

# 无线传感器网络中最小化通信开销的近似监测算法

毕 冉<sup>1)</sup> 李建中<sup>2)</sup> 高 宏<sup>2)</sup>

<sup>1)</sup>(大连理工大学计算机科学与技术学院 辽宁 大连 116000)

<sup>2)</sup>(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

**摘 要** 聚集值的约束违反监测有助于检测监测区域内异常事件的发生. 传感器节点由于电源能量耗尽的原因经常失效或废弃, 因此研究无线传感网的高能效的监测算法具有重要意义. 已有的研究工作给出了启发式的阈值设置方法, 通信能量开销较高并且没有考虑感知数据的特性, 忽略了噪声对感知数据的影响. 鉴于上述原因, 以最小化通信开销为优化目标, 文中开展了过滤器最优阈值的研究. 首先, 文中提出了通信开销模型, 给出了过滤器最优阈值问题的形式化定义, 其次, 通过一维随机游走来模拟感知数据的变化并推导出过滤器失效概率的计算公式, 通过拉格朗日乘子法给出了求解过滤器最优阈值的数学方法, 第三, 提出了低计算复杂度的过滤器最优阈值近似算法并分析了近似算法的近似比. 理论分析和实验结果验证了提出的算法的正确性以及高效性.

**关键词** 无线传感器网络; 监测; 约束违反; 近似算法; 物联网

**中图法分类号** TP393 **DOI号** 10.11897/SP.J.1016.2015.02092

## Approximate Monitoring Algorithm for Minimizing Communication Cost in Wireless Sensor Networks

BI Ran<sup>1)</sup> LI Jian-Zhong<sup>2)</sup> GAO Hong<sup>2)</sup>

<sup>1)</sup>(Department of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116000)

<sup>2)</sup>(Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

**Abstract** Monitoring constraint violation of aggregation result is very useful for detecting abnormal events, which happen in the monitored region. Due to battery energy getting exhausted, sensor node becomes invalid and gets out of use, hence investigating into energy efficient monitoring algorithms for constraint violation plays a significant role in the area of sensor networks. Previous works research on the heuristic schemes for the design of local filters, which bring about higher communication consumption and ignore the effect of measurement error and noise on the sensed data. Due to the above reasons, this paper presents an investigation into the problem of optimal filters for minimizing communication cost. Firstly, this paper presents a model for communication overhead and provides a formal description for the problem of optimal filters. Secondly, based on the random walk in one dimension, the probability of filter failure is derived. According to the method of Lagrange Multiplier, a mathematical approach for calculating optimal filters is provided. And then an approximation algorithm with low computation cost for computing optimal filters is proposed, and the ratio bound of this algorithm is analyzed. The theoretical analysis and performance evaluation demonstrate the accuracy and effectiveness of the proposed algorithm.

**Keywords** wireless sensor networks; monitoring; constraint violation; approximation algorithm; IoT

收稿日期: 2014-07-21; 最终修改稿收到日期: 2015-05-08. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2012CB316200)、国家自然科学基金(61033015, 60933001, 61190115, 61173022)资助. 毕 冉, 女, 1985 年生, 博士研究生, 主要研究方向为无线传感器网络、信息物理融合系统. E-mail: biranhit@163.com. 李建中, 男, 1950 年生, 教授, 博士生导师, 主要研究领域为海量数据管理、数据库、传感器网络、信息物理融合系统等. 高 宏, 女, 1966 年生, 教授, 博士生导师, 主要研究领域为海量数据管理、数据挖掘、传感器网络、信息物理融合系统等.

## 1 引言

集成有传感器, 数据处理单元和通信模块的微小传感器节点以自组织的方式构成了人们广泛熟知的无线传感器网络. 借助于节点中内置的形式多样的感知器件, 传感器节点可以监测包括温度、湿度、噪声、光强度、pH 值、CO<sub>2</sub> 浓度在内的等众多人们感兴趣的物理变量<sup>[1-2]</sup>. 随着嵌入式技术和无线通信技术的不断发展, 传感器网络逐渐走进人们的生活, 在军事、环境、健康、建筑物监测等其他领域得到了广泛地应用.

随着人们对环境科学的日益关注, 传感器网络满足了人们低成本地获取环境科学数据的需求<sup>[3-4]</sup>. 例如 ALERT 系统<sup>①</sup>配置了多种传感器来监测降雨量、河水水位以及土壤水分, 并通过监测数据预测爆发山洪的概率. 具有多种传感器的 FWI<sup>②</sup> 系统用于监测 CO 浓度、CO<sub>2</sub> 浓度、温度和光学烟雾浓度, 并依据监测数据判断森林中是否有火灾发生<sup>[3]</sup>. 鉴于传感器节点的感知数据常常受到硬件误差和环境噪声的影响, 因此在上述基于传感网的连续监测应用中, 用户关心  $t$  时刻监测区域的聚集和(或感知数据的平均值)是否大于(或小于)指定的阈值, 并依此判断监测区域内是否发生异常事件.

由于传感器节点仅以电池为能源供给, 将监测区域内的全部感知数据发送至汇聚(Sink)节点会引起较高的通信开销, 降低了传感网的使用寿命. 因此研究者提出了高能效的监测算法<sup>[5-9]</sup>, 其主要思想是为各节点的感知数据设置局部阈值(或过滤器), 若感知数据小于局部阈值, 那么节点不需向 Sink 节点传输数据. 显然地, 当感知数据大于其局部阈值时, 监测区域内的聚集和(或感知数据的平均值)才有可能大于其指定阈值. 因此, 当节点的感知数据超出局部阈值时, 节点向 Sink 节点发送阈值违反信息. 当 Sink 节点接收到传感器节点的阈值违反信息后, Sink 节点收集监测区域的全部感知数据并计算聚集值. Sink 节点根据聚集值结果判断监测区域内是否发生异常事件.

然而, 已有的研究工作致力于启发式的阈值设置方法, 或者根据启发式的过滤器失败概率值设置局部阈值, 没有考虑感知数据的特性, 忽略了噪声对感知数据的影响, 引起较高的通信开销. 鉴于上述原因, 本文通过一维随机游走来模拟感知数据的变化, 以最小化通信能量开销为优化目标, 开展了过滤器

最优阈值的研究. 本文的主要工作和贡献概括如下:

(1) 提出了通信开销模型, 给出了过滤器最优阈值问题的形式化定义.

(2) 通过一维随机游走来模拟感知数据的变化并推导出过滤器失败概率的计算公式, 根据拉格朗日乘子法给出了求解过滤器最优阈值的数学方法.

(3) 通过分析过滤器最优阈值的解的特性, 提出了低计算复杂度的过滤器最优阈值近似算法, 分析了近似算法的近似比、时间复杂度和通信复杂度.

(4) 通过理论分析和模拟实验, 验证了提出的算法的正确性以及高效性.

本文第 2 节介绍约束违反监测的相关工作; 第 3 节介绍通信开销模型并给出过滤器最优阈值问题的形式化定义; 第 4 节介绍算法的数学基础, 推导过滤器的失败概率, 给出求解过滤器最优阈值的数学方法; 第 5 节提出过滤器最优阈值近似算法, 并对算法的正确性、近似比、时间复杂度和通信复杂度进行分析; 第 6 节通过实验对提出算法的性能进行考察; 最后, 我们总结本文的研究工作, 并对未来的研究进行展望.

## 2 相关工作

无线传感器网络的约束违反监测问题已经引起了人们的广泛关注, 提出了有效的约束违反监测算法<sup>[5-9]</sup>. 针对聚集和的约束违反监测问题, 文献[10]提出了零松弛的阈值设置方法, 当节点的感知值大于  $T/n$  时, 节点向 Sink 节点发送阈值违反信息, Sink 节点收到阈值违反信息后, Sink 节点收集监测区域的感知数据考察其聚集和是否大于阈值  $T$ .

为了监测全局约束违反, 文献[11]研究了基于树的分布式协议. 文献[12]通过考察感知数据的时空关联, 建立了一组概率模型, 使得每个模型对应一个感知节点. 该模型分别在 Sink 节点和传感器节点进行维护. Sink 节点根据维护的概率模型估计感知数据. 基于模型的数据估计方法更适用于数据收集; 在全局约束违反的监测应用中, 该方法的灵活性较低.

由于收集全网感知数据消耗较高的通信能量, 文献[5]以最小化收集全网感知数据的概率为优化目标, 提出了阈值监测的统一框架. 此外, 研究者亦提出了非零松弛的阈值设置方法<sup>[8]</sup>. 以最小化网内

① ALERT. <http://www.altersystem.org>

② Canadian Forest Service (CFS) Web Page. [www.nrcan.gc.ca/cfs](http://www.nrcan.gc.ca/cfs)

通信能量为优化目标,文献[7]提出了启发式的非零松弛阈值设置方法,然而文献[7]缺少关于局部阈值最优性的理论分析,仅给出了通信开销的定性分析. Wuhib 等人<sup>[11]</sup>假设节点的感知数据服从已知的正态分布,基于排队论理论提出了计算局部阈值的算法. Meng 等人<sup>[13]</sup>开发了基于时间窗口的状态监测系统. 然而,上述算法基于的假设条件在传感网中并不适用.

文献[14]研究了通信开销与连续查询结果的准确度的关系,其中文献[14]假设所有节点以聚集方式与中央节点进行通信. 为了跟踪带有误差界的聚集函数结果,文献[15]提出基于自适应过滤器的方法,该方法根据用户的精度要求,给出了局部过滤器阈值. 文献[16]通过累积概率分布函数的 sigmoid 近似将过滤器阈值的计算问题转化为凸优化问题. 文献[17]提出了基于直方图的分布式实时监测算法. 然而,上述方法的优化目标与本文的优化目标最小化通信开销有很大的不同. Tang 等人<sup>[18]</sup>将感知数据视为不确定数据,提出了分布式不确定数据的阈值违反监测算法. 文献[18]研究的分布式不确定数据的监测问题与本文的确定型阈值违反监测问题存在较大的差异.

约束违反监测的一般形式为,对于给定的一般函数  $f(v)$  以及阈值  $T$ ,我们需要确定  $f(v) \leq T$  是否成立,其中  $v$  通常为动态变化的向量. 几何监测方法可以将任意(或非线性)的全局约束分解为局部约束,提供了通用的监测方法论,被广泛地应用于分布式的数据流监测<sup>[19]</sup>. Keren 等人<sup>[19]</sup>考虑了节点上数据分布的差异,以最小化监测向量违反局部安全区域的个数为目标,给出了监测异构数据流的一般方法. Lazerson 等人<sup>[20]</sup>给出了基于凸分解方法的通用性较高的监测框架.

