

基于时间序列的自适应采样机制策略研究

杨明霞^{1,2} 王万良^{1,3} 邵鹏飞^{1,4}

(浙江工业大学信息学院 杭州 310023)¹ (衢州学院电气与信息工程学院 衢州 324000)²

(浙江工业大学计算机学院 杭州 310023)³ (浙江万里学院电子信息学院 宁波 315100)⁴

摘要 在传感器网络中,节点对同一事件采集的数据间存在一定的时空相关性。若有效利用数据相关性,动态调整采样间隔,则能够减少不必要的采样,从而相应地减少采样、计算、传输所消耗的能源,延长网络寿命。采用二次指数平滑法进行预测,参考TCP拥塞控制思想,快速调整采样间隔。实验证明,与普通算法相比,该算法能同时降低错误丢失率和采样率。

关键词 无线传感器网络,自适应采样,双指数平滑法,采样率,错误丢失率

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.7.035

Adaptive Sampling Algorithm Based on TCP Congestion Strategy

YANG Ming-xia^{1,2} WANG Wan-liang^{1,3} SHAO Peng-fei^{1,4}

(College of Information, Zhejiang University of Technology, Hangzhou 310023, China)¹

(College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China)²

(College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China)³

(College of Electronic Information, Zhejiang Wanli University, Ningbo 315100, China)⁴

Abstract We presented the design of a novel adaptive sampling technique based on TCP congestion strategy, in which the temporal data correlations provide an indication of the prevailing environmental conditions and are used to adapt to the sensing rate of a sensor node. It uses irregular data series prediction to reduce sampling rate in combination with change detection to maintain data fidelity. The prediction method employs Wright's extension to Holt's method of exponential double sampling (EDS) coupled with a change detection mechanism based on exponentially weighted moving averages (EWMA). The main advantages are that it does not require heavy computation, incurs low memory and communication overhead and the prediction model can be implemented with ease on resource constrained sensor nodes.

Keywords Wireless sensor networks, Adaptive sampling, Exponential double smoothing, Sampling fraction, Miss ratio

1 引言

WSNs的关键问题之一在于能源,传感器能耗主要由感知(采样)、计算、传输(radio) 3个部分组成,通常认为传输的能耗最关键,节点间通信所耗能量远远大于节点上工作的能量,因此大量研究集中在如何在节点上尽量处理好数据,以减少发送的次数和数据量,但随着研究和应用的深入和普及,发现很多实际应用中面临的问题并非如此,往往节点上采集和计算所消耗的能量也不能小觑^[1,2]。因此,节点上的相关设计对于延长网络的生命周期也同样重要。

文献[1]为监测公路一氧化碳浓度安放了传感器,为延长生命周期,基于双指数平滑法进行预测,提出了一种自适应调整采样间隔的EDSAS算法。文献[3]对能源受限的传感器能源管理策略进行了综述,将节能的策略分为3类:层次感知、自适应采样和基于模型的感知。文献[4]基于累积和的改

变检测技术提出了自适应采样的算法,动态估计信号当前最大频率并在当前最大频率超过阈值时提示改变,且引发更新采样率。该算法采用的是集中式处理方式,即更新算法在基站执行,计算出的采样率结果被发送给每个节点。所以该方法的可扩展性不太好,计算量比较大,通信负载大。另一种方法是统计协议。每个节点适应数据流的特征,自动决定采样率,在给定范围内使用卡尔曼滤波来估计错误。当出现问题时,从基站申请新的采样率。基站考虑现有的能源确定新的采样率,例如在所有活动节点上使用卡尔曼滤波来估计错误最小化,但计算较为复杂。

在传感器网络中,节点对同一事件采集的数据间存在一定时空相关性,若能利用这些相关性来分析数据的特性,则可以避免不必要的采样。在时间相关性方面,可以结合节点的历史读数记录,遵照某种评价指标来对采样进行反馈控制。简单来说,节点对比上一时刻数据,若数据没有发生改变,则

