

面向能量受限传感网的最优数据收集策略

陈朋朋,高守婉,黄丹,牛强

(中国矿业大学计算机科学与技术学院,江苏徐州 221116)

摘要: 本文研究了能量受限条件下无线传感器网络(wireless sensor networks, WSNs)的最优数据收集策略问题. 首先,传感器节点周期性采集数据并通过卡尔曼滤波器(Kalman filter, KF)对信息进行预处理以滤除噪声. 其次,考虑到通信为主要耗能环节,设计最优数据发送策略令节点在特定轮内发送数据,使得满足网络生存周期前提下,基站获得的数据精度最高. 具体来说,针对单跳网络,给出可使基站误差方差最小化的数据发送策略;在此基础上,进一步提出面向多跳网络的改进数据发送策略. 最后,利用仿真和原型实验验证所提策略的有效性.

关键词: 无线传感器网络; 数据收集; 能量效率; 数据精度; 卡尔曼滤波

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2017)08-2049-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.08.035

Optimal Data Collection Strategies for Energy-Constrained Sensor Networks

CHEN Peng-peng, GAO Shou-wan, HUANG Dan, NIU Qiang

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China)

Abstract: This paper investigates the problem of optimal data collection strategies for wireless sensor networks (WSNs) with limited resources. First, a sensor node periodically collects data and utilizes the Kalman filter (KF) to filter noise. Then, considering that communication is a primary source of energy consumption, we develop data transmission strategies to decide at which times the sensors should transmit their data. The proposed strategies can maximize the accuracy of data gathered by the base station for a required network lifetime. To be more specific, for single-hop networks, the optimal data transmission strategy is obtained which can minimize the estimation error. Based on this single-hop optimal strategy, the modified data transmission strategy is proposed for multi-hop networks. Simulation and prototype experiment results show the effectiveness of the presented strategies.

Key words: wireless sensor networks; data collection; energy efficiency; data accuracy; Kalman filtering

1 引言

近年来,无线传感器网络(wireless sensor networks, WSNs)在环境监测^[1]、目标跟踪^[2]、事件探测^[3]等领域得到广泛应用. 众所周知,WSNs的所有应用都基于数据收集来实现,所谓数据收集是指节点感知周围环境采集相关信息并发送至基站的过程. 现有WSNs中的数据收集方法大致可以分为三类:时间驱动数据收集、事件驱动数据收集和查询驱动数据收集. 本文关注第一种数据收集方法,通常也被称为周期性数据收集^[4]. 这

类数据收集方法中,时间被离散化为“轮”,一轮数据收集指节点将该轮时间间隔内采集的信息发送到基站. 在采用周期性数据收集的传感器网络(periodic sensor networks, PSNs)中,所有节点定期向基站汇报信息,因此PSNs适用于需要对指定区域持续进行数据采集的场景.

在PSNs系统设计中,主要面临两方面挑战:(1)节点能量有限使得网络生命周期难以满足实际应用需求;(2)节点采集到的数据含有大量噪声使得数据分析处理十分复杂. 在现有研究成果中,大部分主要关注如

收稿日期: 2016-08-21; 修回日期: 2017-03-29; 责任编辑: 马兰英

基金项目: 国家重点研发计划(No. 2016YFC060908); 国家自然科学基金(No. 51674255); 江苏省自然科学基金(No. BK20160274); 中国博士后科学基金特别资助项目(No. 2016T90523); 江苏省产学研前瞻性联合研究项目(No. BY2016026-03)

何高效利用节点的有限能量以延长网络寿命^[5-7],综合考虑网络寿命与数据精度的成果相对较少.文献[8]针对寿命受限的 PSNs 提出一种自适应数据收集策略,基本思想是节点根据误差情况动态调整数据发送.相较于相关工作,该策略可以在给定网络寿命的前提下显著提高基站数据精度.同样针对 PSNs,文献[9]给出一种基于节点剩余能量的自适应数据收集方法,该方法综合考虑了误差方差和随时间动态变化的节点剩余能量.另外,基于与文献[9]相同的目标,文献[10]针对目标跟踪 WSNs 提出了协同自适应信息收集策略,实验证明该策略可以在保证跟踪精度的同时延长网络寿命.除上述自适应方法,通过挖掘连续时刻和(或)相邻节点数据间的相关性,文献[11]和[12]分别提出了基于时空相关性的高效数据采集策略(EAST)和基于空间相关性的动态可扩展树(YEAST)以高效利用节点能量并保证数据精度.

尽管关于综合考虑节能与数据精度的收集策略研究已经出现一些成果,但现有工作并未给出最优折中算法.本文通过借鉴文献[13~15]提出的面向状态估计的传感器调度策略,研究能量受限条件下 PSNs 的最优数据收集问题,其基本思想是:考虑到基站的误差方差会因连续收不到更新数据而指数增长的事实^[13],在给定网络寿命的前提下,为节点设计合理数据发送策略使基站误差方差最小化.与已有工作相比,本文主要贡献如下:(1)将能量受限 PSNs 中的数据收集问题转化为具有一定限制条件的最优化问题,并给出一个离线算法用于计算最优数据收集策略;(2)针对单跳 PSNs,设计使基站误差方差最小化的最优数据发送策略;针对多跳 PSNs,给出相应改进策略;(3)基于 Matlab 平台进行仿真并利用 CC2530 节点构建原型系统,在数据精度、能量消耗等方面对所提数据收集策略进行性能评估.

2 系统概述

本文的研究对象为典型 PSNs 数据收集系统,主要考虑以下两方面问题:

(1)数据精度.传感器节点感知数据过程中,由于自身精度限制及外界环境影响,量测值与真实值之间存在误差.对传感器采集的数据预处理以提高精度是系统设计需要考虑的一个问题.