文献[21]研究了近似监测问题,其目标为追踪  $f(v)$  的值是否属于用户预先规定的误差界. 文献[22]提出了基于覆盖球体(Covering Spheres)技术的分布式数据的约束违反监测方法. Covering Spheres (CS)方法的主要思想为将全局的阈值约束分解为“安全”的局部约束条件. Sharfman 等人<sup>[23]</sup>对 CS 方法进行了扩展,提出了椭球体概念并引进了更一般的“安全区域(Safe Zones)”概念. 基于 CS 方法, Burdakis 等人<sup>[24]</sup>设计了无线传感器网络中的离群点检测算法. 根据 CS 技术,文献[25]研究了区间-聚集、连接-聚集等分布式数据流查询的近似监测算法. 基于 CS 思想,文献[26]提出了监测系统健康指标的框架.

鉴于其局部监测目标为多维向量,上述基于几何方法的监测算法具有较好的扩展性. 本文结合传感器节点感知数据的特性,针对更具体的应用场景开展了约束违反监测的近似算法的研究.

### 3 最小化过滤器失败概率

由于感知数据具有时空相关性,因此我们根据当前时刻的感知数据以及历史信息为各节点的感知数据分配最优的过滤区间,使得监测区域内节点的过滤器失败概率之和最小,达到降低通信能量开销、延长网络寿命的目的.

#### 3.1 聚集值的约束违反监测

不妨设监测区域  $A$  分布了  $n$  个传感器节点,其节点 id 分别为  $1, 2, \dots, n$ . 令  $d_i(t)$  表示  $i$  节点在  $t$  时刻的感知值,那么  $t$  时刻监测区域的感知数据集合为  $\{d_1(t), d_2(t), \dots, d_n(t)\}$ . 令  $\tau$  表示监测阈值,  $S(t)$  表示  $t$  时刻监测区域  $A$  的聚集和,即  $S(t) = \sum_{i=1}^n d_i(t)$ . 对于给定的监测区域  $A$  和时间区间  $[t_1, t_m]$ , 如果  $S(t) > \tau$ , 那么 Sink 节点向用户返回  $t$  时刻的聚集结果  $S(t)$  并发送警报,其中  $t \in [t_1, t_m]$ .

对  $\forall t \in [t_1, t_m]$ , 将  $t$  时刻监测区域的全部感知数据  $\{d_1(t), \dots, d_n(t)\}$  发送至 Sink 节点会引起较高的通信开销,不适合以电池为能量供给的传感器网络. 因此对任意  $i \in \{1, 2, \dots, n\}$ , 我们为  $i$  节点的感知数据设置边界阈值  $u_i$ , 并且  $u_1, u_2, \dots, u_n$  满足  $\tau = \sum_{i=1}^n u_i$ . 监测算法的基本思想如下. 易知  $S(t) > \tau$  的必要条件为至少存在一个节点的感知数据  $d_j(t)$  大于其边界阈值  $u_j$ . 因此,如果  $t$  时刻各节点的感知数据小于等于其边界阈值,那么必有  $S(t) \leq \tau$  成立,此时各节点不必向 Sink 节点发送感知数据. 不妨设  $t$  时刻  $j$  节点的感知数据  $d_j(t) > u_j$ , 那么  $t$  时刻的聚集和  $S(t)$  可能大于阈值  $\tau$ , 此时  $j$  节点将其感知数据  $d_j(t)$  发送至 Sink 节点, Sink 节点向未违反边界阈值的感知数据的节点发送探测查询,传感器节点接收到探测查询后将其对应时刻的感知数据发送至 Sink 节点. Sink 节点接收到全部感知数据后,计算  $t$  时刻聚集和  $S(t)$  是否大于监测阈值  $\tau$ .

#### 3.2 通信开销模型

由于感知数据具有时空相关性,因此我们根据给定的监测阈值和历史信息为节点分配优化的过滤区间(Filter),达到节省能量、延长网络寿命的目的.

由于用户指定的监测区域是动态变化的,节点位置不确定,于是与文献[27]提出的通信能量开销模型类似,本文假设传感器节点向 Sink 节点发送一个数据包的平均代价为  $c_u$ , Sink 节点向传感器节点发送一个数据包的平均代价为  $c_d$ ;  $N_u$  记为网内传感器节点向 Sink 节点发送的总的数据包的数量,  $N_d$  记为 Sink 节点向传感器节点发送的总的数据包的数量. 那么,通信能量开销为  $c_u N_u + c_d N_d$ . 由于感知数据违反局部过滤器的节点是动态变化的,难以预测的,于是本文研究的问题为求解各节点的局部过滤器阈值使得通信开销的期望最小.

**引理 1.** 若  $t$  时刻  $i$  节点的感知数据  $d_i(t) > u_i$ , 那么因  $i$  节点的过滤器失效而引起的通信开销为  $n(c_u + c_d) - c_d$ .

证明. 如果  $t$  时刻  $i$  节点的感知数据  $d_i(t) > u_i$ , 那么通信开销由下述 3 部分构成. 首先,因  $i$  节点向 Sink 节点发送数据包引起的通信开销为  $c_u$ ; 其次,因 Sink 节点向其他  $n-1$  个节点发送探测查询的通信开销为  $(n-1)c_d$ ; 第三,传感器节点接收到探测查询后,将其  $t$  时刻的感知数据发送至 Sink 节点引起的通信开销为  $(n-1)c_u$ . 综上所述,因  $i$  节点的过滤器失效而引起的通信开销为  $n(c_u + c_d) - c_d$ . 证毕.

令  $Pr_i$  表示  $i$  节点过滤器失效的概率,那么因  $i$  节点过滤器失效而引起的通信开销的期望为  $Pr_i \{n(c_u + c_d) - c_d\}$ . 因此本文研究的问题是为各节点设置优化的过滤器阈值,使得监测区域内的节点通信开销的期望最小. 由于各节点的过滤器失效是相互独立的,优化的过滤器阈值问题可形式化为如下组合优化问题.

$$\begin{aligned} \min \quad & \sum_{i=1}^n Pr_i \{n(c_u + c_d) - c_d\} \\ \text{s. t.} \quad & \sum_{i=1}^n u_i = \tau \end{aligned} \quad (1)$$

由于  $n(c_u + c_d) - c_d$  是常数,易知优化问题(1)等价于下述优化问题.

$$\begin{aligned} \min \quad & \sum_{i=1}^n Pr_i \\ \text{s. t.} \quad & \sum_{i=1}^n u_i = \tau \end{aligned} \quad (2)$$

## 4 过滤器最优阈值

为了求解各节点感知数据过滤器的最优阈值,

首先,本节推导过滤器的失败概率;其次,通过拉格朗日乘法给出求解过滤器最优阈值的数学方法.

### 4.1 一维随机游走模型

已有的研究工作通过估算感知数据属于过滤器的最短时间来估计过滤器的失败概率<sup>[27]</sup>. 由于传感器节点感知硬件的低敏感性,感知数据经常受到硬件误差和环境噪声的影响. 已有的研究工作忽略了噪声对感知数据的影响. 鉴于上述原因,根据感知数据的历史信息,我们通过一维随机游走来模拟感知数据的变化并推导出过滤器的失败概率. 图 1 给出了以 0 为初始位置,在不同步长和不同正向移动概率情况下的一维随机游走<sup>①</sup>. 图 2 给出了来自真实传感网实测的温度变化曲线. 如图 1 所示一维随机游走曲线即体现了感知温度的噪声扰动,又显示出了温度逐渐升高或降低的变化趋势.

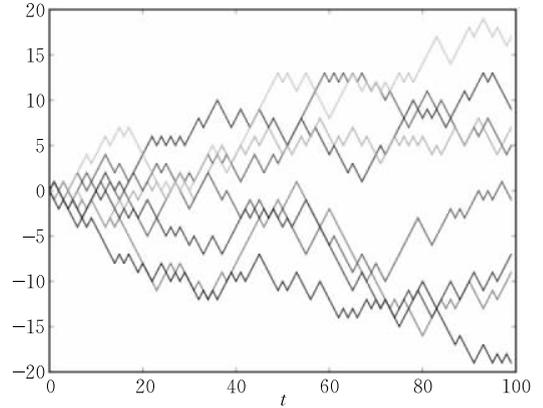


图 1 一维随机游走例子

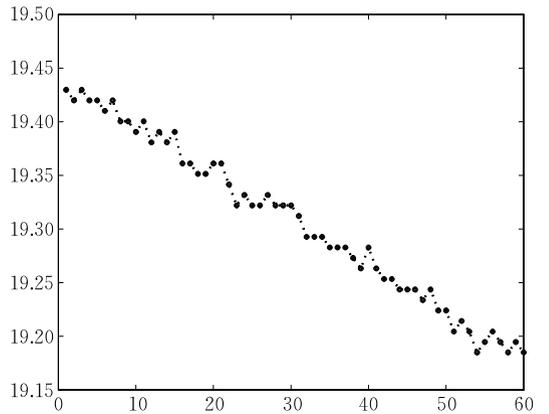


图 2 传感网实测的温度变化曲线

不妨假设当前时刻为  $t_0$ , 传感器节点  $i$  最近一段时间的感知数据为  $\{d_i(t_0), d_i(t_{-1}), \dots, d_i(t_{-m})\}$ , 那

① Random walk. Wikipedia. [http://en.wikipedia.org/wiki/Random\\_walk](http://en.wikipedia.org/wiki/Random_walk)

么对  $\forall h \in \{-1, -2, \dots, -m\}$  有下式成立  $|t_0 - t_h| > |t_0 - t_{h+1}|$ , 即时刻  $t_{h+1}$  比时刻  $t_h$  更接近当前时刻  $t_0$ .

