

基于 ARMA 的无线传感器网络流量预测模型的研究

李捷 刘先省 韩志杰

(河南大学计算机与信息工程学院 开封 475001)

摘要: 无线传感器网络(WSN)的流量预测研究对于 WSN 管理具有至关重要的意义。基于 WSN 的特点, 结合自回归滑动平均(ARMA)模型, 该文提出了一个面向 WSN 的流量预测算法。仿真结果表明, 该算法具有较高的预测精度; 利用本算法对 WSN 的多路径路由机制进行改进后, 可进一步提高 WSN 的生存期。

关键词: 无线传感器网络; 流量; 自回归滑动平均模型; 多路径路由

中图分类号: TP393.07

文献标识码: A

文章编号: 1009-5896(2007)05-1224-04

Research on the ARMA-Based Traffic Prediction Algorithm for Wireless Sensor Network

Li Jie Liu Xian-xing Han Zhi-Jie

(College of Computer & Information Engineering, Henan University, Kaifeng 475001, China)

Abstract: The importance of traffic prediction in Wireless Sensor Network (WSN) should not be disregarded. Concerning the ARMA model, a traffic prediction algorithm is proposed based on its own traffic characters. Simulation results show that the presented algorithm can provide high accuracy. Furthermore, using the proposed algorithm to improve the multi-paths route scheme can extend the lifetime of the whole WSN by remaining load evenly distributed among several paths.

Keywords: Wireless sensor network; Traffic; ARMA model; Multi-path route

1 引言

无线传感器网络(Wireless Sensor Network, WSN)具有很多独特的优点和应用领域, 引起了人们的广泛关注^[1]。传感器节点通常由电池供电, 高效使用有限的电池资源, 尽量延长节点的生命周期是任何WSN相关协议的首要考虑因素。同时由于传感器网络节点的计算能力、存储空间都十分有限, 所设计的路由协议应该简单有效, 使得节点的运算开销尽量的小。与传统网络管理相比, WSN的网络管理不仅包括故障管理, QoS支持等功能域, 还包括了能量管理^[2]。网络实时监测是网络管理基础的部分, 网络监测的目的是为了收集关于网络状态和行为的信息。网络流量的管理是网络监测的一个重要方面。它包括监测流量异常及其诊断、流量异常问题的解决两个阶段。已有的研究成果表明, 对网络流量预测展开研究对网络管理而言具有重要的理论和实践意义。因此面向WSN的网络流量预测的研究对于WSN的网络管理而言是不可或缺的。

在网络流量预测中, 网络流量模型起着非常重要的作用。准确的流量模型能捕获实际网络流量的统计特征。一个模型如果不能捕获实际流量的统计特征, 将引起差的网络性

能, 因为它们要么过高估计网络性能, 要么低估了网络性能。对于传统网络而言建立恰当的网络流量模型也不是一件很容易的事情, 建立什么样的模型, 这个问题是仍然在研究的课题。

与传统网络不同, WSN是以数据为中心的, 由负责sensor节点和sink节点组成, 其数据流量表现出明显的不均衡特征^[3], 具体为大量的数据从sensor节点传向sink节点, 少量的控制消息从sink节点传送至sensor节点, 其主要数据流量集中在sink节点和基站处。WSN网络是应用相关的, 以事件驱动和周期性数据查询为主, 其数据流量相应地表现出随机性和周期性, 对目标数据的采集和跟踪将会产生突然流量, 这种情况不能由CBR或者VBR模型描述^[4]。

流量模型必须有可管理的参数个数, 参数估计必须简单。有许多随机模型用于描述网络流量。通信网络中的流量模型可以分为稳定的和不稳定的两种, 稳定的流量模型又分为两类: 短相关和长相关。短相关模型包括马尔可夫过程和自回归模型(AR), 自回归滑动平均模型(ARMA), 以及自回归求和滑动平均模型(ARIMA); 长相关流量模型包括: F-ARIMA 和 Fractional Brownian Motion(FBM), 我们采用 ARMA 模型, 它不能建立长相关的模型, 因此只能进行短期预测。