(2)节点能量.无线传感器节点通常由电池供电,并且往往处于难以及时补充能量的复杂环境中.如何高效利用节点的有限能量以延长网络生存期是系统设计需要考虑的另一问题.

针对数据精度问题,可以在传感器节点运行卡尔曼滤波器(Kalman filter, KF)^[16,17],将滤波后的估计值

代替真实量测数据传输给基站以减小误差.针对节点能量问题,可设计数据发送策略,使节点仅在指定轮内发送数据以延长网络寿命.然而,减少数据发送次数势必会降低数据精度,因此如何在保证网络生存期的前提下,尽量提高基站的数据精度是一个非常具有挑战性的问题.

本文中,网络生命周期被划分为若干大小相等的时间段,每个时间段称为“轮”.定义 \mathbf{x}_k 为第 k 轮待采集的真实数据, $\hat{\mathbf{x}}_k$ 为相应轮内基站获得的估计数据,并以误差方差 \mathbf{P}_k 来衡量基站的数据精度,即:

$$\mathbf{P}_k = E\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T\} \quad (1)$$

显然,误差方差 \mathbf{P}_k 越小意味着数据精度越高.本文的主要目标为设计最优调度策略决定节点在哪些轮内给基站或中继节点发送数据(称为数据发送策略),以达到在给定网络生存期下最小化 \mathbf{P}_k 的目的.

3 数据预处理——卡尔曼滤波

卡尔曼滤波是基于系统状态空间模型的估计方法,可以从受到干扰的量测信号中尽量准确地估计有用信号.不失一般性,考虑如下系统数学模型:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \quad (2)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \quad (3)$$

其中, $\mathbf{x}_k \in \mathbf{R}^n$ 为系统状态向量, $\mathbf{y}_k \in \mathbf{R}^m$ 为量测向量, \mathbf{A} 和 \mathbf{C} 分别为相应维数的系数矩阵.假设 $\{\mathbf{w}_k\}$ 和 $\{\mathbf{v}_k\}$ ($k = 0, 1, 2, \dots$) 为零均值白噪声序列,其方差分别为 $\mathbf{Q} \geq 0$ 和 $\mathbf{R} > 0$;系统初始状态 \mathbf{x}_0 的均值和方差分别为 \mathbf{u} 和 \mathbf{P}_0 ,且 \mathbf{x}_0 、 $\{\mathbf{w}_k\}$ 和 $\{\mathbf{v}_k\}$ 相互统计独立.基于上述模型的卡尔曼滤波算法递推公式如下^[18]:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} \quad (4)$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \quad (5)$$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}^T (\mathbf{C}\mathbf{P}_k^- \mathbf{C}^T + \mathbf{R})^{-1} \quad (6)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-) \quad (7)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{C})\mathbf{P}_k^- \quad (8)$$

我们以一个室内温度监测实验为例验证卡尔曼滤波对数据的预处理效果,该实验所采用的温度传感模块精度为 $\pm 0.5^\circ\text{C}$.室内温度在某时间段内可认为相对恒定,但受到墙体隔热、通风情况等因素影响,会产生 $\pm 0.1^\circ\text{C}$ 的随机波动.基于上述分析,数学模型(2)和(3)的相关参数如下: $\mathbf{A} = \mathbf{C} = 1$, $\mathbf{Q} = 0.01$, $\mathbf{R} = 0.25$.运行式(4)~(8)给出的卡尔曼滤波算法,得到如图1所示的温度估计曲线,可看出实际测量温度数据经过卡尔曼滤波算法预处理后,准确性得到了大幅度提高.

4 单跳网络最优数据发送策略

在提高节点数据精度的基础上,本节针对单跳网络设计最优数据发送策略,在能量受限条件下最小化

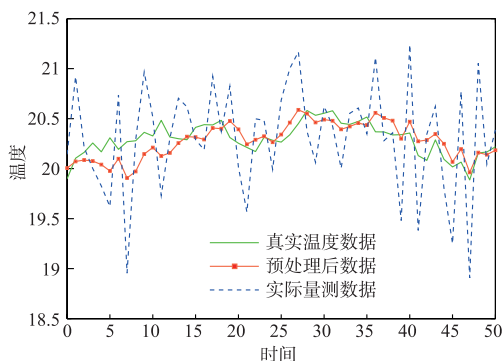


图1 室内温度曲线

基站数据的方差. 该策略可进一步扩展到第 5 节的多跳网络中.

4.1 单跳网络数据发送问题描述

本文采用具有一定计算能力的传感器节点, 节点周期性感知外界数据并进行卡尔曼滤波, 然后将滤波结果发送至远程基站. 为达到节能效果, 设计数据发送策略 θ , 使节点仅在指定轮内发送数据. 利用 $\gamma_k = 1$ 或者 0 表示节点在第 k 轮数据收集中是否发送数据, 具体来说: $\gamma_k = 1$ 时, 节点通过无线方式发送卡尔曼滤波值, 基站接收来自传感器的更新数据; $\gamma_k = 0$ 时, 节点仍感知数据并执行滤波算法, 但不将数据发送出去, 此时, 基站根据自己在 $(k-1)$ 轮获得的数据对第 k 轮数据进行预测. 很明显, 调度策略 θ 的实质就是确定每一轮的 γ 值.