**定义 1.** 感知数据的变化幅值  $\delta_i(t_h)$ .  $i$  节点  $t_h$  时刻感知数据的变化幅值  $\delta_i(t_h) = d_i(t_h) - d_i(t_{h-1})$ .

**定义 2.** 感知数据的平均变化幅度  $a_i$ .  $i$  节点感知数据的平均变化幅度定义为在  $[t_0, t_{-m+1}]$  时间内,

$$a_i = \frac{\sum_{h=-m+1}^0 |\delta_i(t_h)|}{m}$$

变化幅值绝对值的平均值, 即

**定义 3.** 正向移动概率  $Pr_i^+$ .  $i$  节点感知数据正向移动概率定义为在  $[t_0, t_{-m+1}]$  时间内,  $\delta_i(t_h) \geq 0$  的频率,  $Pr_i^+$  的计算公式如下:

$$Pr_i^+ = \frac{|\{k | \delta_i(t_k) > 0, k \in \{0, \dots, -m+1\}\}|}{m} + \frac{1}{2} \frac{|\{k | \delta_i(t_k) = 0, k \in \{0, \dots, -m+1\}\}|}{m}$$

**定义 4.** 负向移动概率  $Pr_i^-$ .  $i$  节点感知数据的负向移动概率定义为在  $[t_0, t_{-m+1}]$  时间内,  $\delta_i(t_h) \leq 0$  的频率,  $Pr_i^-$  的计算公式如下:

$$Pr_i^- = \frac{|\{k | \delta_i(t_k) < 0, k \in \{0, \dots, -m+1\}\}|}{m} + \frac{1}{2} \frac{|\{k | \delta_i(t_k) = 0, k \in \{0, \dots, -m+1\}\}|}{m}$$

由上述讨论可知,  $i$  节点感知数据的变化可以通过以  $d_i(t_0)$  为初始位置的  $[l_i, u_i]$  上的一维随机游走来模拟, 其中  $l_i$  和  $u_i$  为吸收点, 其语义为感知数据小于  $l_i$  或大于  $u_i$  时过滤器失效. 由定义 2 知,  $a_i$  为单位时间的平均移动步长, 令  $x_i = \frac{d_i(t_0) - l_i}{a_i}$ ,  $y_i = \frac{u_i - d_i(t_0)}{a_i}$ , 那么上述一维随机游程等价于质点以  $x_i$  为初始位置, 分别以概率  $Pr_i^+$ 、 $Pr_i^-$  向正、负方向移动单位步长, 以 0 和  $x_i + y_i$  为吸收点的一维随机游过程.

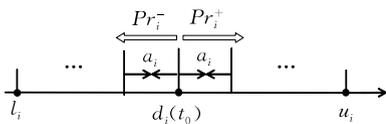


图 3  $i$  节点感知数据过滤器示意图

4.2 推导过滤器失败概率

**引理 2.** 将  $L_0$  记为初始时刻质点所在位置, 将质点到达  $x_i + y_i$  的随机事件记为  $W$ .  $Pr_h$  表示质点初始位置为  $h$  时到达  $x_i + y_i$  的概率, 即  $Pr_h = Pr(W | L_0 = h)$ , 那么有下式成立,

$$Pr_h = \begin{cases} 0, & h=0 \\ 1, & h=x_i+y_i \\ p \times Pr_{h+1} + (1-p) \times Pr_{h-1}, & 0 < h < x_i+y_i \end{cases}$$

其中,  $p$  为质点向正方向移动单位步长的概率.

证明.  $Pr_0 = Pr(W | L_0 = 0) = 0$  比较显然, 若质点的初始位置为 0, 由吸收点的性质知, 质点静止在吸收点, 发生移动的概率为 0, 故  $Pr_0 = 0$ . 若质点的初始位置为  $x_i + y_i$ , 那么质点静止在吸收点  $x_i + y_i$ , 不发生任何移动, 故  $Pr_{x_i+y_i} = 1$ . 当  $0 < h < x_i + y_i$  时, 将第 1 步移动为正方向的随机事件记为  $M^+$ , 将第 1 步移动为负方向的随机事件记为  $M^-$ . 由全概率公式知  $Pr_h = Pr(W \wedge M^+ | L_0 = h) + Pr(W \wedge M^- | L_0 = h)$ . 由条件概率可推出下式成立,

$$\begin{aligned} Pr(W \wedge M^+ | L_0 = h) &= \frac{Pr(W \wedge M^+ \wedge (L_0 = h))}{Pr(L_0 = h)} \\ &= \frac{Pr(W \wedge M^+ \wedge (L_0 = h)) Pr(M^+ \wedge (L_0 = h))}{Pr(M^+ \wedge (L_0 = h)) Pr(L_0 = h)} \\ &= Pr(M^+ | L_0 = h) Pr(W | M^+ \wedge (L_0 = h)) \\ &= p \times Pr_{h+1}. \end{aligned}$$

由于随机游走是无记忆的, 易知  $Pr(W | M^+ \wedge (L_0 = h)) = Pr_{h+1}$ . 同理可推出下式成立  $Pr(W \wedge M^- | L_0 = h) = Pr(M^- | L_0 = h) Pr(W | M^- \wedge (L_0 = h)) = (1-p) \times Pr_{h-1}$ . 因此, 当  $0 < h < x_i + y_i$  时,  $Pr_h = p \times Pr_{h+1} + (1-p) \times Pr_{h-1}$ . 证毕.

定理 1.

$$Pr_{x_i} = \begin{cases} \frac{\gamma^{x_i} - 1}{\gamma^{x_i+y_i} - 1}, \gamma = \frac{1-p}{p}, p \neq \frac{1}{2} \\ \frac{x_i}{x_i+y_i}, p = \frac{1}{2} \end{cases} \quad (3)$$

其中,  $x_i$  为质点的初始位置,  $p$  为质点向正方向移动单位步长的概率.

证明. 由引理 2 知, 对  $\forall m \in \{1, \dots, x_i + y_i - 1\}$ ,  $Pr_m = p \times Pr_{m+1} + (1-p) \times Pr_{m-1}$ , 因此容易推出  $p \times Pr_{m+1} - Pr_m + (1-p) \times Pr_{m-1} = 0$ , 并且其边界条件为  $Pr_0 = 0, Pr_{x_i+y_i} = 1$ . 易知该差分方程的特征方程为  $p \times h^2 - h + (1-p) = 0$ , 其解为  $h_1 = 1, h_2 = \frac{q}{p}$ , 其中  $q = 1-p$ . 令  $\gamma = \frac{q}{p}$ , 那么对  $\forall m \in \{0, 1, \dots, x_i + y_i\}$ , 我们有  $Pr_m = A\gamma^m + B(1)^m$ . 代入边界条件

$$\begin{aligned} Pr_0 = 0 \text{ 和 } Pr_{x_i+y_i} = 1, \text{ 可求解出 } A &= \frac{1}{\gamma^{x_i+y_i} - 1}, B = \\ &= -\frac{1}{\gamma^{x_i+y_i} - 1}. \text{ 于是 } Pr_m = \frac{\gamma^m - 1}{\gamma^{x_i+y_i} - 1}, \text{ 特别地, 当 } m = \\ x_i \text{ 时, } Pr_m &= \frac{\gamma^{x_i} - 1}{\gamma^{x_i+y_i} - 1}. \end{aligned}$$

当  $p = \frac{1}{2}$  时, 可以推出  $Pr_m = \frac{1}{2}Pr_{m+1} + \frac{1}{2}Pr_{m-1}$ ;

其特征方程为  $h^2 - 2h + 1 = 0$ , 易知  $h = 1$  为该特征方程的二重根, 所以对  $\forall m \in \{0, 1, \dots, x_i + y_i\}$ , 我们有  $Pr_m = Am + B(1)^m$ , 代入其边界条件  $Pr_0 = 0$  和

$Pr_{x_i+y_i} = 1$ , 可以推出  $A = \frac{1}{x_i + y_i}$ ,  $B = 0$ . 因此  $Pr_m =$

$\frac{m}{x_i + y_i}$ , 当  $m = x_i$  时,  $Pr_m = \frac{x_i}{x_i + y_i}$ .

综上所述, 质点以  $x_i$  为初始位置,  $p$  为正向移动概率, 到达  $x_i + y_i$  的概率满足式(3). 证毕.

为了方便阅读, 表 1 列出了文中常用符号的说明.

表 1 符号说明

符号	符号说明
$I_N$	节点 id 集合 $\{1, 2, \dots, N\}$
$d_i(t)$	$i$ 节点 $t$ 时刻的感知值
$\delta_i(t_h)$	$i$ 节点 $t_h$ 时刻感知数据的变化幅度
$a_i$	$i$ 节点感知数据的平均变化幅度
$Pr_i^+$	$i$ 节点感知数据的正向移动概率
$Pr_i^-$	$i$ 节点感知数据的负向移动概率
$l$	感知值下界
$u_i$	$i$ 节点感知数据过滤器上界
$F_i$	$i$ 节点感知数据的过滤器 $[l, u_i]$
$\tau$	聚集和的监测阈值
$\Delta$	监测阈值与聚集和之差, $\Delta = \tau - \sum_{i=1}^n d_i(t_0)$

**推论 1.** 不妨设  $l$  为感知值的下界,  $d_i(t_0)$  为  $i$  节点当前时刻的感知值, 若  $[l, u_i]$  为  $i$  节点感知数据的过滤器, 那么过滤器失效的概率满足下式:

$$Pr_i = \begin{cases} \frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i+y_i} - 1}, \gamma_i = \frac{Pr_i^-}{Pr_i^+}, Pr_i^+ \neq \frac{1}{2} \\ \frac{x_i}{x_i + y_i}, Pr_i^+ = \frac{1}{2} \end{cases} \quad (4)$$

其中,  $x_i = \frac{d_i(t_0) - l}{a_i}$ ,  $y_i = \frac{u_i - d_i(t_0)}{a_i}$ .

证明. 不妨设传感器节点正常工作, 由于  $l$  为感知值的下界, 那么感知值必定大于等于  $l$ . 当感知值大于  $u_i$  时, 过滤器失效. 我们根据感知数据的历史信息通过一维随机游走来模拟感知数据的变化, 因此由定理 1 易知, 推论 1 成立. 证毕.

### 4.3 过滤器最优阈值

我们旨在为各节点感知数据的过滤器设置最优的过滤器阈值  $u_i$ , 使得过滤器失效概率之和最小. 不妨设  $t_0$  时刻, 监测区域 A 的感知数据集合为  $\{d_1(t_0), d_2(t_0), \dots, d_n(t_0)\}$ . 由式(4)易知, 对于给定的传感器节点  $i$ ,  $u_i = a_i y_i + d_i(t_0)$ , 则优化问题(2)等价于下述优化问题.

$$\begin{aligned} \min & \sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + y_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + y_{j(k)}} \\ \text{s. t.} & \sum_{i=1}^n a_i y_i = \tau - \sum_{i=1}^n d_i(t_0) \end{aligned} \quad (5)$$

其中, 当  $Pr_i^+ \neq \frac{1}{2}$  时,  $\gamma_i = \frac{Pr_i^-}{Pr_i^+}$ ;  $\{i(1), i(2), \dots, i(h)\}$  与  $\{j(1), \dots, j(g)\}$  互为补集并且  $\{i(1), \dots, i(h)\} \cup \{j(1), \dots, j(g)\} = I_N = \{1, 2, \dots, n\}$ . 在本节, 我们根据拉格朗日乘子法给出了求解过滤器最优阈值的数学方法.

