-滑动均值算法分别在稳态和波动状态下具有与 D-X-R_s 算法近似的压缩比。

表 3 给出了在 0~400 min 的稳定温度状态下采集的 2 462 个温度数据进行算法仿真得到的性能数据。D-X-R_s 方法数据压缩比为 90.0%, 取数据传输阈值为 $0.025 \text{ }^\circ\text{C}$ 、滑动窗口长度为 10, 使阈值算法和 K-滑动均值算法的压缩比近似为 90.0%。

表 3 稳定温度状态的算法性能 ($N = 2\ 462$)

Tab. 3 Algorithm performance in steady temperature state

算法	能耗 C/nJ	压缩比 $R/\%$	误差平方和 $S_e/^\circ\text{C}^2$	t_c $/\text{J} \cdot ^\circ\text{C}^2$
阈值算法	0.938×10^6	90.5	1.629	6.859
K-滑动均值算法	0.986×10^6	90.0	1.191	5.299
D-X-R _s 算法	1.189×10^6	90.0	1.120	6.015

图 5 给出了在稳定温度状态下的原始温度数据的随机波动分布趋势和压缩数据的拟合情况。

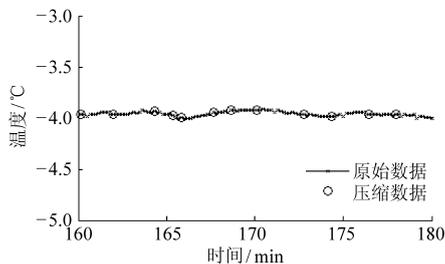


图 5 稳定温度状态的压缩数据拟合

Fig. 5 Compression data fitting map in steady temperature state

表 4 给出了在 400~1 000 min 波动温度状态下采集的 3 863 个温度数据进行算法仿真得到的性能数据。D-X-R_s 算法在这一状态下的数据压缩比为 86.5%, 取数据传输阈值为 $0.12 \text{ }^\circ\text{C}$ 、滑动窗口长度为 7, 使阈值算法和 K-滑动均值法的压缩比近似为 86%。

表 4 波动温度状态的算法性能 ($N = 3\ 863$)

Tab. 4 Algorithm performance in fluctuation temperature state

算法	能耗 C/nJ	压缩比 $R/\%$	误差平方和 $S_e/^\circ\text{C}^2$	t_c $/\text{J} \cdot ^\circ\text{C}^2$
阈值算法	2.133×10^6	86.2	32.670	209.304
K-滑动均值算法	2.213×10^6	85.7	77.465	517.999
D-X-R _s 算法	2.302×10^6	86.5	42.682	294.057

图 6 为波动温度状态下, 原始温度数据的波动分布趋势和压缩数据的拟合情况。图 7 为在波动温度状态下, 压缩数据在温度剧烈变化的顶点区域失去了与原始数据的拟合, 这是由 D-X-R_s 算法的判异流程导致的。

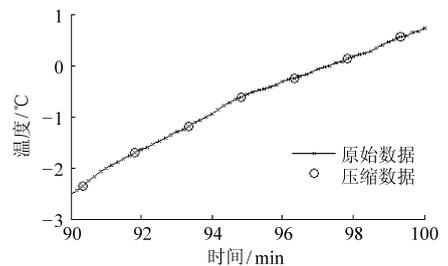


图 6 波动温度状态的压缩数据拟合

Fig. 6 Compression data fitting map in fluctuation temperature state

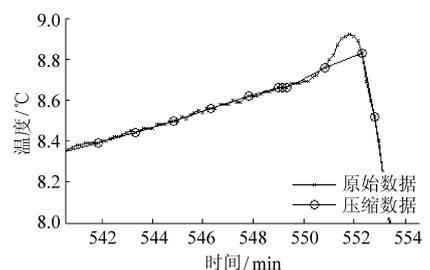


图 7 剧烈温度波动下的压缩数据失拟

Fig. 7 Compression data fitting failure in severe temperature fluctuations

3 结论

(1) 由于 D-X-R_s 算法具有较复杂的判异功能, 在两种状态下, 对温度数据的传输决策判断过程的能耗均比阈值算法和 K-滑动均值算法大, 但由于数据传输部分所占的能耗开销是节点主要能耗开销, 因此在压缩比相同的前提下, 3 种算法的能耗均处于同一数量级上。

(2) 在稳定温度状态下, D-X-R_s 算法通过不断的进行判异决策获得了最高的数据拟合精度, 误差平方和为 1.120。在波动状态下, D-X-R_s 算法的误差平方和为 42.682, 大于阈值算法获得的 32.670。在温度趋势变化剧烈位置的失拟 (如图 7