为方便表述, 下文用 \hat{x}_k^* 和 P_k^* 分别表示传感器节点在第 k 轮的估计值和方差. 进一步, 定义

$$h(X) \triangleq AXA^T + Q \quad (9)$$

根据上述定义和分析, 基站在第 k 轮的估计值 \hat{x}_k 和相应方差 P_k 为:

$$(\hat{x}_k, P_k) = \begin{cases} (\hat{x}_k^*, P_k^*), & \gamma_k = 1 \\ (A\hat{x}_{k-1}, h(P_{k-1})), & \gamma_k = 0 \end{cases} \quad (10)$$

考虑到实际系统的数学模型通常具有能控性和能观性, 不失一般性, 假设 (A, \sqrt{Q}) 能控以及 (C, A) 能观. 基于上述假设, 卡尔曼滤波递推公式 (8) 中的方差 P_k 将很快收敛至唯一半正定矩阵 \bar{P} ^[18]. 另一方面, 数据收集系统往往需要长时间工作, 据此下文忽略 KF 进入稳定状态之前的短暂时间. 假设传感器的 KF 已进入稳定状态, 即 $P_k^* = \bar{P}$, 根据式 (10) 可将基站方差写成如下形式:

$$P_k = \begin{cases} \bar{P}, & \gamma_k = 1 \\ h(P_{k-1}), & \gamma_k = 0 \end{cases} \quad (11)$$

本节关注的单跳数据发送策略问题可以转化为一定约束条件下的最优化问题. 假设节点总的通信能量预算为 J_{tot} , 该节点需要工作 N 轮, 于是可以得到每轮的平均能量预算为:

$$\varphi = \frac{J_{tot}}{N} \quad (12)$$

给定发送策略 θ , 定义 $J(\theta)$ 和 $P_a(\theta)$ 为该策略下节点的实际平均通信能耗和基站的平均误差方差, 即:

$$J(\theta) \triangleq \frac{1}{N} \sum_{k=1}^N \gamma_k \varepsilon \quad (13)$$

$$P_a(\theta) \triangleq \frac{1}{N} \sum_{k=1}^N P_k \quad (14)$$

其中, ε 为每轮数据发送的相关能耗. 我们的目标为寻找周期性数据发送策略 θ , 使如下数学问题具有最优解.

问题 1 $\min_{\theta} P_a(\theta)$ s. t. $J(\theta) \leq \varphi$.

问题 1 的最优解 θ^* 称为最优数据发送策略. 在满足能量预算 φ 的前提下, 对于任意数据发送策略 θ , 有 $P_a(\theta^*) \leq P_a(\theta)$.

4.2 最优数据发送策略

下面针对单跳网络, 给出周期性最优数据发送策略 θ^* . 为方便描述, 定义 $q(\theta)$ 为数据发送率, 即

$$q(\theta) = \frac{n_s}{N} = \frac{1}{N} \sum_{k=1}^N \gamma_k(\theta) \quad (15)$$

其中, n_s 为节点在生命周期 N 内发送数据总次数. 根据发送次数越多 (能耗越大) 数据准确性越高的原则^[13], 可以得到 $J(\theta^*) = \varphi$. 结合式 (12), 可得到

$$q(\theta^*) = \frac{\varphi}{\varepsilon} \quad (16)$$

为简化符号, 下文中将 $q(\theta^*)$ 记作 q^* . 首先把 q^* 写为两个互质整数相比的形式, 即 $q^* = \alpha/\beta$, ($\alpha < \beta$); 进一步, 令 μ 为满足 $\mu \leq \beta/\alpha$ 的最大整数, 并定义 η_1 和 η_2 分别为满足图 2 的数据发送策略.

图 2 中, 箭头 ' \uparrow ' 表示发送数据, 圆圈 ' \circ ' 表示不发送. 在此基础上, 给出如图 3 所示周期为 β 的最优数据发送策略 θ^* .

需要说明的是, 交换图 3 中 θ^* 各个组成部分的位置不影响其最优性. 相应的最小平均误差方差为:

$$P_a(\theta^*) = \frac{1}{\beta} \left[\alpha \sum_{i=0}^{\mu-1} h^i(\bar{P}) + (\beta - \alpha\mu) h^\mu(\bar{P}) \right] \quad (17)$$

其中 $h^0(\bar{P}) = \bar{P}$, $h^{i+1}(\bar{P}) = h(h^i(\bar{P}))$.

证明 上述最优数据发送策略的证明过程类似于文献[13]中定理 5.2 的证明, 故不再赘述.

5 多跳网络改进数据发送策略

由于通信范围有限, 节点的数据往往需要通过多跳方式转发至基站. 即节点周期性感知本地数据, 经过 KF 预处理后发送给中继节点; 中继节点同样也感知并预处理本地数据, 进一步将所有数据融合后发送至基

站. 基于节能考虑, 本节为多跳网络中的节点设计数据发送策略.

5.1 多跳网络数据发送问题描述

η_1	↑	○	...	○	
η_2	↑	○	...	○	○
轮	1	2	...	μ	$\mu+1$

图2 数据发送策略 η_1 和 η_2

$$\mathbf{x}_{k+1}^i = \mathbf{A}_i \mathbf{x}_k^i + \mathbf{w}_k^i \quad (18)$$

$$\mathbf{y}_k^i = \mathbf{C}_i \mathbf{x}_k^i + \mathbf{v}_k^i, i \in \{S_1, S_2\} \quad (19)$$

其中, \mathbf{w}_k^i 和 \mathbf{v}_k^i 的方差分别为 \mathbf{Q}_i 和 \mathbf{R}_i , 除此之外, 系统中随机变量的统计特性沿用第 3 节中假设. 相应地, 定义

$$h_i(\mathbf{X}) \triangleq \mathbf{A}_i \mathbf{X} \mathbf{A}_i^T + \mathbf{Q}_i, i \in \{S_1, S_2\} \quad (20)$$