**引理 3.** 若  $y_1, y_2, \dots, y_n$  是下述方程组的解, 则  $y_1, y_2, \dots, y_n$  是优化问题(5)的极值点.

$$\begin{cases} \frac{\ln(\gamma_{i(1)}) \gamma_{i(1)}^{x_{i(1)}} (1 - \gamma_{i(1)}^{x_{i(1)}}) \gamma_{i(1)}^{y_{i(1)}}}{(\gamma_{i(1)}^{x_{i(1)}} \times \gamma_{i(1)}^{y_{i(1)}} - 1)^2} - \lambda a_{i(1)} = 0 \\ \vdots \\ \frac{-x_{j(1)}}{(x_{j(1)} + y_{j(1)})^2} - \lambda a_{j(1)} = 0 \\ \vdots \\ \lambda = \Delta^{-1} \left( \sum_{k=1}^h \frac{\ln(\gamma_{i(k)}) \gamma_{i(k)}^{x_{i(k)}} (1 - \gamma_{i(k)}^{x_{i(k)}}) \gamma_{i(k)}^{y_{i(k)}}}{(\gamma_{i(k)}^{x_{i(k)}} \times \gamma_{i(k)}^{y_{i(k)}} - 1)^2} y_{i(k)} + \sum_{k=1}^g \frac{-x_{j(k)}}{(x_{j(k)} + y_{j(k)})^2} y_{j(k)} \right) \end{cases} \quad (6)$$

证明. 首先, 根据拉格朗日乘子法定义拉格朗日函数  $\Gamma(y_1, y_2, \dots, y_n, \lambda)$ , 那么有下式成立,

$$\Gamma(y_1, y_2, \dots, y_n, \lambda) = \sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + y_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + y_{j(k)}} - \lambda \left( \sum_{i=1}^n a_i y_i - \Delta \right),$$

其中  $\Delta = \tau - \sum_{i=1}^n d_i(t_0)$ .

其次, 分别计算  $\Gamma$  关于  $y_1, y_2, \dots, y_n, \lambda$  的偏导数, 易知下式成立.

$$\frac{\partial \Gamma}{\partial y_{i(k)}} = \frac{\ln(\gamma_{i(k)}) \gamma_{i(k)}^{x_{i(k)}} (1 - \gamma_{i(k)}^{x_{i(k}}) \gamma_{i(k)}^{y_{i(k)}})}{(\gamma_{i(k)}^{x_{i(k)}} \times \gamma_{i(k)}^{y_{i(k)}} - 1)^2} - \lambda a_{i(k)}, \quad k \in \{1, \dots, h\},$$

$$\frac{\partial \Gamma}{\partial y_{j(k)}} = \frac{-x_{j(k)}}{(x_{j(k)} + y_{j(k)})^2} - \lambda a_{j(k)}, \quad k \in \{1, \dots, g\},$$

$$\frac{\partial \Gamma}{\partial \lambda} = \Delta - \sum_{i=1}^n a_i y_i.$$

由拉格朗日乘子法可知, 若  $y_1, y_2, \dots, y_n$  是下述方程组的解, 那么  $y_1, y_2, \dots, y_n$  是优化问题(5)的极值点.

$$\left\{ \begin{array}{l} \frac{\ln(\gamma_{i(1)})\gamma_{i(1)}^{x_{i(1)}}(1-\gamma_{i(1)}^{x_{i(1)}})\gamma_{i(1)}^{y_{i(1)}}}{(\gamma_{i(1)}^{x_{i(1)}} \times \gamma_{i(1)}^{y_{i(1)}} - 1)^2} - \lambda a_{i(1)} = 0 \\ \vdots \\ \frac{\ln(\gamma_{i(h)})\gamma_{i(h)}^{x_{i(h)}}(1-\gamma_{i(h)}^{x_{i(h)}})\gamma_{i(h)}^{y_{i(h)}}}{(\gamma_{i(h)}^{x_{i(h)}} \times \gamma_{i(h)}^{y_{i(h)}} - 1)^2} - \lambda a_{i(h)} = 0 \\ \vdots \\ \frac{-x_{j(1)}}{(x_{j(1)} + y_{j(1)})^2} - \lambda a_{j(1)} = 0 \\ \vdots \\ \frac{-x_{j(g)}}{(x_{j(g)} + y_{j(g)})^2} - \lambda a_{j(g)} = 0 \\ \Delta - \sum_{i=1}^n a_i y_i = 0 \end{array} \right. \quad (7)$$

由  $\sum_{i=1}^n a_i y_i = \Delta$ , 可知  $\sum_{i=1}^n \lambda a_i y_i = \lambda \Delta$ , 于是  $\lambda =$

$\frac{\sum_{i=1}^n \lambda a_i y_i}{\Delta}$ . 因此容易推出, 若  $y_1, y_2, \dots, y_n$  是方程组(6)的解, 那么  $y_1, y_2, \dots, y_n$  是优化问题(5)的极值点. 证毕.

**引理 4.** 多变量函数  $\Gamma(y_1, y_2, \dots, y_n)$  的海森矩阵是正定的.

证明. 由引理 3 的证明过程易知, 对任意  $i \neq j$ , 可知  $\frac{\partial^2 \Gamma}{\partial y_i \partial y_j} = 0$ . 当  $\gamma \neq 1$  时, 可推出下式成立,

$$\frac{\partial^2 \Gamma}{\partial y_i^2} = \frac{\ln^2(\gamma_i) \gamma_i^{x_i} \gamma_i^{y_i} (1 - \gamma_i^{x_i}) (1 - \gamma_i^{2x_i} \gamma_i^{2y_i})}{(\gamma_i^{x_i} \gamma_i^{y_i} - 1)^4}.$$

如果  $\gamma_i > 1$ , 易知  $(1 - \gamma_i^{x_i}) (1 - \gamma_i^{2x_i} \gamma_i^{2y_i}) > 0$ ; 类似地, 如果  $0 < \gamma_i < 1$ , 易知  $(1 - \gamma_i^{x_i}) (1 - \gamma_i^{2x_i} \gamma_i^{2y_i}) > 0$ ; 于是可知  $\frac{\partial^2 \Gamma}{\partial y_i^2} > 0$ .

当  $\gamma_i = 1$  时, 易知  $\frac{\partial^2 \Gamma}{\partial y_i^2} = \frac{2x_i}{(x_i + y_i)^3}$ ; 那么我们容易推出  $\frac{\partial^2 \Gamma}{\partial y_i^2} > 0$ . 因此,  $\Gamma(y_1, y_2, \dots, y_n)$  关于变量

$y_1, y_2, \dots, y_n$  的海森矩阵  $\frac{\partial^2 \Gamma}{\partial y_i \partial y_j}$  为对角元素皆为正数的对角矩阵. 于是我们可知, 海森矩阵  $\frac{\partial^2 \Gamma}{\partial y_i \partial y_j}$  的特征值都为正数, 所以多变量函数  $\Gamma(y_1, y_2, \dots, y_n)$  的海森矩阵是正定的. 证毕.

**定理 2.** 若  $y_1, y_2, \dots, y_n$  是方程组(6)的解, 那么  $y_1, y_2, \dots, y_n$  是优化问题(5)的最优解.

证明. 由引理 3 可知, 如果  $y_1, y_2, \dots, y_n$  是方程组(6)的解, 那么  $y_1, y_2, \dots, y_n$  是优化问题(5)的极值点; 由引理 4 知目标函数  $\Gamma(y_1, y_2, \dots, y_n)$  的海森矩阵是正定的, 因此函数  $\Gamma$  在极值点取得最小

值. 综上可知  $y_1, y_2, \dots, y_n$  使得优化问题(5)的目标函数取得最小值, 因此  $y_1, y_2, \dots, y_n$  是优化问题(5)的最优解. 证毕.

## 5 求解过滤器最优阈值的近似算法

第 4 节通过拉格朗日乘子法给出了求解过滤器最优阈值的数学方法并证明了方法的正确性. 由定理 2 可知, 为监测区域内的各传感器节点计算其感知数据过滤器的最优阈值等价于求解非线性方程组(6)的解. 我们注意到不易给出非线性方程组(6)的解的一般表达式, 而且求解非线性方程组的近似解的数值分析方法计算开销较高, 且不易分析其方法的收敛性. 鉴于上述原因, 我们分析了非线性方程组(6)的解的特性, 提出了低计算复杂度的近似算法并分析了该算法的近似比.

### 5.1 可行解的存在性

在本节, 我们给出了构造优化问题(5)的可行解的数学方法, 证明了构造解的存在性和可行性.

**引理 5.** 非线性方程组(6)的系数  $\lambda < 0$ .

证明. 对  $\forall i \in \{1, 2, \dots, n\}$ , 如果  $\gamma_i \neq 1$  我们根据方程组(6)容易推出  $\lambda = \frac{\gamma_i^{x_i} \ln(\gamma_i) (1 - \gamma_i^{x_i}) \gamma_i^{y_i}}{a_i (\gamma_i^{x_i} \times \gamma_i^{y_i} - 1)^2}$ . 令

$\kappa_i = \ln(\gamma_i) (1 - \gamma_i^{x_i})$ , 当  $\gamma_i > 1$  时, 容易验证  $\kappa_i < 0$ ; 类似地, 易验证当  $0 < \gamma_i < 1$  时,  $\kappa_i < 0$ . 由定义可知  $a_i > 0$ ,  $\gamma_i > 0$ , 因此我们容易推出  $\lambda < 0$ . 如果  $\gamma_i = 1$ , 根据方程组(6), 可推出  $\lambda = \frac{-x_i}{a_i (x_i + y_i)^2}$ , 由定义可知  $a_i > 0$ ,  $x_i > 0$ , 因此容易推出  $\lambda < 0$ . 证毕.