到稿日期:2014-07-01 返修日期:2014-12-16 本文受国家自然科学基金面上项目(61379123),浙江省自然科学基金(LQ12F03011, LQ14F020005, LY13F030011),宁波市自然科学基金(2012A610016),2013浙江省重点实验室开放基金项目(2013026),衢州学院师资队伍建设基金(XNZQN201308)资助。

杨明霞(1979-),女,硕士,讲师,主要研究方向为智能信息处理, E-mail: 37847098@qq.com; 王万良(1957-),男,教授,主要研究方向为智能控制; 邵鹏飞(1978-),男,副教授,主要研究方向为网络通信。

节点可以延长下一个采样时间。而在另外一些方案中,系统默认在一个相对较短的时间范围内,传感器节点所采集到的数据之间的浮动是不大的。依据一段时间的平均值来划定一个置信区间,以此来判断传感器读数是否处于正常范围,并进一步判断是否需要调整采样频率来适应变化。这些方案在相对稳定的环境中非常有效,在其采样协议中,各个节点根据读数的变化和下一组观测数据的预测来自行调节采样频率^[8]。

本文考虑在单个节点上基于时间序列相关性来调整采样率,目标是在减少采样次数的同时不丢失重要信息,甚至还要考虑提高数据特征提取的精确性。主要采用 Wright's 对 Holt's 双指数平滑法的改进方法进行预测,采用指数加权平均方法(Exponentially Weighted Moving Averages, EWMA)进行事件检测^[2],将算法性能与文献[1]的 EDSAS 算法的性能做比较。第 2 节介绍双指数平滑法;第 3 节是本文算法的描述;第 4 节给出仿真及结果分析;最后总结全文。

2 指数平滑法

指数平滑法是布朗(Robert G. Brown)提出的,布朗认为时间序列的态势具有稳定性或规则性,所以时间序列可被合理地顺势推延;他认为最近的过去态势,在某种程度上会持续到最近的未来,所以将较大的权数放在最近的资料。简单的全期平均法是对时间数列的过去数据一个不漏地加以同等利用;移动平均法则不考虑较久远的数据,并在加权移动平均法中给予近期资料更大的权重;指数平滑法则兼容了全期平均和移动平均的优势,不舍弃过去的资料,但是仅给予逐渐减弱的影响程度,即随着数据的远离,赋予逐渐收敛为零的权数。

已知时间序列为: x_1, x_2, \dots, x_n , n 为时间序列总期数,指数平滑法的基本公式是:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}$$

式中, S_t 为 t 时刻的平滑值; y_t 为 t 时刻的实际值; S_{t-1} 为 $t-1$ 时刻的平滑值; α 为平滑常数,其取值范围为 $[0, 1]$ 。平滑常数 α 越接近于 1,远期实际值对本期平滑值影响程度的下降越迅速;平滑常数 α 越接近于 0,远期实际值对本期平滑值影响程度的下降越缓慢。由此可知,当时间数列相对平稳时,可取较大的 α ;当时间数列波动较大时,应取较小的 α 。生产预测中,平滑常数的值取决于产品本身和管理者对良好响应率内涵的理解。

当时间序列没有明显的趋势变动时,使用第 t 周期一次指数平滑就能直接预测第 $t+1$ 期之值。但当时间序列的变动呈直线趋势时,用一次指数平滑法来预测仍存在着明显的滞后偏差,一次指数平滑不能用于有明显趋势变动的现象的预测,只能向未来预测一期,因此,其也需要进行修正。修正的方法也是在一次指数平滑的基础上再作二次指数平滑,利用滞后偏差的规律找出曲线的发展方向和发展趋势,然后建立直线趋势预测模型。故其称为二次指数平滑法。