我们用 $\gamma_k^{S_2} = 1$ 或 0 表示叶子节点 S_2 在第 k 轮是否发送数据. 另外, 假设所有节点的 KF 已达到稳定状态, 即节点 i 的误差方差为稳定状态误差方差 $\bar{\mathbf{P}}_i, i \in \{S_1, S_2\}$. 根据 4.1 节中的分析, 不难发现中继节点 S_1 在第 k 轮收集到的数据和误差方差分别为:

$$\hat{\mathbf{x}}_k^{S_1} = [\hat{\mathbf{x}}_k^{S_1, loc}; \hat{\mathbf{x}}_k^{S_2}] = \begin{cases} [\hat{\mathbf{x}}_k^{S_1, loc}; \hat{\mathbf{x}}_k^{S_2, loc}], & \gamma_k^{S_2} = 1 \\ [\hat{\mathbf{x}}_k^{S_1, loc}; \mathbf{A}_{S_2} \hat{\mathbf{x}}_{k-1}^{S_2}], & \gamma_k^{S_2} = 0 \end{cases} \quad (21)$$

$$\mathbf{P}_k^{S_1} = [\bar{\mathbf{P}}_{S_1}; \mathbf{P}_k^{S_2}] = \begin{cases} [\bar{\mathbf{P}}_{S_1}; \bar{\mathbf{P}}_{S_2}], & \gamma_k^{S_2} = 1 \\ [\bar{\mathbf{P}}_{S_1}; h_{S_2}(\mathbf{P}_{k-1}^{S_2})], & \gamma_k^{S_2} = 0 \end{cases} \quad (22)$$

其中, $\hat{\mathbf{x}}_k^{i, loc}$ 为节点 i 对本地量测数据预处理后得到的估计值. 进一步, 利用 $\gamma_k^{S_1} = 1$ 或 0 表示 S_1 在第 k 轮是否发送数据, 基站获得的数据和误差方差可以表示为:

$$\hat{\mathbf{x}}_k = [\hat{\mathbf{x}}_k^{S_1'}; \hat{\mathbf{x}}_k^{S_2'}] = \begin{cases} [\hat{\mathbf{x}}_k^{S_1, loc}; \hat{\mathbf{x}}_k^{S_2}], & \gamma_k^{S_1} = 1 \\ [\mathbf{A}_{S_1} \hat{\mathbf{x}}_{k-1}^{S_1'}; \mathbf{A}_{S_2} \hat{\mathbf{x}}_{k-1}^{S_2'}], & \gamma_k^{S_1} = 0 \end{cases} \quad (23)$$

$$\mathbf{P}_k = [\mathbf{P}_k^{S_1'}; \mathbf{P}_k^{S_2'}] = \begin{cases} [\bar{\mathbf{P}}_{S_1}; \mathbf{P}_k^{S_2}], & \gamma_k^{S_1} = 1 \\ [h_{S_1}(\mathbf{P}_{k-1}^{S_1'}); h_{S_2}(\mathbf{P}_{k-1}^{S_2'})], & \gamma_k^{S_1} = 0 \end{cases} \quad (24)$$

5.2 改进数据发送策略

本小节以图 4 所示的简单两跳网络为例, 在不同能量预算下, 给出多跳网络中节点的数据发送策略. 首先, 假设图 4 中节点 S_1 和 S_2 的数据发送率为 $q_{S_1}^* = q_{S_2}^* = q^* = \frac{3}{10}$, S_1 和 S_2 均采用 4.2 节给出的周期性单跳最优数据发送策略, 如图 5 所示(以一个周期为例), 每个节点都在第 1、4、7 轮发送数据.

与单跳网络不同, 多跳网络中对于中继节点来说, 除与基站通信需要消耗能量外, 接收来自叶子节点的数据也要消耗能量; 而对于叶子节点来说, 仅需考虑将数据发送至中继节点的能量消耗. 如果期望图 4 中网络的生存期为 N , 忽略睡眠耗能, 节点 S_1 和 S_2 的能量

考虑图 4 所示的简单两跳网络, 假设各传感器节点的 KF 对应如下数据采集模型:

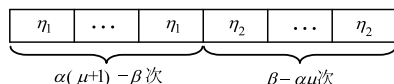


图3 最优数据发送策略 θ^*



图4 简单两跳网络

S_1	↑	○	○	↑	○	○	↑	○	○	○
S_2	↑	○	○	↑	○	○	↑	○	○	○
轮	1	2	3	4	5	6	7	8	9	10

图5 直接应用单跳数据发送策略($q_{S_1}^* = q_{S_2}^*$)

预算应该分别满足:

$$J_{S_1} = (\varepsilon_{S_1} + \varepsilon_{rec}) \times N \times q^* \quad (25)$$

$$J_{S_2} = \varepsilon_{S_2} \times N \times q^* \quad (26)$$

其中, ε_{S_1} 为中继节点 S_1 将数据发送给基站消耗的能量, ε_{rec} 为 S_1 接收来自叶子节点 S_2 的数据消耗的能量. 相应地, ε_{S_2} 为叶子节点 S_2 发送一次数据所消耗的能量. 值得注意的是, 图 5 给出的数据发送策略对于图 4 中的每个节点都是最优的, 而保证最优性的前提条件为 S_1 和 S_2 的数据发送率相同. 换言之, 若令 $\xi = J_{S_1}/J_{S_2}$, 如果实际给出的 S_1 和 S_2 的能量预算 $\bar{J}_{S_1}/\bar{J}_{S_2} = \xi$, 则可以在每个节点直接运行单跳最优数据发送策略; 如果实际能量预算 $\bar{J}_{S_1}/\bar{J}_{S_2} \neq \xi$, 则需要对数据发送策略重新讨论.