实现预测技术的有 LMS 和 Karhman 过滤器, 神经网络、自回归模型、自回归滑动平均以及分形自回归求和滑动平均

模型(F-ARIMA)和小波模型(Wavelet)。LMS 过滤器不适宜描述不稳定的网络流量;Karlman 过滤器在预测实现之前要求过程的统计信息,这在实时性的预测过程中不能保证;神经网络技术适宜描述流量的不稳定性,但是它的计算量通常都很大;F-ARIMA 和小波模型则是用于捕获长相关特性。

现有的 WSN 流量预测算法中主要有 Poisson^[5], Markov^[6], AR^[7]。Poisson 已被证明不适于 WSN 的流量特点^[9], Markov 预测模型缺点是预测精度低,误差大,适用范围小。应用 ARMA 模型可以进行短期的预测,它比 AR 模型具有更小的预测误差方差。

因 WSN 中以平稳性,周期性数据流量为主,ARMA 模型对平稳性数据序列能够有效地分析其数据序列相关性,所以本文采用了 ARMA 预测模型对 WSN 的流量进行分析。理论分析和仿真结果都取得了较好的效果。根据 WSN 的流量特点,采用 ARMA(2, 1),设计了一种 WSN 的流量预测模型,使得计算复杂性降低,提高预测精度。将该模型应用于 WSN 的多路路由协议中,实现了基于流量预测的路径切换,通过网络负载的均衡,延长了 WSN 的生存期。

本文组织如下,第 2 节给出了传感器节点流量的 ARMA 预测模型,并结合实例对 ARMA 预测模型的预测精度进行了仿真验证;第 3 节中利用本算法对多路路由由算法进行改进并对改进前后的 WSN 生存期等数据进行了分析说明;最后在第 4 节总结全文。

2 传感器节点数据流量的 ARMA 预测模型

2.1 ARMA 预测模型

2.1.1 平稳性判定 假设数据流量序列为

$X'_0, X'_1, \dots, X'_i, \dots, X'_n$, 采用平稳化序列方法对 $X'_0, X'_1, \dots, X'_i, \dots, X'_n$ 进行取对数后得到 $X_0, X_1, \dots, X_i, \dots, X_n$ 。

首先判定 $X_0, X_1, \dots, X_i, \dots, X_n$ 是否为稳定时间序列。计算 $X_0, X_1, \dots, X_i, \dots, X_n$ 的自相关函数和偏相关函数,如果呈现出明显的拖尾现象,即可认定其为 ARMA 序列。

2.1.2 定阶 这里采用文献[8]中的 AIC 定阶法。

AIC 准则函数如下:

$$AIC = 2k - 2L(\hat{\beta}) \quad (1)$$

其中 k 为独立参数个数, $\hat{\beta}$ 为参数的最大似然估计值, $L(\bullet)$ 似然函数。ARMA 模型的似然函数近似为

$$\left. \begin{aligned} L(\hat{\beta}) &= -\frac{n}{2} \lg 2\pi - \frac{n}{2} \lg \hat{\sigma}^2 - \frac{S(\hat{\beta})}{2\hat{\sigma}^2} \\ \hat{\sigma}^2 &= \frac{1}{n} S(\hat{\beta}) \\ \hat{\beta} &= (\hat{\varphi}, \theta)^T = (\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_p, \theta_1, \theta_2, \theta_3 \dots \theta_q) \end{aligned} \right\} \quad (2)$$

结合式(1), 式(2)最终可得

$$AIC(p, q) = n \lg \hat{\sigma}^2 + (p + 1) \quad (3)$$

结合以上分析^[9], 这里采用 ARMA(2, 1)^[3, 8]来进行预测, 其模型为

$$\varphi(B)X_t = \theta(B)a_t \quad (4)$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 \quad (5)$$

$$\theta(B) = 1 - \theta_1 B \quad (6)$$

其中 B 是后移算子, a_t 是白噪声。 $\varphi_1, \varphi_2, \theta_1, \sigma_a^2$ (白噪声方差)为估计参数。ARMA 相关矩估计方法主要有最小二乘估计方法, 最大似然估计, 最大熵估计等等, 考虑到传感器节点的计算能力, 这里我们采用最小二乘估计方法可解出 $\hat{\varphi}_1, \hat{\varphi}_2, \hat{\theta}_1, \hat{\sigma}_a^2$ 。