尽管双指数平滑法可以用于具有线性趋势的数据集,但它要求数据采集间隔是规则的。而 Wright's 改进的 EDS 可以用于不规则采样时间序列。根据该算法,下一个采样间隔 k 的计算方法如下:首先计算估计值 y'_{t+k} :

$$y'_{t+k} = L_t + KM_t \quad (1)$$

其中, t 代表时间序列, $t=1, 2, \dots, n$; L_t 和 M_t 分别代表 t 时刻

的估计和趋势; k 代表时刻 t 到 $t+1$ 的时间间隔(步长), k 可以动态变化,这使得 Wright's 改进的方法可以用于不规则采样的时间序列。 L_t 和 M_t 遵循下面的基本规则:

$$\text{new estimate} = (\text{parameter}) \text{new information} + (1 - \text{parameter}) \text{previous estimate}$$

具体由下式给出:

$$L_{t+k} \leftarrow (1-V_t)(L_t + KM_t) + V_t y_{t+k} \quad (2)$$

$$M_{t+k} \leftarrow (1-U_t)M_t + U_{t+k}(L_{t+k} - L_t)/k \quad (3)$$

L_t 的前一个估计值是 $L_t + KM_t$,新的信息是实际的读数 y_{t+k} ; M_t 的前一个估计值是 $(L_{t+k} - L_t)$,所以,根据式(2)和式(3),参数 V_t 和 U_t 如下:

$$V_{t+k} = V_t / (b_{t+k} + V_t) \quad (4)$$

其中 $b_{t+k} = \text{power}((1-\alpha), k)$ 。

$$U_{t+k} = U_t / (d_{t+k} + U_t) \quad (5)$$

其中 $d_{t+k} = \text{power}((1-\beta), k)$ 。

V_t 和 U_t 是归一化因子, α 和 β 是平滑参数, $0 < \alpha < 1, 0 < \beta < 1$ 。根据时间序列的初始水平和趋势值设置 L_t 和 M_t 的初值, V_t 和 U_t 初值的设定依据如下公式:

$$U_t = 1 - (1-\beta)^q, V_t = 1 - (1-\alpha)^q \quad (6)$$

式中, q 代表平均时间间隔。

3 算法描述

调整采样率的算法分成两个阶段执行,如图 1 所示。第一个阶段用于计算步长,它是基于预测误差(预测值和真实值之间的差)的。步长 k 是两次采样点之间的时间间隔,用于时间序列的预测,产生下一个采样点的预测值;第二阶段用于步长调整,当检测到异常时,可能有重要的事件发生,或是遇到了错误,需要及时减小采样率。下面详细描述。

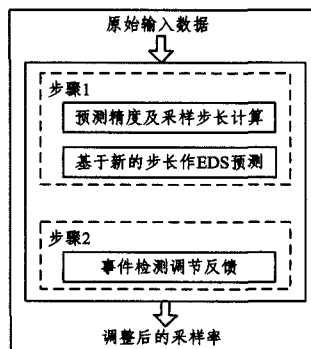


图 1 算法流程

在第一阶段中,每个采样点通过比较实际采样值和预测值,得到一个预测误差值 δ ;基于 δ 计算下一个采样间隔 k ,得到下一个采样点。当误差 δ 在允许的范围内时,可以加大 k 的值,但不能超过 k 的最大值(记为 S_{\max} ,由实际应用程序的需求决定)。文献[1]采用简单的加 1 来改变 k 的值, k 的值在任何情况下都线性加 1;事实上,在数据平稳的情况下,可以快速加大采样间隔。本文参考 TCP 拥塞控制的思想: k 的初始值设为 1,在 $k < S_{\max}/2$ 的范围内,采用指数增长的方式, $k=2 * k, k$ 成倍增长;当 k 到达 $S_{\max}/2$ 时,减缓 k 的增长速度,改用 $k=k+1$ 的线性增长方式。当 δ 超出范围时, k 不直接减为 1,而是先减半,即 $k=k/2$ 。