(1) S_1 和 S_2 的能量预算比例小于 ξ

S_1 和 S_2 的能量预算比例小于 ξ 意味着 S_1 的数据发送率小于 S_2 . 然而事实上, 对于 4.2 节提出的最优数据发送策略来说, 叶子节点 S_2 的数据发送率无需高于中继节点 S_1 , 即便高出, 对提高基站数据精度也没有帮助, 下面将举例说明. 假设 S_1 的数据发送率 $q_{S_1}^* = 3/11$, S_2 的数据发送率 $q_{S_2}^* = 6/11$, 在 S_1 和 S_2 上分别运行单跳最优数据发送策略, 得到图 6. 观察图 6, 在 1、4、8 轮, S_1 和 S_2 都发送数据, 在 2、6、10 轮, 仅 S_2 发送数据. 因为 S_1 在 2、6、10 轮并不发送数据, 结合式 (22) 和 (24) 可以推断出 S_2 在这三轮里是否发送数据对基站的误差方差没有影响; 另外, 假设 S_1 接收 S_2 发来的全部数据, S_2 在上述三轮发送数据也会增大 S_1 的能耗.

S_1	↑	○	○	↑	○	○	○	↑	○	○	○
S_2	↑	↑	○	↑	○	↑	○	↑	○	↑	○
轮	1	2	3	4	5	6	7	8	9	10	11

图6 直接应用单跳数据发送策略($q_{S_1}^* < q_{S_2}^*$)

因此,可以选择让 S_2 在 2、6、10 轮不发送数据,即让 S_2 运行与 S_1 相同的数据发送策略。

基于上述分析,对于 $q_{s_1}^* < q_{s_2}^*$ 的情况,在中继节点 S_1 上直接运行单跳最优数据发送策略,而令叶子节点 S_2 采用与 S_1 相同的发送策略。采用上述方法,显然 S_2 的能量会有冗余,即对于 S_2 来说,该数据发送策略并不是最优的,但是在中继节点能量受限的前提下,该方法可以尽量提高基站的数据精度。

(2) S_1 和 S_2 的能量预算比例大于 ξ

下面讨论 S_1 和 S_2 的能量预算比例大于 ξ 的情况。假设 $q_{s_1}^* = 2/3, q_{s_2}^* = 2/5$, 在 S_1 和 S_2 上分别运行单跳最优数据发送策略,得到图 7。

根据图 7,并结合式(22)和(24)可以得到,基站的平均误差方差为:

$$P_a = \left\{ \frac{1}{3}(2\bar{P}_{S_1} + h_{S_1}(\bar{P}_{S_1})), \right. \\ \left. \frac{1}{15}(4\bar{P}_{S_2} + 6h_{S_2}(\bar{P}_{S_2}) + 4h_{S_2}^2(\bar{P}_{S_2}) + h_{S_2}^3(\bar{P}_{S_2})) \right\} \quad (27)$$

S_1	↑	○	↑	↑	○	↑	↑	○	↑	↑	○	↑	↑	○	↑
S_2	↑	○	↑	○	○	↑	○	○	↑	○	↑	○	○	↑	○
轮	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

图7 直接应用单跳数据发送策略($q_{s_1}^* > q_{s_2}^*$)

注意到 $h_i^t(\bar{P}_i) \leq h_i^{t_2}(\bar{P}_i), 0 \leq t_1 \leq t_2, i \in \{S_1, S_2\}$ (文献[13]中引理 3.1), 很明显,若要减小 P_a , 必须尽量减小高次 $h_i(\cdot)$ 出现的次数。而式(27)中出现了 $h_{S_2}^3(\bar{P}_{S_2})$ 和 4 次 $h_{S_2}^2(\bar{P}_{S_2})$, 其原因是在第 8 和 11 轮, 节点 S_2 向中继节点发送数据, 但是中继节点并未向基站转发数据。为了使 S_2 每轮数据发送都对基站的数据精度有最大程度贡献, 应该改进 S_2 的数据发送策略, 尽量让 S_2 在 S_1 发送数据的轮内发送数据。

为方便下文描述, 定义序列 $\{t_k\}_{k \geq 1}$ 表示节点 S_1 在第 t_k 轮发送数据:

$$t_1 = 1 \\ t_2 = \inf\{k | k > t_1, \gamma_k^{S_1} = 1\} \\ \vdots \\ t_k = \inf\{k | k > t_{k-1}, \gamma_k^{S_1} = 1\}$$

基于尽量提高基站数据精度的考虑, 节点 S_2 只能在指定的 t_k 轮内发送数据, 换言之, S_2 的改进发送策略 $\bar{\theta}_{S_2}^*$ 的实质为确定 $\gamma_k^{S_2} = 1$ 还是 0。

下面将给出改进发送策略 $\bar{\theta}_{S_2}^*$ 。首先, 将 $q_{s_1}^*$ 和 $q_{s_2}^*$ 分别写成互质数相比的形式, 即 $q_{s_1}^* = \alpha_{s_1}/\beta_{s_1}, q_{s_2}^* = \alpha_{s_2}/\beta_{s_2}$ 。令 $\bar{\beta}$ 为 β_{s_1} 和 β_{s_2} 的最小公倍数, $\bar{\beta}$ 即为改进发送策略 $\bar{\theta}_{S_2}^*$ 的周期。在周期 $\bar{\beta}$ 内, 节点 S_1 发送数据的次数为 $n_{s_1} =$

$\alpha_{s_1}\bar{\beta}/\beta_{s_1}, S_2$ 发送数据的次数为 $n_{s_2} = \alpha_{s_2}\bar{\beta}/\beta_{s_2}$ 。因为改进发送策略是在 t_k 的基础上进行调度, 所以可认为其数据发送率 $\bar{q}_{s_2}^* = n_{s_2}/n_{s_1}$, 进一步, 将 $\bar{q}_{s_2}^*$ 写成两个互质数相比的形式 $\bar{q}_{s_2}^* = \alpha'/\beta'$ 。在上述分析和单跳最优数据发送策略的基础上, 同样针对 $q_{s_1}^* = 2/3, q_{s_2}^* = 2/5$ 的情况, 得到如图 8 所示的改进发送策略。

