

基于预测的节能数据分发机制

胡 宁, 孙钦东, 张德运

(西安交通大学电子与信息工程学院, 西安 710049)

摘要:为减少无线传感器网络的数据流量, 提出了一种基于预测的数据分发机制。在源节点和 sink 节点分别配置相同参数的预测器, 用以预测未来采样值。当在源节点处预测值和实际采样值的差异大于某门限值时, 当前采样数据才会被传送到 sink 节点。预测器采用双指数平滑预测模型, 根据历史数据去预测未来数据。实验表明, 该节能的数据分发机制可以滤除 83% 的冗余数据而仍然保持较高的精度。

关键词:无线传感器网络; 预测模型; 节能策略

Energy-saving Data Dissemination Based on Forecast

HU Ning, SUN Qindong, ZHANG Deyun

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

【Abstract】 In order to decrease the traffic in the wireless sensor networks, a data dissemination scheme based on forecast is presented. Two matching predictors with the same parameters are deployed at the source and the sink respectively, which predict the next immediate sampling value simultaneously. The sampling value is sent to the sink only when the difference between the actual and the predicted value at the source exceeds a certain threshold. The predictors adopt the double exponential smoothing forecast model to estimate the future values based on the past samples. The experiment results show that this energy-saving data dissemination scheme is very effective, which can filter about 83% data and remain high precision yet.

【Key words】 Wireless sensor networks; Forecast model; Energy-saving policy

随着微电子技术和精密制造业的发展, 具备感应、无线通信以及信息处理能力的微型无线传感器开始出现。这些廉价、低功耗的传感器节点被撒布在传感区域, 以自组织的方式构成无线传感器网络, 收集需要的数据并协作将这些数据传送给用户。无线传感器网络具有容错能力强、配置简单、感知信息准确等优点, 在军事侦察、设备监控、森林防火、环境监测等很多领域都具有广阔的应用前景^[2]。

与传统的无线网络不同, 无线传感器网络能源有限且难于补充, 为其设计的算法和协议应能够有效地利用节点有限的能量, 以延长网络的生命期。传感器节点的能量主要由无线数据通信消耗, 因此应减少数据的传送量和传送次数是一种最为有效的节能方式, 现有的很多协议和算法都是基于这一思想设计的^[5,6]。通过对检测对象的分析, 发现在通常情况下, 被监测对象的观测值是前后相关的。例如检测某物体或区域的温度时, 其温度的变化是连续的, 一般不可能出现温度跳变的情况; 而监测移动目标时, 目标的运动也遵循一定的轨迹, 不可能出现位置的跳跃。因此, 本文利用这种监测对象观测值在时间上的连续性和相关性, 对前后连续的数据建立关联模型, 采用预测的方法进一步滤除需要传送的数据。

1 基于预测的数据滤除体系结构

本文的设计思想是: 首先对特定应用建立预测模型, 构造采用此模型的预测器, 在传感器节点和 sink 节点采用同样的预测器, 并且配置参数相同; 然后在工作阶段利用此预测器去估计下一个观测值, 如果此观测值与预测值匹配, 则不需要进行当前观测值的传送, 仅当观测值与预测值不匹配时, 当前观测值才需要从传感器节点传送到 sink 节点。这样的数

据传送机制可以减少需要传送的数据量, 从而达到节省能量的目的。该机制的实现由传感器节点和 sink 节点联合完成。传感器节点中的预测器结构如图 1 所示, 由预测模块和决策模块组成。

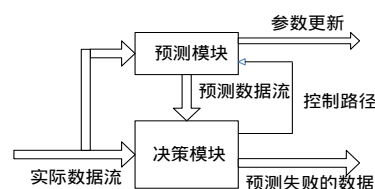


图 1 传感器节点的预测器

预测模块实现从应用数据流识别并建立预测模型, 它在线完成预测模型的参数估计, 然后利用该预测模型预测将要产生的数据值。决策模块负责对预测模块产生的数据进行评估, 即比较预测数据值和实际观测值。如果满足一定的精度, 表明预测匹配, 当前的观测值不需要传送; 而如果二者差异显著, 表明预测不匹配, 则发送实际观测值。当预测不匹配时, 决策模块通过图 1 中的控制路径向预测模块发出预测失败通告, 此时预测模块将重新进行预测模型参数的估计, 以适应当前数据流的特性。当预测模型的新参数估算完成时, 应向 Sink 节点发送预测器参数更新消息, 告知传感器节点当前预测器的参数设置, 以保持二者中预测器参数的一致。在

基金项目: 国家“863”计划基金资助项目(2003AA148010)

作者简介: 胡 宁(1978 -), 男, 博士生, 主研方向: 无线传感器网络; 孙钦东, 博士后; 张德运, 教授、博导

收稿日期: 2006-05-25

E-mail: huning@xanet.edu.cn

参数估算期间,实际采样的数据将全部被发送到 Sink 节点。

Sink 节点中的预测器结构如图 2 所示,仅包含预测模块。该模块采用与传感器节点中预测模块相同的参数,这些参数来自传感器节点的通告。当没有收到参数更新的时候, Sink 节点的预测器认为在传感器节点处预测模块产生的数据能够很好地匹配实际观测值,因此采用预测器所预测的数据作为采样数据观测值;而当收到预测更新消息的时候,它则知道在传感器节点处预测器预测失效,并且正在进行参数估算,因此采用收到的预测失效时才发送过来的数据作为观测值(这时的数据就是传感器节点真正采样得到的数据)。



图 2 Sink 节点的预测器

预测模型的识别和建立可以在传感器网络部署前事先完成,网络部署后,预测模块仅需要进行在线的模型参数估计。然而,传感器节点的计算资源有限,复杂预测模型的采用是不现实的,应该选择简单实用的方法。下文给出一个基于双指数平滑的预测方法,并取得了较好的效果。

