

预测修正下的动态数据传送机制

杨军^{1,2}, 张德运¹

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 宁夏大学数学与计算机科学学院, 750021, 银川)

摘要: 利用无线传感器网络中节点感知数据的相关性, 提出了一种基于预测修正的动态数据传送机制. 核心思想是: 将数据预测与模型计算分离, 接收节点先对数据进行平稳化处理, 然后建立模型或动态更新模型, 再把模型参数发送给采样节点. 采样节点用精简的预测修正算法预测数据, 通过比较法确定需要发送的采样数据, 从而减少了数据传送的次数, 延长了网络的生命期. 仿真结果表明, 所提算法可以滤除真实采样观测序列中 83% 的冗余数据, 预测精度较一般预测算法提高了 22%, 它适用于能量约束性较强的无线传感器网络.

关键词: 无线传感器网络; 时间序列; 平稳化; 预测修正

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2008)06-0655-04

Dynamic Data Transmission Mechanism Based on Prediction Revision

YANG Jun^{1,2}, ZHANG Deyun¹

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. School of Mathematics and Computer Science, Ningxia University, Yinchuan 750021, China)

Abstract: A dynamic data transmission mechanism based on prediction revision is presented according to the inherent correlation of sampling data between nodes in WSNs. The core of the dynamic data transmission mechanism is to separate the data prediction and the model computing. The sampling data in sink node are firstly stabilized, the model is built or updated dynamically, and then the parameters are sent to sensor nodes by sink. The predictions in sensor nodes are made using a simplified prediction revision algorithm, and then, the sampling data that needs to be sent out is determined based on a comparison between the prediction data and the sampling data. The times of data transmission can be reduced and the lifetime of WSNs can be prolonged. Simulation studies show that the proposed algorithm can decrease times on sending data by more than 83% and the prediction precision can be increased by about 22% compared with the traditional prediction algorithms in real sampling data series. The algorithm is suitable for WSNs which are strictly constrained by energy.

Keywords: wireless sensor network; time-series; stabilizing; prediction revision

在无线传感器网络中, 节点感知到的被测对象属性数据往往在时间上是相关的, 是一个时间序列数据, 直接发送这类数据会浪费有限的传输带宽, 造成能量消耗. 为了解决这类问题, 学术界提出了相关算法, 如以数据为中心的调度算法(DCS)^[1], 利用统计概率模型实现交互式查询法^[2], 通过节点对时序

数据建模^[3]等, 但它们忽略了以下问题: ①采样的时序数据的平稳性; ②模型的动态可调整性; ③预测方法的简化性. 为此, 本文提出了一种面向无线传感器网络的基于预测修正的动态数据传送机制, 以减少网络的数据传送次数, 延长网络的生命期.

1 问题描述

在无线传感器网络中,采样节点(Sensor)需感知被测对象的属性数据,并将感知数据传送给接收节点(Sink),而 Sink 会利用算法聚合接收到的数据,并将控制消息传送给 Sensor. 因此,在保证 Sink 获得一定精度采样数据的前提下,应尽可能减少 Sink 与 Sensor 之间的数据交换量或交换次数,才是减少无线传输能耗的关键. 假定时间是离散的,在最近的一段滑动时间窗口 $T=\{1,2,\dots,n\}$ 之内,若每一个整数代表一次采样的序号,采样数据的值域为实数域 $\mathbf{R}$,则 Sensor 相应的采样序列为 $X=\{X_1,\dots,X_n|X_i\in\mathbf{R}\}$.

1.1 采样数据序列中确定性信息的提取

采样数据大多属非平稳序列,为了建立时间序列模型,首先要处理序列中的确定性信息. 根据 Cramer 分解定理^[4],任何一个方差齐性的非平稳序列 $\{X_t\}$ 都可以分解为由两部分叠加的序列,其一是由多项式决定的具有确定性趋势的成分,另一是平稳的 0 均值误差成分,即

$$X_t = \sum_{j=1}^d \beta_j t^j + \phi(B)a_t \quad (1)$$

而 $\{a_t\}$ 为 0 均值的白噪声序列. 文献[5]证明了差分法是一种非常简便、有效的确定性信息提取方法,而离散序列的 d 阶差分就相当于连续变量的 d 阶求导,即

$$\nabla^d \sum_{j=1}^d \beta_j t^j = C \quad (2)$$

式中: C 为常数. 由式(2)可知,通过 d 阶差分可以将 $\{X_t\}$ 中蕴含的确定性信息充分地提取出来. 时间序列蕴含着两类趋势,一类是长期性趋势,另一类是季节性趋势. 对于长期性趋势,可利用多阶差分

$$Y_t = (1-B)^d X_t \quad (3)$$

来提取确定性信息,而对于季节性趋势,可利用

$$Y_t = (1-B_d)X_t \quad (4)$$

进行多步差分来提取确定性信息,其中的 B 为延迟算子.