采用图 8 所示的改进数据发送策略, 基站的平均误差方差为:

$$P_a = \left\{ \frac{1}{3}(2\bar{P}_{S_1} + h_{S_1}(\bar{P}_{S_1})), \right. \\ \left. \frac{1}{15}(6\bar{P}_{S_2} + 5h_{S_2}(\bar{P}_{S_2}) + 4h_{S_2}^2(\bar{P}_{S_2})) \right\} \quad (28)$$

S_1	↑	○	↑	↑	○	↑	↑	○	↑	↑	○	↑	↑	○	↑
S_2	↑	○	↑	○	○	↑	○	○	↑	↑	○	○	↑	○	○
轮	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

图8 改进数据发送策略($q_{s_1}^* > q_{s_2}^*$)

通过与式(27)进行比较, 并注意到 $h_i^{t_1}(\bar{P}_i) \leq h_i^{t_2}(\bar{P}_i), 0 \leq t_1 \leq t_2$, 我们发现: 节点 S_1 采用的数据发送策略不变, 所以在基站端由 S_1 采集的相关数据精度没有变化; 而对于节点 S_2 来说, 通过运用改进数据发送策略, 在能耗不变的前提下, 基站相应的数据精度得到了明显提高。

观察图 8 可以得到, 在 $q_{s_1}^* > q_{s_2}^*$ 的情况下, S_1 接收数据的次数小于发送数据的次数。因此, 根据节点能量预算和网络生命周期 N 计算 $q_{s_1}^*$ 和 $q_{s_2}^*$ 时, 首先通过 $n_{s_2} = \bar{J}_{S_2}/\varepsilon_{S_2}$ 计算叶子节点 S_2 在生命周期 N 内发送数据的总次数; 然后, 利用 $n_{s_1} = (\bar{J}_{S_1} - n_{s_2}\varepsilon_{rec})/\varepsilon_{S_1}$ 计算中继节点 S_1 发送数据的总次数; 最后, 根据 $q_{s_1}^* = n_{s_1}/N$ 和 $q_{s_2}^* = n_{s_2}/N$ 分别求出 S_1 和 S_2 的数据发送率。需要指出的是, 本节仅给出具有一个叶子节点的简单情况, 对于采用经典 LEACH 协议的网络, 可以在簇首上运行单跳最优数据发送策略, 在所有成员节点上运行改进数据发送策略。

6 性能评估

在这一节里, 我们通过 Matlab 仿真和原型实验对本文提出的数据收集策略进行性能评估。

6.1 仿真评估

本节仿真中, 根据文献[8], 默认网络中所有节点发送一次数据耗能 20nAh, 接收一次数据耗能 8nAh。另外, 某些参数设置可能会随仿真目标或场景的变化而发生改变, 下文中将具体说明。

6.1.1 单跳网络数据发送策略性能评估

本节中, 将节点的初始能量设置为允许节点发送

2000 次数据. 为验证 4.2 节提出的单跳最优数据发送策略 θ^* 的有效性, 构建周期性数据发送策略 θ 与之比较. 考虑到公平性, 令 θ 与 θ^* 具有相同的数据发送率 $q^* = \alpha/\beta$. 同时, 为简化描述, 使 θ 的形式为: 在周期 β 内, α 次数据发送 '↑' 集中在前半部分.

在仿真中, 数据发送率 q^* 随网络的寿命需求 N 发生变化, 具体可由 $q^* = 2000/N$ 计算. 设置系统参数 $A = C = 1, Q = 0.25, R = 1$, 图 9 给出了不同数据发送率下, 两种策略的平均误差方差 $P_a(\theta^*)$ 与 $P_a(\theta)$ 的比较. 可以看出, 当数据发送率很大 ($q^* = 11/11$ 或者 $9/11$) 或者很小 ($q^* = 1/11$) 时, 两种策略的误差基本相同; 当数据发送率适中时 ($q^* = 3/11$ 或 $5/11$ 或 $7/11$), $P_a(\theta^*)$ 明显小于 $P_a(\theta)$. 实际应用中, 通常不会将数据发送率设置得很大 (能耗高) 或者很小 (精度差), 因此, 单跳最优数据发送策略 θ^* 在实际应用中具有较大的优势.

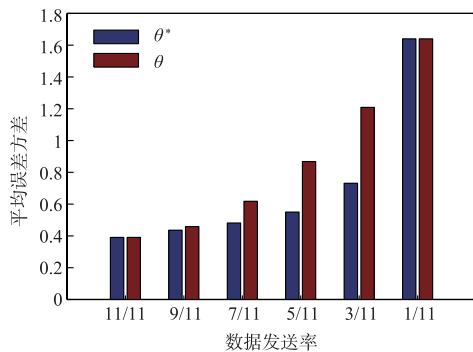


图9 θ 与 θ^* 平均误差方差比较

6.1.2 多跳网络数据发送策略性能评估

本节中, 分别针对中继节点 S_1 和叶子节点 S_2 的能量预算比例 $\bar{J}_{S_1}/\bar{J}_{S_2} < \xi$ 以及 $\bar{J}_{S_1}/\bar{J}_{S_2} > \xi$ 的情况, 评估 5.2 节给出的改进数据发送策略的性能.