**引理 6.** 对于给定的  $\lambda < 0$ ,  $x_i > 0$ ,  $a_i > 0$ ,  $\Delta_i$  如下定义:

$$\Delta_i = \gamma_i^{2x_i} (4\lambda a_i \ln \gamma_i (1 - \gamma_i^{x_i}) + \ln^2 \gamma_i (1 - \gamma_i^{x_i})^2) \quad (8)$$

那么, 对  $\forall \gamma_i > 0$ , 如果  $\gamma_i \neq 1$ , 可以推出  $\Delta_i > 0$ .

证明. 对  $\forall \gamma_i > 0$  且  $\gamma_i \neq 1$ , 易验证  $\ln(\gamma_i) (1 - \gamma_i^{x_i}) < 0$ , 由于  $\lambda$  为给定的负数, 于是可知  $4\lambda a_i \ln \gamma_i (1 - \gamma_i^{x_i}) > 0$ . 因此, 对于给定的  $\lambda < 0$ ,  $x_i > 0$ , 对  $\forall \gamma_i > 0$  且  $\gamma_i \neq 1$ , 有  $\Delta_i > 0$  成立. 证毕.

**引理 7.** 对于给定的  $\lambda < 0$ ,  $x_i > 0$ ,  $Y_i$  如下定义:

$$Y_i = \frac{(2\lambda a_i \gamma_i^{x_i} + \gamma_i^{x_i} \ln \gamma_i (1 - \gamma_i^{x_i})) - \sqrt{\Delta_i}}{2\lambda a_i \gamma_i^{2x_i}} \quad (9)$$

那么, 对  $\forall \gamma_i > 0$ , 如果  $\gamma_i \neq 1$ , 可以推出  $Y_i$  是下述方程的一个根, 并且  $Y_i > 0$ ,

$$\frac{\ln(\gamma_i) \gamma_i^{x_i} (1 - \gamma_i^{x_i}) Y_i}{(\gamma_i^{x_i} \times Y_i - 1)^2} - \lambda a_i = 0 \quad (10)$$

其中  $\Delta_i$  按式(8)定义.

证明. 求解方程(10)的根等价于求解下述方程的根.

$$\lambda a_i (\gamma_i^{x_i} \times Y_i - 1)^2 - \ln(\gamma_i) \gamma_i^{x_i} (1 - \gamma_i^{x_i}) Y_i = 0 \quad (11)$$

经化简, 易知方程(11)与下述一元二次方程等价.

$$\lambda a_i \gamma_i^{2x_i} Y_i^2 + (-2\lambda a_i \gamma_i^{x_i} - \ln(\gamma_i) \gamma_i^{x_i} (1 - \gamma_i^{x_i})) Y_i + \lambda a_i = 0 \quad (12)$$

由引理 6 可知  $\Delta_i > 0$ , 所以式(9)是有意义的. 若  $Y_i$  满足式(9), 容易验证  $Y_i$  是方程的根. 由于  $\lambda < 0$ ,  $a_i > 0$ , 所以  $2\lambda a_i \gamma_i^{x_i} < 0$ ; 由于  $x_i > 0$ , 那么对  $\forall \gamma_i > 0$  且  $\gamma_i \neq 1$ , 可知  $\gamma_i^{x_i} \ln \gamma_i (1 - \gamma_i^{x_i}) < 0$ . 因此式(9)的分子为负数. 类似地, 易知式(9)的分母为负数, 综上所述,  $Y_i > 0$ . 证毕.

**定理 3.** 对于给定的  $\lambda < 0$ ,  $x_i > 0$ ,  $a_i > 0$ , 对  $\forall \gamma_i > 0$ ,  $\gamma_i \neq 1$ ,  $y_i$  定义如下:

$$y_i = \ln Y_i / \ln \gamma_i \quad (13)$$

其中,  $Y_i$  按式(9)定义, 那么  $y_i$  为下述方程的解.

$$\frac{\ln(\gamma_i) \gamma_i^{x_i} (1 - \gamma_i^{x_i}) \gamma_i^{y_i}}{(\gamma_i^{x_i} \times \gamma_i^{y_i} - 1)^2} - \lambda a_i = 0 \quad (14)$$

证明. 由引理 5 至引理 7 易知, 定理 3 结论成立. 证毕.

**定理 4.** 对于给定的  $\lambda < 0$ ,  $a_i > 0$  以及  $x_i > 0$ , 令

$$y_i = \sqrt{-x_i (\lambda a_i)^{-1}} - x_i \quad (15)$$

那么  $y_i$  是下述方程的一个根

$$\frac{-x_i}{(x_i + y_i)^2} - \lambda a_i = 0 \quad (16)$$

证明. 由方程(16)易知  $(x_i + y_i)^2 = (-x_i) (\lambda a_i)^{-1}$ , 由于  $\lambda < 0$ ,  $a_i > 0$ ,  $x_i > 0$ , 因此容易验证  $y_i = \sqrt{-x_i (\lambda a_i)^{-1}} - x_i$  是一元二次方程(16)的一个根. 证毕.

## 5.2 近似算法

由上述分析可知, 为监测区域内的各传感器节点的感知数据求解过滤器的最优阈值可形式化为下面的问题.

输入:

1.  $t_0$  时刻监测区域内的感知数据集,  $\{d_1(t_0), \dots, d_n(t_0)\}$ ;
2. 感知数据的历史信息  $\{\gamma_1, \dots, \gamma_n\}, \{a_1, \dots, a_n\}$ ;
3. 感知数据下界  $l$ , 监测阈值  $\tau$ .

输出:

过滤器最优阈值  $\{u_1, u_2, \dots, u_n\}$ .

由定理 2 可知, 计算过滤器的最优阈值等价于求解非线性方程组的精确解, 由于不易求解非线性方程组(6)的精确解, 本节提出计算最优阈值的近似

算法. 算法的基本思想如下.

不妨假设  $z$  为给定的整数并且  $a_\pi = \min\{a_1, a_2, \dots, a_n\}$ .

首先, 令  $y'_\pi = 1/z$ , 如果  $\gamma_\pi \neq 1$ , 将  $y'_\pi = 1/z$  代入式(14)中, 计算  $\lambda = \frac{\ln(\gamma_\pi) \gamma_\pi^{x_\pi} (1 - \gamma_\pi^{x_\pi}) \gamma_\pi^{y'_\pi}}{a_\pi (\gamma_\pi^{x_\pi} \times \gamma_\pi^{y'_\pi} - 1)^2}$ ; 如果  $\gamma_\pi = 1$ ,

将  $y'_\pi = 1/z$  代入式(16)中, 计算  $\lambda = \frac{-x_\pi}{a_\pi (x_\pi + y'_\pi)^2}$ , 其中  $x_\pi = (d_\pi(t_0) - l) / a_\pi$ .

其次, 对任意  $i \neq \pi$ , 如果  $\gamma_i \neq 1$ , 将  $\lambda$  代入式(13)求解  $y'_i$ ; 如果  $\gamma_i = 1$ , 将  $\lambda$  代入式(15)求解  $y'_i$ .

第三, 令  $\Delta' = \sum_{i=1}^n a_i y'_i$ ,  $\rho = \Delta / \Delta'$ , 对任意  $i \in \{1, 2, \dots, n\}$ , 令  $y_i = \rho y'_i$  并计算优化问题(5)的目标函数值, 即过滤器失效概率之和.

令  $y'_\pi = y'_\pi + 1/z$ , 重复上述步骤直到  $a_\pi y'_\pi \geq \Delta$ . 不妨设在上述计算过程中,  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n$  使得目标函数取得最小值, 那么对任意  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点感知数据过滤器的阈值  $u_i = d_i(t_0) + a_i \bar{y}_i$ .

## 5.3 算法性能分析

**算法正确性分析:** 过滤器最优阈值近似算法给出的近似解  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n$  是优化问题(5)的可行解. 由算法步骤易知, 对任意  $i \in \{1, 2, \dots, n\}$ , 我们对  $y'_i$  进行了线性变换, 即  $y_i = \rho y'_i$ . 因此可知  $\sum_{i=1}^n a_i y_i = \rho \sum_{i=1}^n a_i y'_i = \Delta$ , 即  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n$  是优化问题(5)的可行解.

近似比分析:

**定理 5.** 过滤器最优阈值近似算法的近似比上界为  $\alpha$ , 其中  $\alpha$  满足如下公式:

$$\alpha = \max \left[ \max_{\gamma_i \neq 1} \left( \frac{\gamma_i^{x_i + u_i/a_i} - 1}{\gamma_i^{x_i + \bar{y}_i} - 1} \right), \max_{\gamma_i = 1} \left( \frac{x_i + u_i/a_i}{x_i + \bar{y}_i} \right) \right] \quad (17)$$

证明. 不妨设  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$  为优化问题(5)的最优解,

由于优化目标为最小化过滤器失效概率之和, 那么必定存在  $i \in \{1, 2, \dots, n\}$ , 使得  $\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} \leq$

$\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \bar{y}_i} - 1}$  或  $\frac{x_i}{x_i + \hat{y}_i} \leq \frac{x_i}{x_i + \bar{y}_i}$  成立; 否则通过反证法

易证明  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$  不是最优解.

不妨设  $\gamma_i \neq 1$ , 那么容易推出  $\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \bar{y}_i} - 1}$

$$\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} = \frac{\gamma_i^{x_i + \hat{y}_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} \leq \frac{\gamma_i^{x_i + u_i/a_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1}. \text{ 类似地, 若 } \gamma_i =$$

$$1, \text{ 则 } \frac{x_i}{x_i + \hat{y}_i} / \frac{x_i}{x_i + \hat{y}_i} = \frac{x_i + \hat{y}_i}{x_i + \hat{y}_i} \leq \frac{x_i + u_i/a_i}{x_i + \hat{y}_i}. \text{ 令}$$

$$\alpha = \max \left( \max_{\gamma_i \neq 1} \left\{ \frac{\gamma_i^{x_i + u_i/a_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} \right\}, \max_{\gamma_i = 1} \left\{ \frac{x_i + u_i/a_i}{x_i + \hat{y}_i} \right\} \right),$$

那么对任意  $i \in \{1, 2, \dots, n\}$ , 如果  $\gamma_i \neq 1$ , 可以推出

$$\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} / \frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + \hat{y}_i} - 1} \leq \alpha; \text{ 如果 } \gamma_i = 1, \text{ 可以推出}$$

$$\frac{x_i}{x_i + \hat{y}_i} / \frac{x_i}{x_i + \hat{y}_i} \leq \alpha. \text{ 优化问题(5)的目标函数定义为}$$

$$T(y_1, y_2, \dots, y_n) = \sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + y_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + y_{j(k)}},$$

那么我们容易推出下式成立:  $\frac{T(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)}{T(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)} \leq \alpha$ ,

即  $\alpha$  为过滤器最优阈值近似算法的近似比上界.