判断时间序列的稳定性, 其稳定性条件为 $\hat{\varphi}_1 + \hat{\varphi}_2 < 1$, $\hat{\varphi}_2 - \hat{\varphi}_1 < 1, |\hat{\varphi}_2| < 1$, 如果满足此条件则视为平稳序列。则得出 ARMA 拟和模型为

$$X_t = \hat{\varphi}_1 X_{t-1} + \hat{\varphi}_2 X_{t-2} + a_t - \hat{\theta}_1 a_{t-1} \quad (7)$$

进一步地, 我们利用逆函数法进行一步预测, ARMA 的逆函数记为 I_1, I_2, \dots, I_n , 有

$$\left. \begin{aligned} I_1 &= \hat{\varphi}_1 - \hat{\theta}_1 \\ I_2 &= \hat{\varphi}_2 - I_1 \hat{\theta}_1 \\ I_3 &= I_j \hat{\theta}_1 \dots (j > 3) \end{aligned} \right\} \quad (8)$$

则一步预测模型为

$$\hat{X}_t(1) = \sum_{j=1}^m I_j X_{t+1-j} \quad (9)$$

其中 m 为 X_t 之前 m 次观测值, 可根据预测精度的要求取值。

其多步预测模型为

$$\hat{X}_t(l) = \hat{\varphi}_1 \hat{X}_t(l-1) + \hat{\varphi}_2 \hat{X}_t(l-2) \quad (10)$$

2.2 预测精度仿真

如图 1(a)所示, 为一个 WSN 中主路由上的某一传感器节点的真实网络流量, 利用 ARMA(2, 1)分析其时间序列, 对其利用上述模型进行计算, 可估计出模型参数如下:

$\hat{\varphi}_1 = 0.86579, \hat{\varphi}_2 = -0.07356, \hat{\theta}_1 = 0.68954, \hat{\sigma}_a^2 = 0.00186$ 。在这里 $\hat{\varphi}_1 + \hat{\varphi}_2 = 0.79223 < 1, \hat{\varphi}_2 - \hat{\varphi}_1 = -0.93935 < 1, |\hat{\varphi}_1| < 1$, 满足序列平稳性条件, 可以推出其拟和模型为

$$X_t = 0.86579 X_{t-1} + 0.07356 X_{t-2} + a_t - 0.68954 a_{t-1} \quad (11)$$

进一步地, 令 $m = 3$, 得到其一步预测模型:

$$X_t(1) = \sum_{j=1}^3 I_j X_{t+1-j} \quad (12)$$

其多步预测模型为

$$\hat{X}_t(l) = 0.86579 \hat{X}_t(l-1) - 0.07356 \hat{X}_t(l-2) \quad (13)$$

从任意时刻起在 250min 内, 每分钟采样其数据流量, 利用本模型可计算出其估计值, 其一步预测、二步预测和五步预测的比较图如图 1(c)–1(d)所示。由于 ARMA 为短相关模型, 因此在一步预测中取得了较好的预测效果, 多步预测的精度有所下降。