1.2 采样数据的自回归模型

1.2.1 自回归模型 采样数据的自回归(AR)模型可根据一定的时间序列模型,在满足误差要求的情况下(本算法采用最小二乘估计)来求解模型参数. 由于传感器节点可以使用的能量是有限的,所以节点应采用简单的中心化 AR 模型^[6],即

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (5)$$

式中: X_i 是第 i 次采样值; ϕ_i 是回归系数; a_t 是服从 $N(0, \sigma_a^2)$ 标准正态分布的白噪声. 式(5)表明,第 t 次采样值可以用其前 p 次的线性组合加上一个随机的白噪声来表示. 令

$$\begin{aligned} Y &= [X_{p+1}, X_{p+2}, \dots, X_N]^T \\ \phi &= [\phi_1, \phi_2, \dots, \phi_p]^T \\ \varepsilon &= [a_{p+1}, a_{p+2}, \dots, a_N]^T \end{aligned} \quad (6)$$

$$X = \begin{bmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & & \vdots \\ X_{N-1} & X_{N-2} & \dots & X_{N-p} \end{bmatrix} \quad (7)$$

由式(6)、式(7)可得

$$Y = X\phi + \varepsilon \quad (8)$$

由最小二乘法可得系数矩阵和方差估计,即

$$\hat{\phi} = (X^T X)^{-1} X^T Y \quad (9)$$

$$\begin{aligned} \hat{\sigma}_a^2 &= \frac{1}{(N-p)} \sum_{t=p+1}^N (X_t - \hat{\phi}_1 X_{t-1} - \\ &\quad \hat{\phi}_2 X_{t-2} - \dots - \hat{\phi}_p X_{t-p})^2 \end{aligned} \quad (10)$$

1.2.2 预测修正算法

(1) AR 模型数据预测的一般方法. 就 AR 模型来说, t 时刻的 l 步前向预测及方差分别为

$$\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \dots + \phi_p \hat{X}_t(l-p) \quad (11)$$

$$\text{var}[e_t(l)] = (1 + G_1^2 + \dots + G_{l-1}^2) \sigma_a^2 \quad (12)$$

式中

$$\begin{aligned} \hat{X}_t(k) &= \begin{cases} \hat{X}_t(k) & k \geq 1 \\ X_{t+k} & k \leq 0 \end{cases} \\ G &= \begin{cases} G_0 = 1 \\ G_j = \sum_{k=1}^j \phi'_k G_{j-k} & j = 1, 2, \dots \end{cases} \end{aligned}$$

其中

$$\phi'_k = \begin{cases} \phi_k & k \leq p \\ 0 & \text{其他} \end{cases}$$

由此可知,对于平稳时间序列的预测,实质上就是根据所有已知的历史信息 $X_t, X_{t-1}, \dots$ 对序列未来的某个时期的发展水平 $X_{t+l} (l=1, 2, \dots)$ 做出估计.

(2) 预测修正算法(PRA). 假定 t 时刻的 l 步前向预测为 $\hat{X}_t(l)$, 而在 $t+1$ 时刻,若 X_{t+1} 已知,则 $\hat{X}_{t+2}, \hat{X}_{t+3}, \dots$ 的预测精度可以得到修正,其表达式为^[6]

$$\begin{aligned} \hat{X}_{t+1}(l) &= \hat{X}_t(l+1) + G_l a_{t+1} \\ a_{t+1} &= X_{t+1} - \hat{X}_t(1) \\ \text{var}[e_{t+1}(l)] &= (1 + G_1^2 + \dots + G_{l-1}^2) \sigma_a^2 = \text{var}[e_t(l)] \end{aligned} \quad (13)$$