证毕.

**算法 1.** 过滤器最优阈值近似算法 AACOF (Approximation Algorithm for Computing Optimal Filter).

输入:  $\tau, l, z, \Delta = \infty, \{d_i(t_0), \dots, d_n(t_0)\}, \{\gamma_1, \dots, \gamma_n\}, \{a_1, \dots, a_n\}$

输出:  $u_1, u_2, \dots, u_n$

1. For ( $i=1$  to  $n$ )

2.  $x_i = (d_i(t_0) - l)/a_i$ ;

3. End For

4.  $\Delta = \tau - \sum_{i=1}^n d_i(t_0)$ ,  $\pi = \arg \min \{a_1, \dots, a_n\}$ ,  $y'_\pi = 0$ ;

5. While ( $a_\pi y'_\pi \leq \Delta$ )

6. If  $\gamma_\pi \neq 1$

$$7. \quad \lambda = \frac{\ln(\gamma_\pi) \gamma_\pi^{x_\pi} (1 - \gamma_\pi^{x_\pi}) \gamma_\pi^{y'_\pi}}{a_\pi (\gamma_\pi^{x_\pi} \times \gamma_\pi^{y'_\pi} - 1)^2};$$

8. Else

$$9. \quad \lambda = \frac{-x_\pi}{a_\pi (x_\pi + y'_\pi)^2};$$

10. End If

11. For ( $i=1$  to  $n \wedge i \neq \pi$ )

12. If  $\gamma_i \neq 1$

$$13. \quad y'_i = \ln Y_i / \ln \gamma_i;$$

14. Else

$$15. \quad y'_i = \sqrt{-x_i (\lambda a_i)^{-1}} - x_i;$$

16. End If

17. End For

$$18. \quad \Delta' = \sum_{i=1}^n a_i y'_i, \quad \rho = \Delta / \Delta';$$

19. For ( $i=1$  to  $n$ )

20.  $y_i = \rho y'_i$ ;

21. End For

$$22. \quad \text{If } \sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + y_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + y_{j(k)}} < \Delta$$

23.  $\hat{Y} = \{y_1, y_2, \dots, y_n\}$ ,

$$24. \quad \Lambda = \sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + y_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + y_{j(k)}};$$

25. End If

26.  $y'_\pi = y'_\pi + 1/z$ ;

27. End While

28. For ( $y_i \in \hat{Y}$ )

29.  $u_i = d_i(t_0) + a_i y_i$ ;

30. End For

31. Return  $u_1, u_2, \dots, u_n$ ;

算法时间复杂度: 如算法伪代码所示, 执行第 6~26 行计算步骤的时间复杂度为  $O(n)$ , 而且容易得知上述运算执行了  $\Delta/a_\pi$  次, 因此算法的计算开销为  $O(n\Delta/a_\pi)$ .

算法通信复杂度: 对于给定的监测区域 A, 不妨假设传感器节点发送数据包到 Sink 节点的平均跳数为  $h$ , 那么节点将感知数据  $d_i(t_0)$  以及感知数据的历史信息  $\gamma_i, a_i$  发送至 Sink 节点的通信开销为  $O(nh)$ ; Sink 节点将阈值发送至所有传感器节点的通信开销为  $O(nh)$ , 因此算法的通信开销为  $O(nh)$ .

## 6 实验结果

为了衡量过滤器最优阈值近似算法(AACOF)的性能, 在本节我们通过模拟实验对算法性能进行了测试. 我们分别在合成数据集和真实感知数据集上对算法性能进行了考察. 模拟实验中的温度感知数据来源于 Intel Berkeley 实验室的真实传感器网络实测. 所有实验均在主频为 2.5GHz 的 Intel Core i5 CPU 和内存为 4GB 的 PC 机上完成. 我们使用的操作系统为 Windows 7, 我们用 Matlab 语言在 MATLAB R2010a 环境下实现了 AACOF 算法.

(1) 第 1 组实验考察不同的监测阈值以及不同的监测节点数量对过滤器的平均失败概率的影响, 计算过滤器的平均失败概率的公式如下,

$$Ave = \frac{\sum_{k=1}^h \frac{\gamma_{i(k)}^{x_{i(k)}} - 1}{\gamma_{i(k)}^{x_{i(k)} + \hat{y}_{i(k)}} - 1} + \sum_{k=1}^g \frac{x_{j(k)}}{x_{j(k)} + \hat{y}_{j(k)}}}{n},$$

其中  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$  为近似算法的输出结果,  $n$  为监测节点的数量.

图 4 和图 5 显示了近似算法在合成数据集上的运行结果,其中各节点的数据集以随机游走的方式生成.“○”表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走,其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布;其移动步长服从正态分布  $N(0.5, 1)$ .“◇”表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走,其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布;其移动步长服从正态分布  $N(1, 1)$ . 我们在上述随机生成的数据集上计算过滤器的平均失败概率. 对于给定的监测阈值与聚集和之差  $\Delta$ 、以及节点数量  $n$ ,我们运行上述算法 500 次,并将其平均失败概率的均值作为实验结果. 从图中我们可知,近似算法输出的过滤器阈值使得平均失败概率较小. 如图 4 所示,当监测阈值与聚集和之差  $\Delta=14$ , 监测节点数量  $n=10$ , 移动步长服从正态分布  $N(0.5, 1)$  时,过滤器的平均失败概率小于 0.2.

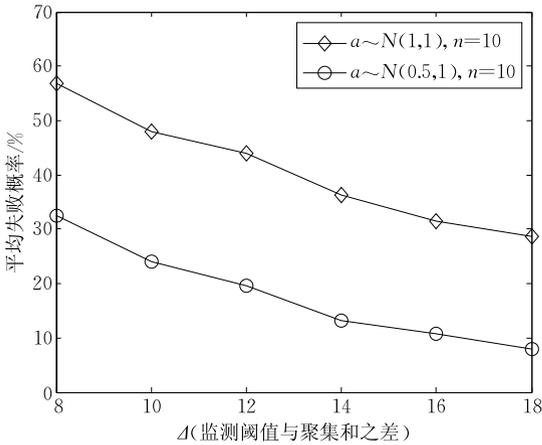


图 4 Δ 不同时过滤器失效概率平均值

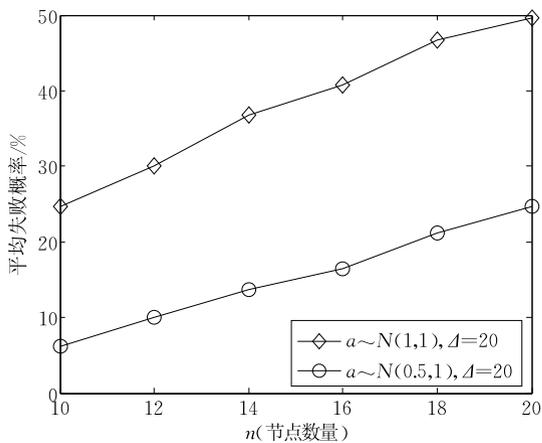


图 5 n 不同时过滤器失效概率平均值

图 6 显示了近似算法在真实传感网实测的温度感知数据集上的运行结果. 我们分别考察了在温度

变化剧烈和变化缓慢情况下的平均失败概率结果,其中变化剧烈的感知数据采用上午 10 点~12 点的温度感知数据;变化缓慢的感知数据采用凌晨 4 点~6 点的温度感知数据. 如图 6 所示,真实数据集上的平均失败概率略高于合成数据集上的平均失败概率,其原因分析如下. 首先,感知数据选取于温度随时间逐渐升高的过程,因此对任意  $i \in \{1, 2, \dots, n\}$ ,  $\gamma_i < 1$ ;其次,对于给定的  $i$ ,当  $y_i \rightarrow \infty$  时,  $\frac{\gamma_i^{x_i} - 1}{\gamma_i^{x_i + y_i} - 1}$  趋近于  $1 - \gamma_i^{x_i}$ ,其中  $x_i = (d_i(t_0) - l) / a_i$ ,因此失败概率取决于  $\gamma_i$  以及  $d_i(t_0) - l$ ,并随  $\Delta$  的增大逐渐趋近其极限,变化不明显.

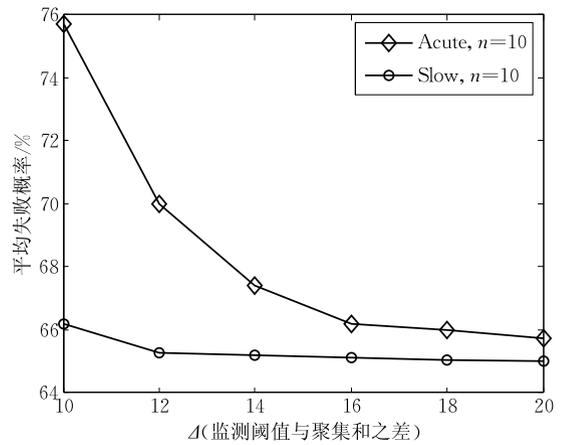


图 6 Δ 不同时过滤器失效概率平均值

(2) 第 2 组实验考察不同的  $\Delta$  以及不同的监测节点数量对近似算法的通信开销的影响,其中通信开销百分比的计算公式如下:

$$\text{通信开销}\% = \frac{\text{\#数据包数量(近似算法)}}{\text{\#数据包数量(基本算法)}} \times 100\%$$

基本算法的思想如下:监测区域内的所有节点将全部感知数据发送至 Sink 节点. 由通信开销模型,我们可知近似算法的通信开销由 2 部分组成:①因监测节点过滤器失效引起的数据包传输;②因 Sink 节点收集监测区域的感知数据引起的数据包传输.

图 7 和图 8 显示了近似算法在合成数据集上的运行结果,其中各节点的数据集以随机游走的方式生成.“○”表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走,其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布;其移动步长服从正态分布  $N(0.5, 1)$ .“◇”表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走,其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布;其移动步长

服从正态分布  $N(1,1)$ . 我们在上述随机生成的数据集上考察近似算法的通信开销百分比. 对于给定的监测阈值与聚集和之差  $\Delta$ 、以及节点数量  $n$ , 我们运行上述算法 500 次, 并将其通信开销百分比的平均值作为实验结果. 从图 7 和图 8 中不难看出近似算法仅需较低的通信能量开销就能完成对监测区域内的聚集和约束违反的监测. 例如, 如图 7 所示, 当  $\Delta=20$ , 监测节点数量  $n=10$ , 移动步长服从正态分布  $N(0.5,1)$  时, 通信开销百分比为 32%. 根据图 8 给出的实验结果, 我们可知随着节点数量的增大, 近似算法所需的通信开销比逐渐降低.