(1) S_1 和 S_2 能量预算比例小于 ξ 的情况

首先, 将叶子节点 S_2 的通信能量设置为 $20 \times 1000 = 2.0 \times 10^4$ nAh, 即允许 S_2 向 S_1 发送 1000 次数据; 另外, 将中继节点 S_1 的通信能量设置为 $(20 + 8) \times 500 = 1.4 \times 10^4$ nAh, 即允许 S_1 向基站发送 500 次数据, 并允许 S_1 接收 500 次来自 S_2 的数据. 进一步, 设置网络寿命需求为 $N = 1300$. 为了评估 5.2 节中数据发送策略 θ^* 的性能, 按照文献[15]给出的 OPS 算法构建数据发送策略 θ 与之进行比较. 图 10 给出了两种策略下的数据误差对比情况. 仿真结果表明, θ^* 可以使基站获得较高的数据精度, 但在该策略下, 中继节点接收到的来自 S_2 的数据精度有所降低. 通常情况下, 仅需考虑基站的数据精度, 因此, θ^* 的有效性得到验证.

(2) S_1 和 S_2 能量预算比例大于 ξ 的情况

下面仿真中, 叶子节点 S_2 通信能量预算为 $20 \times$

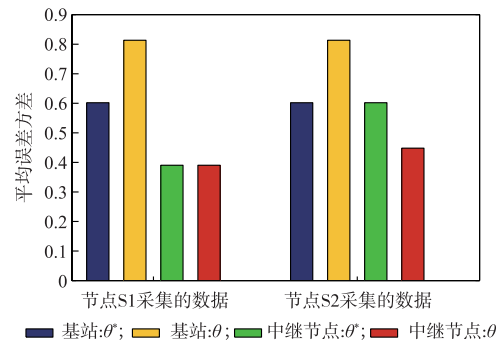


图10 θ 与 θ^* 平均误差方差比较($\bar{J}_{S_1}/\bar{J}_{S_2} < \xi$)

500 = 1.0×10^4 nAh, 即允许 S_2 向 S_1 发送 500 次数据; 中继节点 S_1 的通信能量预算为 $20 \times 900 + 8 \times 500 = 2.2 \times 10^4$ nAh, 即允许 S_1 向基站发送 900 次数据, 同时允许 S_1 接收 500 次来自叶子节点的数据. 假设网络寿命需求为 $N = 1500$. 根据 OPS 算法^[15] 和本文 5.2 节给出的改进数据发送策略分别构建发送策略 θ 和 θ^* , 并在下图中对两种策略的平均误差方差进行比较. 由图 11 可知, 对于所关注的基站数据, 其中来源于 S_1 的部分在两种发送策略下精度基本相同; 而来源于 S_2 的部分在本文给出的策略 θ^* 下精度明显提高, 说明了 θ^* 的有效性.

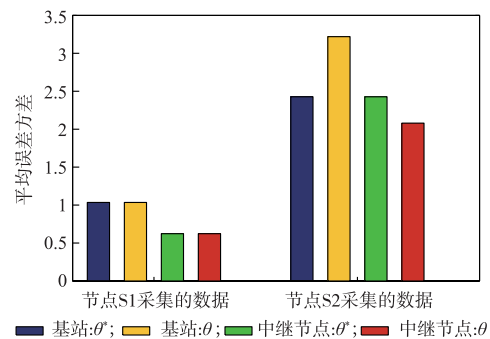


图11 θ 与 θ^* 平均误差方差比较($\bar{J}_{S_1}/\bar{J}_{S_2} > \xi$)

6.2 原型实验

本文利用 8 个 CC2530 节点在实验室搭建一个小型室内环境监测原型系统. 如图 12(a) 所示, 利用一个节点充当基站, 其余 7 个节点负责采集室内环境数据 (以温度为例). 系统拓扑结构如图 12(b) 所示. 传感器节点配置的温度感知模块量程为 $-20^\circ\text{C} \sim +130^\circ\text{C}$, 精度为 $\pm 0.5^\circ\text{C}$. 设置传感器每 5s 采集一次数据, 图 13 比较了数据发送率为 70% 和 30% 时基站端的误差曲线, 其中第 k 轮误差的计算方法为:

$$e_k = |\hat{\mathbf{x}}_k^1 - \hat{\mathbf{x}}_k^q|, k = 1, 2, \dots \quad (29)$$

上式中, $\hat{\mathbf{x}}_k^1$ 表示发送率为 100% (即每轮都发送数据) 时基站获得的数据, $\hat{\mathbf{x}}_k^q$ 表示发送率为 $q < 1$ 时的基站数据. 为使实验结果具有代表性, 图 13 中的曲线为 7 个传感器的平均结果. 进一步, 每个传感器运行 500 轮, 得到

发送率为 70% 和 30% 两种情况下的平均误差分别为: $\bar{e}_a^{0.7} = 0.0628$, $\bar{e}_a^{0.3} = 0.1779$. 考虑到节点精度为 0.5, 利用 $\lambda = \frac{\bar{e}_a}{0.5}$ 衡量精度损失, 可以算出当节能 30% 和 70% 时, 基站端数据精度损失分别 12.6% 和 35.6%.

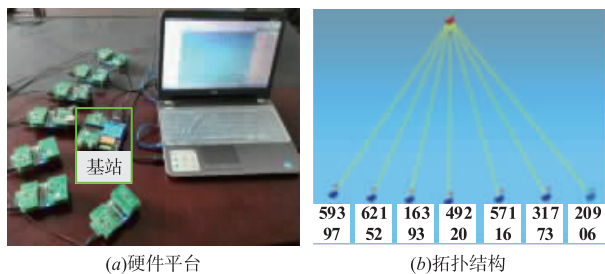


图12 原型系统

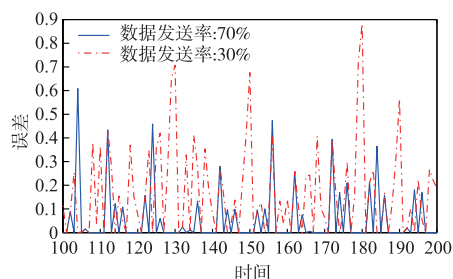


图13 基站误差比较

需要指出的是, 通过大量实验和仿真研究, 我们发现在节点节能程度相同的情况下, 基站在实验中的数据精度要低于其仿真精度. 原因在于, 实验中的系统建模以及噪声统计特性提取过程中均不可避免存在误差, 而本文给出的数据收集策略会受到上述两个因素的影响.