其中 G_l 可以由式(12)中的格林公式计算,式(13)表明, $t+1$ 时刻的预测值可以由 t 时刻第 $l+1$ 步的预测值和 G_l 倍的 a_{t+1} 获得,而且一期修正后的第 $l+1$ 步预测方差就等于修正前的第 l 步预测方差,它比修正前的同期预测方差减少了 $G_l^2\sigma_a^2$ 。上述算法表明,若已获得了新的观察值,要获得 X_{t+l} 更精确的预测值并不需要对历史记录进行反复计算,只要利用新观察值去修正预测值即可。

2 算法的动态调整

动态调整是通过 Sink 和 Sensor 的互动来完成的,即将模型计算与数据预测分离,由 Sink 完成模型的计算和更新(见图 1),并根据设定的比率将模型参数发给 Sensor,而 Sensor 通过数据采样和误差比较(见图 2),将采样的真实值发送给 Sink,这样可把计算量较大计算模块放在能量相对充足的 Sink 端, Sensor 只进行简单的预测和比较,从而极大地减少了 Sensor 的计算能耗,延长了 Sensor 的生存能力。

```
While(1){
    Wait for data from Sensor;
    If (no data receive){
        predict  $l$  step data using predict model;
        Continue;
    }
    If ( the ratio of number of data coming more than threshold  $\alpha$  ){
        Update model and send parameter to sensor;
        Continue;
    } Else
        Save receiving data;
}
```

图1 Sink 计算或更新算法

```
While(1){
    Wait for model's parameter from Sink;
    If (received model's parameter){
        Update model;
    }
    Do{
        Sampling and predict data;
    } While(prediction error less than or equal to threshold);
    Send data to Sink;
}
```

图2 Sensor 采样和误差比较算法

3 实验结果与分析

实验数据取自 2001-01-01 至 2001-12-31 赤道附近海洋表面的空气温度^[7],随机选取其中的 4 000 个连续观测值,采样间隔为 10 min,采样数据具有周期性趋势和线性趋势。实验中将本文算法与一般预测算法(GPA)进行了比较,其中预测失败的温度误差门限设定为 $0.1\text{ }^{\circ}\text{C}$,模型自动调整的参数为 α 。

3.1 评价指标及采样数据的平稳化处理

评价指标包括:数据发送率(RDS)

$$R_{DS} = \frac{\sum_{i=1}^n \delta(|X_i - \hat{X}_i| > 0.1)}{n}$$

是指因预测失败而发送的数据占全部采样数据的比例,其中 $\delta(x)$ 为 Delta 函数;平均相对误差(MRD)

$$E_{MRD} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|X_i - \hat{X}_i|}{X_i} \right)$$

平均绝对误差(MAD)

$$E_{MAD} = \frac{1}{n} \sum_{i=1}^n |X_i - \hat{X}_i|$$

先对原始的采样数据进行 144 步差分处理,以消除周期性影响,再通过 1 阶差分消除线性趋势。

3.2 数据分析

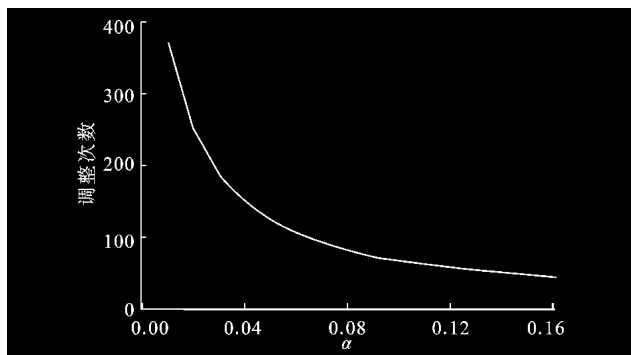
(1)温度误差门限对算法的影响。GPA 与 PRA 的精度比较如表 1 所示。从中看出,当温度误差门限为 $0.05\text{ }^{\circ}\text{C}$ 时, GPA 仅能过滤大约 43% 的冗余数据,而 PRA 则能滤除 66% 的冗余数据, PRA 较之 GPA 的预测精度提高了 23%。当温度误差门限为 $0.1\text{ }^{\circ}\text{C}$ 时, PRA 能够滤除 83% 的冗余数据,较之 GPA 的预测精度提高了 22.8%。当温度误差门限为 $0.15\text{ }^{\circ}\text{C}$ 时, PRA 能过滤大约 89% 的冗余数据,而 GPA 能过滤大约 80% 的冗余数据,但二者对应的 MRD、MAD 也比较大。由此可知,只有温度误差门限为 $0.1\text{ }^{\circ}\text{C}$ 时, PRA 在效率和精度上的优势才比较明显。