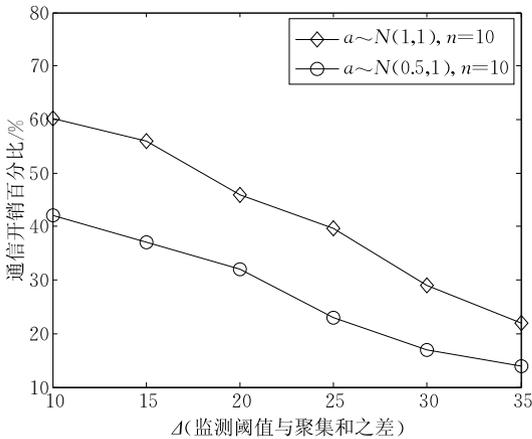


图 7  $\Delta$  不同时通信开销百分比

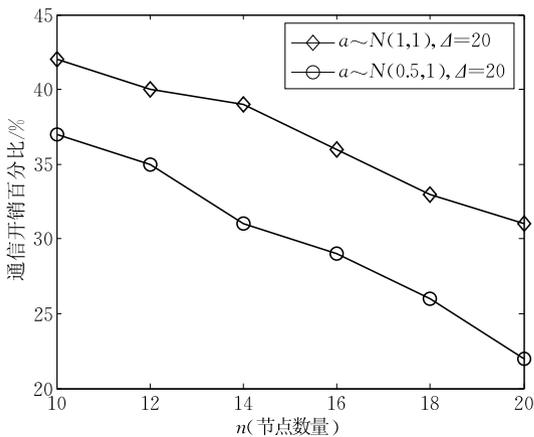


图 8  $n$  不同时通信开销百分比

图 9 显示了近似算法在真实传感网实测的温度感知数据集上的运行结果. 类似地, 我们分别考察了在温度感知值变化剧烈和变化缓慢情况下的通信开销百分比. 如图 9 所示, 真实数据集上的通信开销百分比明显低于合成数据集上的通信开销百分比, 且随节点数量的增加变化不大, 其原因分析如下. 监测节点的分布具有地域性, 温度感知值差异不大; 过滤器失效通常是因聚集和大于约束阈值引起的. 第 2

组实验说明近似算法具有较高的能效性, 降低了传感网的通信开销, 延长了网络寿命.

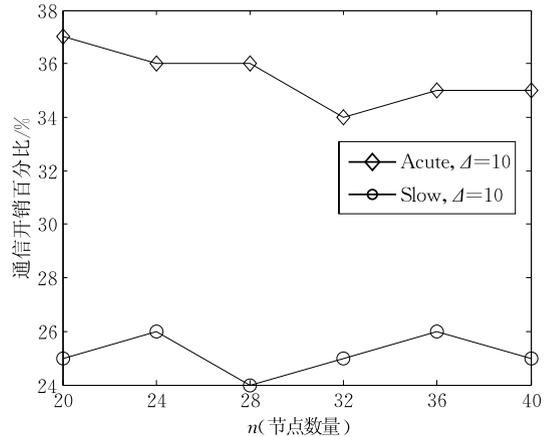


图 9  $n$  不同时通信开销百分比

(3) 第 3 组实验考察不同的  $\Delta$  以及不同的监测节点数量对近似算法的约束违反的准确率的影响, 其中, 计算约束违反的准确率的公式如下:

$$\text{准确率} = \frac{\text{聚集值大于监测阈值时刻的数量}}{\text{过滤器失效的时刻的数量}} \times 100\%.$$

准确率考察了当过滤器失效时, 监测区域内的聚集和大于指定阈值的概率. 图 10 和图 11 显示了近似算法在合成数据集上的运行结果, 其中各节点的数据集以随机游走的方式生成. 类似地, “○” 表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走, 其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布; 其移动步长服从正态分布  $N(0.5, 1)$ ; “◇” 表示对于给定的节点  $i \in \{1, 2, \dots, n\}$ ,  $i$  节点以 1 为初始位置在数轴上进行随机游走, 其中  $i$  节点的正向移动概率  $Pr_i^+$  服从  $[0, 1]$  上的均匀分布; 其移动步长服从正态分布  $N(1, 1)$ . 我们在上述随机生成的数据集上计算约束违反的准确率. 对于给定的监测阈值与聚集和之差  $\Delta$ 、以及节点数量  $n$ , 我们运行上述算法 500 次, 并将其准确率的平均值作为实验结果. 从图 10 和图 11 我们可以观察到, 近似算法具有较高的约束违反准确率, 即当节点向 Sink 节点发送过滤器失效的消息时, 用户可知监测区域的聚集和极可能大于其指定的阈值. 例如, 如图 11 所示, 当  $\Delta=20$ , 监测节点数量  $n=18$ , 移动步长服从正态分布  $N(0.5, 1)$  时, 准确率为 82%.

图 12 显示了近似算法在真实传感网实测的温度感知数据集上的运行结果. 我们分别考察了在温度感知值变化剧烈和变化缓慢情况下的准确率. 如图 12 所示, 随节点数量的增加, 约束违反的准确率

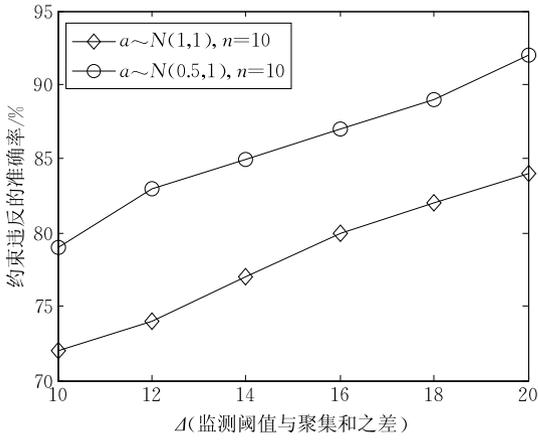


图 10 Δ 不同时约束违反的准确率

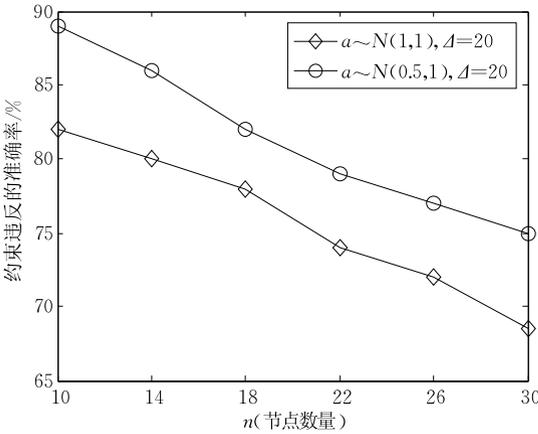


图 11 n 不同时约束违反的准确率

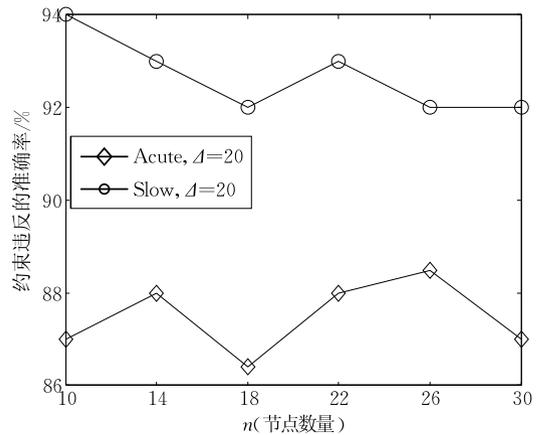


图 12 n 不同时约束违反的准确率

变化不大.其原因分析如下.首先,监测节点的感知数据具有地域相关性,温度感知值差异性较低.其次,感知数据来源于连续变化的物理过程,因此当某(几)个节点的感知数据的过滤器失效时,意味着监测区域的聚集和大于指定的阈值.

(4)为了衡量过滤器最优阈值近似算法的性能,我们将其与零松弛算法<sup>[7]</sup>和自适应的非零松弛算法<sup>[10]</sup>在真实传感网实测的温度感知数据集上进行

了对比实验.首先,我们考察了在感知温度变化缓慢情况下的平均失败概率结果.如图 13 所示,AACOF 代表本文提出的算法; $T/n$  和 Adaptive 分别代表零松弛方法和非零松弛方法.与传统方法相比,根据本文提出的近似算法设置的阈值使得过滤器具有较低的失败概率.由于非零松弛方法的监测阈值为  $\tau - \delta$ ,其中  $\delta$  为给定的非零松弛值,因此该方法具有较高的过滤器失败概率.

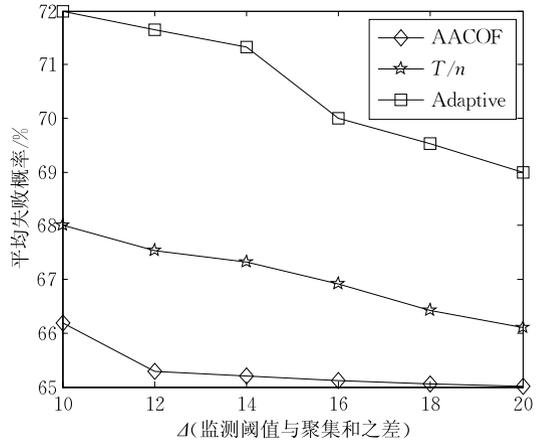


图 13 Δ 不同时过滤器失效概率平均值

其次,在感知温度变化剧烈情况下,我们对各算法引起的通信开销进行了对比实验.如图 14 所示,由于一维随机游走过程可以较好地模拟感知数据的变化,因此 AACOF 算法引起的通信开销较低.在非零松弛算法中,Sink 节点维护的“松弛” $\delta$  虽然增大了局部过滤器失败的概率,但是降低了收集监测区域内的全部感知数据的概率.由于收集全部感知数据的通信开销较高,因此,非零松弛方法引起的通信开销低于零松弛方法.