3 改进多路径路由协议

3.1 改进思路

在传感器网络中,引入多路径路由是为了提高数据传输

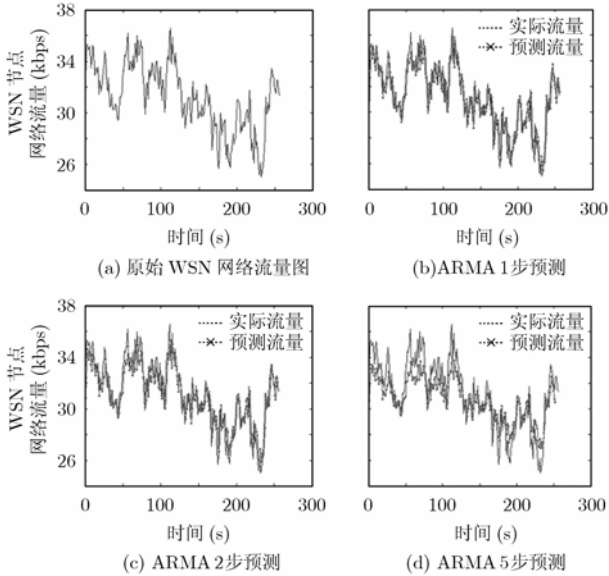


图 1 ARMA 预测结果

的可靠性和实现网络负载平衡。文献[6]提出一种多路径路由机制,它预先建立和维护一组数据,无需周期性洪泛就能够从数据源节点到汇聚节点的传输路径。其基本思想是:首先建立从数据源节点到汇聚节点的主路径,然后再建立多条备份路径;数据通过主路径传送数据来维护数据的有效性;当主路径失败时,从备用路径中选择次优路径。这种由主路由失败触发的切换机制并不适用于 WSN 的需求,因为负载的不均衡将会导致传感器节点资源消耗的不均衡,从而导致 WSN 全网的生存期降低。可以通过引入预测机制,主路由在预测其将要传输失败前切换至次优路径,从而实现了 WSN 多路径之间的负载均衡。其关键在于设计一种基于预测的阈值超越判断方法。

3.2 基于预测的阈值超越判断方法

设阈值为 Max , 设在 l 步预测值超过阈值概率为

$$P_i(l) = (y(t) > \text{Max} | X_t, X_{t-1}, X_{t-2}, \dots, X_{t-m}) \quad (14)$$

此处 m 为 X_t 之前 m 次观测值。

据式(11), X_t 的变化受 a_t 的变化制约,若 a_t 为正态分布,则 X_t 也为正态分布,对式(10)进行标准化,则得

$$Z_t(l) = \frac{X_t(l) - \hat{X}_t(l)}{\sigma(l)} \quad (15)$$

其中 $Z_t(l)$ 为标准化后随机变量, $\sigma(l)$ 为线性最小方差。则

$$\begin{aligned} P(X_{t+l} \leq \text{Max}) &= P\left(\frac{X_{t+l} - \hat{X}_t(l)}{\sigma(l)} \leq \frac{\text{Max} - \hat{X}_t(l)}{\sigma(l)}\right) \\ &= P\left(Z_t(l) \leq \frac{\text{Max} - \hat{X}_t(l)}{\sigma(l)}\right) \\ &= \Phi\left(\frac{\text{Max} - \hat{X}_t(l)}{\sigma(l)}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{\text{Max} - \hat{X}_t(l)}{2\sigma(l)}\right) dx \quad (16) \end{aligned}$$

其中 $\phi(x)$ 标准正态分布函数。进一步地,

$$\begin{aligned} P_i(l) &= 1 - P(X_{t+l} \leq \text{Max}) \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{\text{Max} - \hat{X}_t(l)}{2\sigma(l)}\right) dx \quad (17) \end{aligned}$$

3.3 仿真实证

利用 NS2 对改进前后的多路径路由协议进行了性能比较。实验参数如表 1 所示。

表 1 仿真实验参数表

节点数	100
区域面积	100m×100m
基站坐标	(0, 0)
信道带宽	1Mbps
节点初始电能	2J
采样周期	10s

感知节点每次采样生成的数据量为 4096byte, 分别用改进前后的多路径路由协议对数据进行传送。试验中对全网的能耗和生存节点个数在每个采样周期进行了统计。统计结果如图 2 所示。

图 2(a)为改进前后的存活节点的采样值的统计图。在前 400s 内,均未发生存活节点的减少,在 400s 之后,改进前的多路径协议导致了存活节点数的急剧下降,在 600s 内即减少为 0,而在引入预测机制后节点的消亡速度较小,特别在 400s-500s 之间,节点的存活数减少走势较为平缓,最终在 700s 处无存活节点。改进后的协议可有效地减少节点的减员。