7 总结

本文分析了能量受限条件下 WSNs 的数据收集策略问题. 在给定节点能量预算和网络生存期的前提下, 设计相关策略决定节点应该在哪些轮内发送数据, 使得基站数据精度最高. 针对单跳网络和多跳网络两种情况, 分别给出了单跳最优数据发送策略和多跳改进发送策略, 并通过仿真和实验结果验证了所提策略的有效性. 考虑到实际应用, 该方法还存在若干问题, 例如路由协议对算法的影响以及无线通信的不可靠性等. 在后续工作中, 将对算法进一步研究完善.

参考文献

- [1] Al-Turjman F, Hassanein H, Ibnkahla M. Towards prolonged lifetime for deployed WSNs in outdoor environment monitoring [J]. *Ad Hoc Networks*, 2015, 24 (PA): 172 – 185.
- [2] Cao N, Brahma S, Varshney P K. Target tracking via crowdsourcing: A mechanism design approach [J]. *IEEE Transactions on Signal Processing*, 2015, 63 (6): 1464 – 1476.
- [3] Xie M, Hu J, Guo S. Segment-based anomaly detection with approximated sample covariance matrix in wireless sensor networks [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2015, 26(2): 574 – 583.
- [4] Harb H, Makhoul A, Tawil R, et al. A suffix-based enhanced technique for data aggregation in periodic sensor networks [A]. *Proceedings of the 10th International Wireless Communications and Mobile Computing Conference [C]*. New York: IEEE, 2014. 494 – 499.
- [5] Luo D, Zhu X, Wu X, et al. Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks [A]. *Proceedings of the 2011 IEEE INFOCOM [C]*. Piscataway, NJ: IEEE, 2011. 1566 – 1574.
- [6] Imon S K A, Khan A, Di Francesco M, et al. Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks [J]. *IEEE/ACM Trans on Networking*, 2015, 23(5): 1401 – 1415.
- [7] Abdulla A E A A, Fadlullah Z M, Nishiyama H, et al. Toward fair maximization of energy efficiency in multiple UAS-Aided networks: A game-Theoretic methodology [J]. *IEEE Transactions on Wireless Communications*, 2015, 14 (1): 305 – 316.
- [8] Tang X, Xu J. Adaptive data collection strategies for lifetime-constrained wireless sensor networks [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2008, 19(6): 721 – 734.
- [9] Makhoul A, Harb H, Laiymani D. Residual energy-based adaptive data collection approach for periodic sensor networks [J]. *Ad Hoc Networks*, 2015, 35: 149 – 160.
- [10] Feng J, Lian B, Zhao H. Coordinated and adaptive information collecting in target tracking wireless sensor networks [J]. *IEEE Sensors Journal*, 2015, 15 (6): 3436 – 3445.
- [11] Villas L A, Boukerche A, Guidoni D L, et al. An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks [J]. *Computer Communications*, 2013, 36(9): 1054 – 1066.
- [12] Villas L A, Boukerche A, De Oliveira H A, et al. A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks [J]. *Ad Hoc Networks*, 2014, 12(1): 69 – 85.
- [13] Shi L, Cheng P, Chen J. Sensor data scheduling for optimal state estimation with communication energy constraint [J]. *Automatica*, 2011, 47(8): 1693 – 1698.
- [14] Zhang H, Cheng P, Shi L. Optimal denial-of-service attack scheduling with energy constraint [J]. *IEEE Transactions*

- on Automatic Control, 2015, 60 (11): 3023 – 3028.
- [15] Ren Z, Cheng P, Chen J, et al. Optimal periodic sensor schedule for steady-state estimation under average transmission energy constraint [J]. IEEE Transactions on Automatic Control, 2013, 58(12): 3265 – 3271.
- [16] Jain A, Chang E Y, Wang Y F. Adaptive stream resource management using Kalman Filters [A]. Proceedings of the ACM SIGMOD International Conference on Management of Data [C]. New York: ACM, 2004. 11 – 22.
- [17] Min J-K, Chung C-W. EDGES: Efficient data gathering in sensor networks using temporal and spatial correlations [J]. Journal of Systems and Software, 2010, 83(2): 271 – 282.
- [18] Anderson B D O, Moore J. Optimal Filtering [M]. Englewood Cliffs, NJ, USA: Prentice – Hall, 1979.

作者简介



陈朋朋 男, 1983 年 11 月出生于山东省济宁市. 现为中国矿业大学计算机科学与技术学院副教授、硕士生导师, 研究方向为传感器网络、大数据等.
E-mail: chenp@cumt.edu.cn



高守婉 (通信作者) 女, 1982 年 10 月出生于山东省烟台市. 现为中国矿业大学计算机科学与技术学院讲师, 研究方向为网络控制系统、信号估计等.
E-mail: gaoshouwan@cumt.edu.cn



黄丹 女, 1991 年 10 月出生于江西省上饶市. 现为中国矿业大学计算机科学与技术学院硕士研究生, 研究方向为传感器网络.
E-mail: 791843924@qq.com



牛强 男, 1974 年 7 月出生于辽宁省沈阳市. 现为中国矿业大学计算机科学与技术学院教授、博士生导师, 研究方向为物联网、智能信息处理等.
E-mail: niuq@cumt.edu.cn