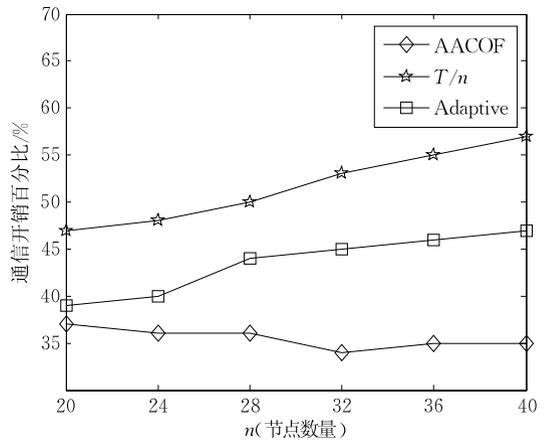


图 14 n 不同时通信开销百分比

第三,在感知温度变化缓慢情况下,我们对各算法的约束违反的准确率进行了对比实验.如图 15 所

示,与已有方法相比,AACOF 在约束违反的准确率方面显示了较好的性能.在非零松弛算法中,Sink 节点维护的“松弛” $\delta$  降低了局部过滤器的阈值上界,导致较高的过滤器失败概率.因此,零松弛方法的约束违反准确率高于非零松弛方法.

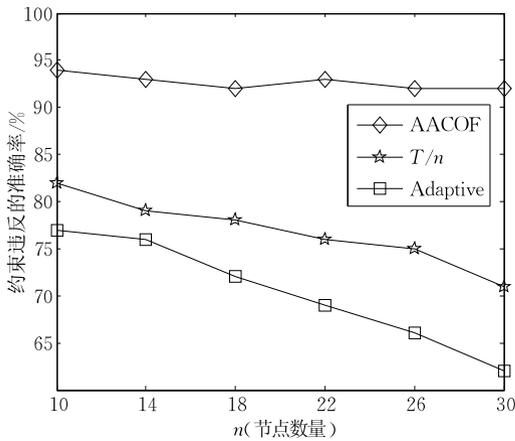


图 15  $n$  不同时约束违反的准确率

## 6 结 论

研究无线传感网的高能效的监测算法具有重要意义.本文以最小化通信开销为优化目标,开展了过滤器最优阈值的研究.首先,本文提出了通信开销模型,给出了过滤器最优阈值问题的形式化定义;其次,通过一维随机游走来模拟感知数据的变化并推导出过滤器失效概率的计算公式,通过拉格朗日乘子法给出了求解过滤器最优阈值的数学方法;第三,提出了低计算复杂度的过滤器最优阈值近似算法并分析了近似算法的近似比.理论分析和实验结果验证了提出的算法的正确性以及高效性.在未来的工作中,我们将进一步开展较复杂的通信能量开销模型下的监测算法的研究.

## 参 考 文 献

- [1] Li Jian-Zhong, Li Jin-Bao, Shi Sheng-Fei. Concepts, issues and advance of sensor networks and data management of sensor networks. *Journal of Software*, 2003, 14(10): 1717-1727(in Chinese)  
(李建中, 李金宝, 石胜飞. 传感器网络及其数据管理的概念、问题与进展. *软件学报*, 2003, 14(10): 1717-1427)
- [2] Ren Feng-Yuan, Huang Hai-Ning, Lin Chuang. Wireless sensor networks. *Journal of Software*, 2003, 14(7): 1282-1291(in Chinese)  
(任丰原, 黄海宁, 林闯. 无线传感器网络. *软件学报*, 2003, 14(7): 1282-1291)
- [3] Bonnet P, Gehrke J, Seshadri P. Querying the physical world. *IEEE Personal Communication*, 2000, 7(5): 10-15
- [4] Hefeeda M, Bagheri M. Wireless sensor networks for early detection of forest fires//*Proceedings of the IEEE 4th International Conference on Mobile Adhoc and Sensor Systems*. Pisa, Italy, 2007: 1-6
- [5] Agrawal S, Deb S, Naidu K V M, et al. Efficient detection of distributed constraint violations//*Proceedings of the IEEE 23rd International Conference on Data Engineering*. Istanbul, Turkey, 2007: 1320-1324
- [6] Sharfman I, Schuster A, Keren D. A geometric approach to monitoring threshold functions over distributed data streams. *ACM Transactions on Database Systems*, 2007, 32(4): 23
- [7] Kashyap S, Ramamirtham J, Rastogi R, et al. Efficient constraint monitoring using adaptive thresholds//*Proceedings of the IEEE 24th International Conference on Data Engineering*. Cancun, Mexico, 2008: 526-535
- [8] Dilman M, Raz D. Efficient reactive monitoring. *IEEE Journal on Selected Areas in Communications*, 2002, 20(4): 668-676
- [9] Cormode G, Muthukrishnan S, Yi K. Algorithms for distributed functional monitoring. *ACM Transactions on Algorithms*, 2011, 7(2): 21
- [10] Keralapura R, Cormode G, Ramamirtham J. Communication-efficient distributed monitoring of thresholded counts//*Proceedings of the ACM SIGMOD International Conference on Management of Data*. Chicago, USA, 2006: 289-300
- [11] Wuhib F, Dam M, Stadler R. Decentralized detection of global threshold crossings using aggregation trees. *Computer Networks*, 2008, 52(9): 1475-1761
- [12] Chu D, Deshpande A, Hellerstein J M, Hong W. Approximate data collection in sensor networks using probabilistic models//*Proceedings of the IEEE 22th International Conference on Data Engineering*. Atlanta, USA, 2006: 48-59
- [13] Meng S, Wang T, Liu L. Monitoring continuous state violation in datacenters: Exploring the time dimension//*Proceedings of the IEEE 26th International Conference on Data Engineering*. Long Beach, USA, 2010: 968-979
- [14] Olston C, Loo B T, Widom J. Adaptive precision setting for cached approximate values//*Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*. Santa Barbara, USA, 2001: 355-366
- [15] Olston C, Jiang J, Widom J. Adaptive filters for continuous queries over distributed data streams//*Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*. New York, USA, 2003: 563-574
- [16] Talebi M, Khonsari A, Mohtasham A, et al. Cost-aware monitoring of network-wide aggregates in wireless sensor networks. *Computer Networks*, 2011, 55(6): 1276-1790
- [17] Jurca D, Stadler R. H-GAP: Estimating histograms of local variables with accuracy objectives for distributed real-time monitoring. *IEEE Transactions on Network and Service Management*, 2010, 7(2): 1-13

- [18] Tang M, Li F, Phillips J M, et al. Efficient threshold monitoring for distributed probabilistic data//Proceedings of the IEEE 28th International Conference on Data Engineering. Washington, DC, USA, 2012: 1120-1131
- [19] Keren D, Sagy G, Abboud A, et al. Geometric monitoring of heterogeneous streams. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(8): 1890-1903
- [20] Lazerson A, Sharfman I, Keren D, et al. Monitoring distributed streams using convex decompositions. *Proceedings of the VLDB Endowment*, 2015, 8(5): 545-556
- [21] Garofalakis M, Keren D, Samoladas V. Sketch-based geometric monitoring of distributed stream queries. *Proceedings of the VLDB Endowment*, 2013, 6(10): 937-948
- [22] Sharfman I, Schuster A, Keren D. A geometric approach to monitoring threshold functions over distributed data streams//Proceedings of the ACM SIGMOD International Conference on Management of Data. Chicago, USA, 2006: 301-312
- [23] Keren D, Sharfman I, Schuster A, et al. Shape sensitive geometric monitoring. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 24(8): 1520-1535
- [24] Burdakis S, Deligiannakis A. Detecting outliers in sensor networks using the geometric approach//Proceedings of the IEEE 28th International Conference on Data Engineering. Washington, DC, USA, 2012: 1108-1119
- [25] Garofalakis M, Keren D, Samoladas V. Sketch-based geometric monitoring of distributed stream queries. *Proceedings of the VLDB Endowment*, 2013, 6(10): 937-948
- [26] Giatrakos N, Deligiannakis A, Garofalakis M, et al. Prediction-based geometric monitoring over distributed data streams//Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. Scottsdale, USA, 2012: 265-276
- [27] Zhang Z, Cheng R, Papadias D, et al. Minimizing the communication cost for continuous skyline maintenance//Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data. Rhode Island, USA, 2009: 495-508



**BI Ran**, born in 1985, Ph.D. candidate. Her current research interests include wireless sensor networks and cyber-physical systems.

**LI Jian-Zhong**, born in 1950, professor, Ph.D. supervisor. His current research interests include massive data management, data base, wireless sensor networks and cyber-physical systems etc.

**GAO Hong**, born in 1966, professor, Ph.D. supervisor. Her current research interests include massive data management, data mining and wireless sensor networks and cyber-physical systems etc.

## Background

Wireless Sensor Networks (WSNs) attract more and more attention and have been widely applied in many areas. Detecting abnormal events in a monitored area is one of the fundamental applications in WSNs. Accidents and property damage can be avoided if accurate alarms are informed on time. The threshold based constraint violation monitoring method is the most popular one to detect abnormal events. In these threshold based constraint violation monitoring strategies, a predefined threshold is given and an alarm is triggered when the aggregation result of sensor readings in a monitored area exceeds this threshold. Previous research works focus on heuristic schemes for the design of local filters. However, most of these strategies bring about higher communication overhead and ignore the effect of measurement error and noise on the sensed data. Due to battery energy getting exhausted, sensor node becomes invalid and gets out of use, thus researching on energy efficient monitoring algorithms plays a significant role in the area of sensor networks.

In this paper, we investigate into the problem of optimal filters for minimizing communication cost. Firstly, a formal description for the problem of optimal filters is given. Based on the method of Lagrange Multiplier, a mathematical approach

for calculating optimal filters is provided. And then we propose an approximation algorithm with low computation cost for computing optimal filters, and the ratio bound of this approximation algorithm is analyzed. The theoretical analysis and performance evaluation demonstrate the effectiveness of our algorithm.

This work was supported in part by the National Basic Research Program (973 Program) of China under Grant No. 2012CB316200, the Major Program of the National Natural Science Foundation of China under Grant No. 61190115, the Key Program of the National Natural Science Foundation of China under Grant Nos. 61033015, 60933001, and the General Program of the National Natural Science Foundation of China under Grant No. 61173022.

Our group Massive data computing research center in Harbin Institute of Technology has been working on the research of wireless sensor networks for many years. Some results about congestion control, network capacity, scheduling, and data management including data collection, querying, online processing and mining are achieved. Many high quality papers have been published in worldwide conferences and transactions, such as VLDB, KDD, INFOCOM, MOBIHOC, TKDE, TPDS, VLDB Journal, IPL et al.