(2)参数 α 对算法的影响。 α 为预测失败的次数与一轮样本数之比(一轮样本数为 100,总样本数为

表1 PRA、GPA 算法的精度比较

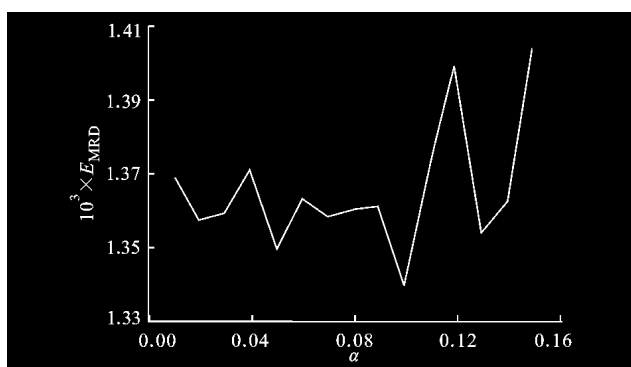
误差门限/ $^{\circ}\text{C}$	算法	E_{MRD}	$E_{MAD}/^{\circ}\text{C}$	R_{DS}
0.05	GRA	0.000 728	0.018 842	0.569 085
	PRA	0.000 560	0.014 703	0.336 500
0.10	GRA	0.001 701	0.044 452	0.282 000
	PRA	0.001 322	0.034 707	0.166 750
0.15	GRA	0.002 656	0.069 462	0.197 054
	PRA	0.002 065	0.054 228	0.110 500

4 000). 算法通过调整参数 α 来权衡能量消耗和预测精度. α 与模型调整次数的关系如图 3 所示, 从中看出, 模型调整次数将随着 α 的增大而减小. MRD 随 α 的变化如图 4 所示, 从中看出, α 在 0.01~0.10 之间, MRD 随 α 的增大而减小, 而当 $\alpha > 0.1$ 时, MRD 随 α 的增大而增大. 由此可见, 当 α 在 0.1 左右时, 算法在调整次数和预测精度之间取得平衡, 并获得比较好的性能.



误差门限为 0.1 °C; 一轮样本数为 100; 总样本数为 4 000

图 3 参数 α 对算法调整的影响



误差门限为 0.1 °C; 一轮样本数为 100; 总样本数为 4 000

图 4 参数 α 对 MRD 的影响

4 结 论

在无线传感器网络中, 节点感知的采样数据往往具有时间相关性, 这种相关性表现为数据的可预

测性, 即可以根据数据的历史观测值来预测未来的观测值. 在这一条件下, 对于一个稳定的温度过程或者一个按照某种轨迹运动的目标, 预测其未来值是完全可行的. 本文利用观测值在时间上的连续性和相关性对连续的数据建立关联模型, 采用预测修正算法进一步滤除传送的冗余数据, 从而降低了传输能耗, 实验结果也验证了算法的可行性.

参考文献:

- [1] MA Chi, MA Ming, YANG Yuanyuan. Data-centric energy efficient scheduling for densely deployed sensor networks [C] // Proceedings of IEEE International Conference on Communications. Piscataway, NJ, USA: IEEE, 2004:3652-3656.
- [2] DESHPANDE A, GUESTRIN C, MADDEN S R, et al. Model-driven-data acquisition in sensor networks [J]. The VLDB Journal, 2004, 14(4):417-443.
- [3] LAZARIDIS I, MEHROTRA S. Capturing sensor-generated time series with quality guarantees [C] // Proceedings of IEEE International Conference on Data Engineering. Piscataway, NJ, USA: IEEE, 2003:429-440.
- [4] 王燕. 应用时间序列分析 [M]. 北京: 中国人民大学出版社, 2005:90-97.
- [5] BOX G, JENKINS G M, REINSEL G. Time series analysis forecasting and control [M]. 3rd ed. San Francisco, CA, USA: Prentice Hall, 1994:89-120.
- [6] PANDIT S M, WU Shienming. Time series and system analysis with applications [M]. New York, USA: John Wiley and Sons, 1983:118-199.
- [7] MCPHADEN M J. Tropical atmosphere ocean project [EB/OL]. [2007-02-06]. <http://www.pmel.noaa.gov/tao/index.shtml>.

(编辑 苗凌)