第二阶段主要的任务是探测事件。当 k 到达最大值,并以最大值持续采样时,需要监测采样过程中是否有异常发生。

表 1 算法参数列表

参数	值
α	0.9
β	0.6
α_{long}	0.01
α_{short}	0.9
δ	0.07
S_{max}	8
q	1

异常意味着发生了事件,或者是传感器错误,如果发现异常,则需要降低 k 的值。异常检测的方法参照文献[5]的指数加权平均方法。EWMA 方法具有两个平滑参数: α_{long} , α_{short} , 其中 α_{long} 用于计算长期平滑值 L_{long} , α_{short} 用于计算短期平滑值 L_{short} , 而 $\theta = L_{short} / L_{long}$, 当 θ 超过了一定阈值时,就代表有事件发生,此时相应地降低 k 的值。算法的伪码如下:

Algorithm adaptive temporal sampling

```

Initialize variables  $k \leftarrow 1, \delta, S_{max}$ 
Initialize EDS smoothing parameters  $\alpha, \beta$ 
Initialize EWMA smoothing parameters  $\alpha_{long}, \alpha_{short}$ 
Initialize Wright's extension variables
Initialize forecast
function senseData(t+k)
while (dataCollection == TRUE)
 $y_{t+k} \leftarrow$  collect data sample
adaptiveSampling( $y_{t+k}, k, \delta, S_{max}$ )
end while
end function
function adaptiveSampling( $y_{t+k}, k, \delta, S_{max}$ )
计算预测值  $y'_{t+k}$ 
% 检查误差值是否超过预设值的阈值
if ( $y_t - y'_t < \delta$ ) then
if ( $k < Sec$ ) %采用 TCP 控制的思想调整 k
 $k = 2 * k$ ;
else if ( $k < S_{max}$ )
 $k = k + 1$ ;
Endif
Else %超过阈值时,令 k 减半
 $Sec = fix(k/2 + 0.5)$ ;
if  $k \geq 2$ 
 $k = k - 1$ ;
end
endif
%以下为事件检测部分
if ( $k = S_{max}$ ) then
if (consecutive  $S_{max}$  predictions) then
update  $L_{long}, L_{short}$ 
% 探测事件
Evaluate  $\theta = L_{short} / L_{long}$ 
%若发现事件,则调整 k 变小
if ( $\theta > 1$ ) then
 $k \leftarrow 1$ 
end if
else
reset  $L_{long}, L_{short}$ 
end if
end if
senseData(t+k)
end function
    
```

4 算法分析及仿真结果

本节通过实验来详细分析算法的特性。该算法在 Matlab 中实现,数据集由一个 24 小时的室内自来水水温数据组成。由于水温变化不大,数据较为平稳,算法中用到的各参数如表 1 所列。

采样率调整的目标在于避免不必要的采样,从而节省能源。但由于拉长了采样间隔,更容易漏掉一些重要的变化和信 息,因此衡量算法性能最基本的指标为采样率和事件丢失率,此处主要从这两个指标做算法的具体比较和分析。

(1) 采样率: 定义为用算法取得的数据样本点个数与实际数据集中样本点个数之比。这代表了采样算法节省的能耗。

(2) 事件丢失率: 在给定的相对阈值下漏检测的事件比例,即数据库中实际事件数与算法检测到的事件数之差的比。如果在某个时间点,事件发生了,但该点没有被采样到,则认为这是一个事件丢失。用 n_f 代表事件丢失数目, n 表示整个采样点数目,则事件丢失率计算如下:

$$MR = \frac{n_f}{n} \tag{7}$$

采样性能分析的计算中,参数的设置是非常重要的。容错率的范围为 0.01 至 0.11,最大步长的范围设为 1s 至 20s。容错率的选择依据应用场景中的实际情况而定,例如,室内水温的变化多大时是异常,此处设置为 0.01 至 0.11。

取数据集集中的 9000 个数据点运行程序,相对阈值 $\delta = 0.07, S_{max} = 8$ 时,对不同数据段进行实验,所得部分结果如表 2 所列。 n_2 代表实际的事件个数, n_1 代表判断出的事件数, t 为采样数。