折线顶点部分)是造成误差平方和增大的重要原因。判异准则的进一步完善和判异流程的优化是D-X-R_s算法性能进一步改善的关键。

(3) 在 t_c 值的对比上,阈值算法和K-滑动均

值算法分别显示出其对稳态和波动态的适应性,D-X-R_s算法在两种状态上并非最优,但 t_c 值均接近阈值算法和K-滑动均值算法所得结果,显示出了良好的性能平衡性和状态适应性。

参 考 文 献

- 1 鲍长生. 冷链物流运营管理研究[D]. 上海:同济大学,2007.
Bao Changsheng. On operational management of cold chain [D]. Shanghai:Tongji University, 2007. (in Chinese)
- 2 应晓书. 基于射频识别(RFID)技术在冷链物流中的应用研究[D]. 武汉:武汉理工大学,2008.
Ying Xiaoshu. Application research of RFID technology applied in cold chain logistics [D]. Wuhan:Wuhan University of Technology, 2008. (in Chinese)
- 3 Wang Ning, Zhang Naiqian, Wang Maohua. Wireless sensors in agriculture and food industry—recent development and future perspective[J]. Computers and Electronics in Agriculture, 2006, 50(1):1~14.
- 4 张荣标,余功江,张磊,等. WSN土壤湿度采集节点电压补偿方法研究[J]. 农业机械学报,2010,41(11):178~182.
Zhang Rongbiao, Yu Gongjiang, Zhang Lei, et al. Voltage compensation method of WSN soil moisture acquisition nodes[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(11):178~182. (in Chinese)
- 5 郭文川,程寒杰,李瑞明,等. 基于无线传感器网络的温室环境信息监测系统[J]. 农业机械学报,2010,41(7):181~185.
Guo Wenchuan, Cheng Hanjie, Li Ruiming, et al. Greenhouse monitoring system based on wireless sensor network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(7):181~185. (in Chinese)
- 6 Tan K K, Huang S N, Zhang Y, et al. Distributed fault detection in industrial system based on sensor wireless network[J]. Computer Standards & Interfaces, 2009, 31(3):573~578.
- 7 Vellidis G, Tucker M, Perry C, et al. A real-time wireless smart sensor array for scheduling irrigation[J]. Computers and Electronics in Agriculture, 2008, 61(1):44~50.
- 8 Nadimi E S, Sogaard H T. Observer kalman filter identification and multiple-model adaptive estimation technique for classifying animal behavior using wireless sensor networks[J]. Computers and Electronics in Agriculture, 2009, 68(1):9~17.
- 9 Pierce F J, Elliott T V. Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington[J]. Computers and Electronics in Agriculture, 2008, 61(1):32~43.
- 10 Estrin D. Wireless sensor networks tutorial part IV: sensor network protocols[R]. Georgia: Mobicom, 2002.
- 11 王继良. 无线传感器网络中的小波数据压缩算法研究[D]. 长沙:湖南大学,2007.
Wang Jiliang. A research on wavelet data compression algorithms for wireless sensor networks [D]. Changsha: Hunan University, 2007. (in Chinese)
- 12 刘少强,汪立林. 一种面向WSNs节点的数据压缩简化算法[J]. 传感技术学报,2009,22(9):1333~1336.
Liu Shaoqiang, Wang Lilin. A simplified algorithm of data compression for wireless sensor network node [J]. Chinese Journal of Sensors and Actuators, 2009, 22(9):1333~1336. (in Chinese)
- 13 尹震宇,赵海,徐久强,等. WSNs中基于分簇路由的多维度数据压缩算法研究[J]. 电子学报,2009,37(5):1109~1114.
Yin Zhenyu, Zhao Hai, Xu Jiuqiang, et al. Research on multi-dimensional data compression algorithm for cluster-based routing in wireless sensor network [J]. Acta Electronica Sinica, 2009, 37(5):1109~1114. (in Chinese)
- 14 范祥辉,李士宁,杜鹏雷,等. WSNs中一种自适应无损数据压缩机制[J]. 计算机测量与控制,2010,18(2):463~465.
Fan Xianghui, Li Shining, Du Penglei, et al. Simple algorithm for self-adapting lossless data compression in WSNs [J]. Computer Measurement & Control, 2010, 18(2):463~465. (in Chinese)
- 15 刘璐,岳峻,张健,等. 水产品冷链管理决策模型的构建[J]. 农业工程学报,2010,26(8):379~384.
Liu Lu, Yue Jun, Zhang Jian, et al. Construction of cold chain decision support system for aquatic products [J]. Transactions of the CSAE, 2010, 26(8):379~384. (in Chinese)
- 16 Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks [C]//Hyun-Kook K. Proceedings of the Hawaii International Conference on System Sciences. Berlin: Springer-Verlag, 2003:181~191.