图 2(b)为改进前后的全网能耗的统计图。全网的初始功率为 200J。改进后的协议无论是同采样时刻中的能耗,还是全网的生存期,都表现出较改进前的更好的性能。这是因为在改进前的多路径路由协议中,主路径失败后方切换至次优路径中,而在引入预测机制后,在本路径即将失败时,就实现了路径的切换,这样就使得全网的负载得以有效地平衡分配至各个路径中。这种机制类似于固定网中常见的重路由机制。二者的区分在于固定网是通过在路由的某一结点处调整网络流量的物理路径,而在 WSN 中是通过整条路径的切换实现的。

4 结束语

本文提出了一种基于 ARMA 的面向 WSN 的流量预测模型。将本模型应用于 WSN 的多路径路由协议, 仿真结果

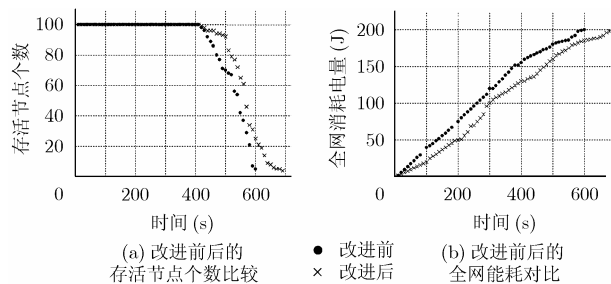


图2 仿真数据结果

表明改进后的多路路由协议, 由于引入了预测机制, 使得网络负载得以有效地均衡分配, 提高了全网的生存期。由于 WSN 的计算资源的受限, 对其预测模型的研究应注意以下几点: (1)算法的计算复杂度应尽可能的低; (2)预测尽量在本节点内进行, 否则将会由于参数等数据的交换引起新的网络负载。

要更好地实现 WSN 的流量预测功能, 还有许多工作要做。首先, 应该研究更多体现 WSN 性能的网络参数的预测; 其次, 研究具有一般性的时间序列的平稳化方法; 第三, 在实际应用中, 还需要模型参数估计的在线完成。

参考文献

[1] 孙利民, 李建中等. 无线传感器网络[M]. 北京: 清华大学出版社, 2005.5.
 [2] 崔莉, 鞠海玲等. 无线传感器网络研究进展[J]. 计算机研究与发展, 2005, 42(1): 163-174.
 [3] Deng Jing, Han Richard, and Shivakant Mishra. Defending against path-based DoS attacks in wireless sensor networks [C]. ACM press, New York, USA, 2005: 89-96.
 [4] Ilker Demirkol, Fatih Alag'oz, Hakan Delic, and Cem Ersoy. Wireless sensor networks for intrusion detection: packet traffic modeling[EB/OL]. www.cmpe.boun.edu.tr/~ilker/

IlkerDEMIRKOL_COMML_ext_abstract.pdf

- [5] Ma Y and Aylor J. System lifetime optimization for heterogeneous sensor networks with a hub-spoke topology, *IEEE Trans. on Mobile Computing*, 2004, 3(3): 286-294.
 [6] Ganesan D and Govindan R, *et al.* Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *Mobile Computing Review*, 2002, 1(2): 11-25.
 [7] Lisa A and Shay. The wireless network environment sensor: A technology independent sensor of faults in mobile wireless network links[D]. Rensselaer Polytechnic Institute Troy, New York, USA, 2002.
 [8] 张树京, 齐立心. 时间序列分析简明教程[M]. 北京: 清华大学出版社, 2003, 9: 98-122.
 [9] 薛可, 李增智, 刘浏等. 基于 ARIMA 模型的网络流量预测[J]. 微电子与计算机, 2004, 21(7): 84-87.

李 捷: 男, 1975 年生, 博士, 副教授, 研究方向为无线传感器网络、网络管理。

刘先省: 男, 1964 年生, 博士, 教授, 研究方向为传感器管理、网络建模。

韩志杰: 男, 1979 年生, 博士生, 研究方向为无线传感器网络。