表 2 基本数据结果

数据起点	方法	n_1	n_2	t	SF	MR
3000	原有方法	56	101	1359	0.1509	0.0331
	改进方法	56	101	1196	0.1328	0.0376
8000	原有方法	57	105	1491	0.1557	0.0322
	改进方法	60	105	1196	0.1328	0.0376
12000	原有方法	36	80	1514	0.1682	0.0291
	改进方法	33	80	1178	0.1309	0.0399

为减少随机性,分别取不同的数据段做实验,在获取事件数方面,EDSAS 与基于 TCP 的方法相差不大;而在采样数方面,基于 TCP 的方法的采样数均明显低于 EDSAS 方法。因此,基于 TCP 的方法在保持识别事件的基础上,进一步降低了采样率。

接下来衡量两种自适应采样方法的参数问题,即最大步长和容错率对于采样率和错误丢失率的影响程度。首先,在相对阈值 $\delta = 0.07$ 时,对不同的 S_{max} 做实验,结果如图 2 所示。由图 2 可见,当最大步长增加时,采样率下降;在不同的 S_{max} 下,基于 TCP 的方法的采样率均低于 EDSAS。

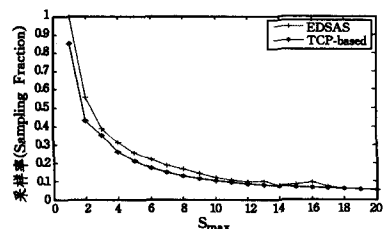


图 2 不同最大步长下的采样率比较

I; Information Flow)[J]. Foundations of Security Analysis and Design-Tutorial Lectures, Springer-Verlag, volume 2171 of LNCS, 2001; 331-396

- [7] Focardi R, Rossi S. Information Flow Security in Dynamic Contexts [J]. Journal of Computer Security, 2006, 14 (1): 65-110
- [8] Oheimb D. Information Flow Control Revisited; Noninfluence = Noninterference + Nonleakage [C] // Proceedings of European Symposium on Research in Computer Security 2004 (ESORICS'04). Springer-Verlag, Vol. LNCS 3193, 2004; 225-243
- [9] Riely J, Matthew H. Distributed Processes and Location Failures [J]. Theoretical Computer Science Archive, 2001, 266 (122): 693-735
- [10] 王立斌, 陈克非. 可移动系统安全模型统一框架[J]. 电子学报, 2002, 30(12A): 2108-2110
Wang Li-bin, Chen Ke-fei. A Uniform Framework of Security Model for Mobile Systems [J]. Acta Electronica Sinica, 2002, 30

(12A); 2108-2110

- [11] 余万涛, 胡光锐. 考虑位置失效的移动系统安全模型[J]. 计算机应用研究, 2006, 10; 128-129
Yu Wan-tao, Hu Guang-rui. Security Model for Mobile Systems with Location Failures [J]. Application Research of Computers, 2006, 10; 128-129
- [12] Castagna G, Vitek J, Zappa Nardelli F. The Seal Calculus [J]. Information and Computation, 2005, 201(1): 1-54
- [13] Davide Sangiorgi, David Walker. The Pi-Calculus; A Theory of Mobile Processes [M]. Cambridge University Press, 2002
- [14] Cardelli L, Gordon A D. Mobile Ambients [J]. Foundations of Software Science and Computation Structures, LNCS, 1998, 1378; 140-155
- [15] Bugliesi M, Gallina L, Hamadou S, et al. Behavioral Equivalences and Interference Metrics for Mobile Ad-Hoc Networks [J]. Performance Evaluation, 2014, 73; 41-72

(上接第 164 页)