2 双指数平滑预测模型

假定观测序列在一段时间内具有线性趋势,以 t 为起点的第 $t+k$ 期的预测值可以表示为

$$y_{t+k} = a_t + b_t k \quad (1)$$

这里的参数 a_t 和 b_t 随 t 的变化而变动,以使预测值有较好的适应性。以下讨论如何确定这两个参数的问题。

假设到时刻 t 为止有无穷多个观测数据 $\dots, y_{t-2}, \dots, y_{t-1}, y_t$ 。根据“厚近而薄远”的原则,构造指数形式的权系数,则以往观测数据的加权平均值为

$$S_t = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j y_{t-j} \quad (2)$$

其中, $0 < \alpha < 1$ 。A 称为记忆因子, α 越大,则越强调新数据的作用。经整理

$$\begin{aligned} S_t &= \sum_{j=0}^{\infty} \alpha(1-\alpha)^j y_{t-j} \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \dots \\ &= \alpha y_t + (1-\alpha)[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \\ &\quad \alpha(1-\alpha)^2 y_{t-3} + \dots] \\ &= \alpha y_t + (1-\alpha)S_{t-1} \end{aligned} \quad (3)$$

得到递推形式的指数加权平均值,并且可以证明 S_t 关于初值 S_0 是渐近稳定的,即 S_t 的值渐近地与初值 S_0 的选取无关。

式(3)为一次指数加权平均公式,为区别于后面的二次指数加权平均公式,改记为

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} \quad (4)$$

对 $S_t^{(1)}$ 再作一次指数平均,得到二次指数平均值

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \quad (5)$$

在原序列为线性趋势假定下,对于相同的时间坐标 t ,一次指数平均值滞后于序列观测值,而二次指数平均值又滞后于一次指数平均值,且滞差相等,即

$$y_t - S_t^{(1)} = S_t^{(1)} - S_t^{(2)} \quad (6)$$

这样,对于以 t 为起点的预测公式,其截距是

$$a_t = 2S_t^{(1)} - S_t^{(2)} \quad (7)$$

而 b_t 可用比较平滑的二次指数平均值单位变化率定义

$$b_t = S_t^{(2)} - S_{t-1}^{(2)} \quad (8)$$

为计算方便,用同期的二次指数平均值改写式(8),即

$$\begin{aligned} b_t &= S_t^{(2)} - S_{t-1}^{(2)} \\ &= \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} - S_{t-1}^{(2)} \\ &= \alpha(S_t^{(1)} - S_{t-1}^{(2)}) \\ &= \frac{\alpha}{1-\alpha}[(1-\alpha)S_t^{(1)} - (1-\alpha)S_{t-1}^{(2)}] \\ &= \frac{\alpha}{1-\alpha}[S_t^{(1)} - (\alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)})] \\ &= \frac{\alpha}{1-\alpha}[S_t^{(1)} - S_{t-1}^{(2)}] \end{aligned} \quad (9)$$

所以,利用式(7)和式(9),可以在第 t 期数据已知的前提下,预测第 $t+k$ 期的数据

$$\hat{y}_{t+k} = a_t + b_t k \quad (10)$$

3 仿真实验

本文使用 Matlab 编程实现上述双指数平滑预测器。实验数据取自 TAO(Tropical Atmosphere Ocean)测量的 2000 年 1 月 1 日~2000 年 12 月 31 日全年赤道附近海洋表面空气温度数据^[7],随机选取其中的 4 000 个连续观测值,采样精度为 10min。

决策模块使用最简单的阈值方法,即设定预测失败的判断门限为 0.1,当预测值和真实采样值的差异大于 0.1 时,判断为预测失败,重新调整预测模型的参数,并在新参数计算好之前传送所有的采样数据。

双指数平滑预测器需要事先设定记忆因子 α ,这是一个与应用特性紧密相关的经验参数,因此需要通过实验选取较为合适的 α 值。我们考察了不同 α 设定下,双指数平滑预测器对于样本序列的预测质量。评价指标包括数据发送率和平均相对误差(Mean Relative Deviation, MRD)。数据发送率是指因预测失败而发送的数据占全部采样数据的比例;而平均相对误差则定义为

$$\sum_{i=1}^w \frac{|f_i - \hat{f}_i|}{w f_i}$$

其中 \hat{f}_i 为对第 i 次采样的预测值, f_i 为第 i 次采样的真实值, w 为采样次数。

实验表明,当 α 在 0.3~0.5 之间时,预测失败的次数最少,数据发送率最低(见图 3);从平均相对误差的角度考虑,当 α 大于 0.2 后,随着 α 的增大,预测误差缓慢减小(见图 4)。因此,对于这样的监测应用,考虑到预测失败门限设置为 0.1 已经很小,从减少数据传送量的角度, α 取 0.4 较为合适。此时,可以滤除 83% 的冗余数据。图 5 是对随机选取的一段包含 200 个观测值的采样序列的预测结果展示,可以直观地看到其具有较好的预测质量(采样精度为 10min)。

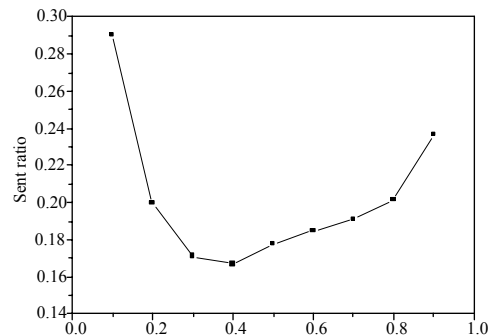


图 3 记忆因子对数据发送率的影响

(下转第 107 页)