由于两种方法实际采样的次数不一样, 因此评价 Miss Ratio 不具有可比性, 因而本文直接比较两种方法实际获取的事件数, 结果如图 3 所示。

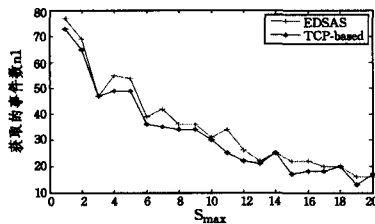


图 3 不同最大步长下获取的事件数对比

由图可见, 基于 TCP 的方法获取的事件数与 EDSAS 方法基本一致, 没有丢失重要的事件。接下来, 衡量不同阈值对采样率和事件丢失率的影响。取 $S_{max} = 8$ 的情况, 基于不同的相对阈值得到的结果如图 4、图 5 所示。

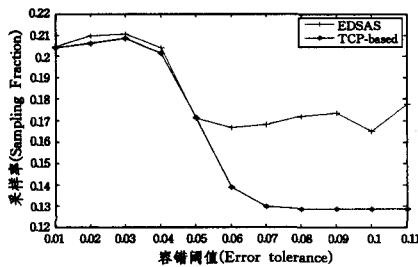


图 4 不同最大容错率下的采样率比较

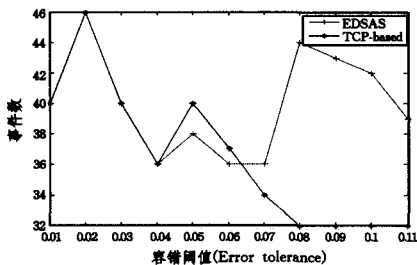


图 5 不同容错阈值下获取的事件数对比

当容错率增加时, 采样率快速下降, 在容错率从 0.05 增至 0.07 时, EDSAS 方法的采样率从 18% 降至 17%, 而基于 TCP 的采样方法的采样率则从 18% 降至 13%, 性能优于 ED-

SAS; 在获取事件数方面, 两种方法的性能仍基本一致。由此可见, 容错率对于算法的影响比较大。针对不同的具体应用场景, 需要通过实验来设置合理的参数。对于数据精度与能量损耗两个因素, 从图中可以发现适合的参数取值。

结束语 算法的参数设置依赖于具体的应用需求, 该算法针对场景选择合适的参数, 能够很好地动态调整采样间隔并尽可能地获取重要信息。当预测误差低于阈值时, 步长 k 增大, 这样可以使传感器节点避免对不需要的数据点进行采样。另一方面, 如果预测误差超过了阈值, 或者是检测到了事件, 算法将降低 k 的值, 在较小的采样间隔上采样, 以避免丢失重要的信息。改进后的算法参考 TCP 拥塞控制的思想, 使得 k 的调整能够快速响应变化。实验结果表明, 该算法进一步提高了采样率调整的有效性。该算法计算简单, 在资源受限的节点上易于实现。

参 考 文 献

- [1] Gupta M, Shum L V, Bodanese E, et al. Design and evaluation of an adaptive sampling strategy for a wireless air pollution sensor network [C] // 2011 IEEE 36th Conference on Local Computer Networks (LCN). IEEE, 2011; 1003-1010
- [2] Werner-Allen G, et al. Monitoring volcanic eruptions with a wireless sensor network [C] // Proceedings of the Second European Workshop on Wireless Sensor Networks, 2005. 2005; 108-120
- [3] Alippi C, et al. Energy management in wireless sensor networks with energy-hungry sensors [J]. Instrumentation & Measurement Magazine, IEEE, 2009, 12(2); 16-23
- [4] Alippi C, et al. Adaptive Sampling for Energy Conservation in Wireless Sensor Networks for Snow Monitoring Applications [C] // IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS 2007). 2007; 1-6
- [5] Alippi C, Roveri M. An adaptive CUSUM-based test for signal change detection [C] // Proceedings of 2006 IEEE International Symposium on Circuits and Systems (ISCAS 2006). 2006; 5752-5755
- [6] Wright D J. Forecasting Data Published at Irregular Time Intervals Using an Extension of Holt's Method [J]. Management Science, 1986, 32(4): 499-510