

无线传感器网络中基于熵评判的关联支配集构造算法

于瑞云¹⁾ 王兴伟²⁾ 刘永和³⁾

¹⁾(东北大学软件学院 沈阳 110819)

²⁾(东北大学信息科学与工程学院 沈阳 110819)

³⁾(德克萨斯大学阿灵顿分校计算机科学与工程系 美国 德克萨斯 76019)

摘 要 无线传感器网络通常是密集分布的,因此相邻网络节点感知的数据之间具有很大的相关性.为了在无线传感器网络中进行数据冗余缩减,文中提出了一个基于熵评判的关联支配集构造算法(EECDs),算法首先通过评价高斯随机变量的熵值来判断网络节点间的数据相关性,然后分布式地构造一个关联图,最后根据关联图信息移除网络中的冗余节点,构建一个连通关联支配集.基于连通关联支配集的数据采集策略能在密集无线传感器网络中进行高效的数据冗余缩减,显著降低网络的能量消耗,延长网络的生命周期.

关键词 无线传感器网络;信息熵;微分熵;关联图;关联支配集;冗余缩减

中图法分类号 TP393 **DOI号**: 10.3724/SP.J.1016.2011.00087

Correlation Dominating Set Construction Based upon Entropy Evaluation in Wireless Sensor Networks

YU Rui-Yun¹⁾ WANG Xing-Wei²⁾ LIU Yong-He³⁾

¹⁾(Software College, Northeastern University, Shenyang 110819)

²⁾(College of Information Science and Engineering, Northeastern University, Shenyang 110819)

³⁾(Computer Science and Engineering Department, the University of Texas at Arlington, TX 76019, USA)

Abstract Wireless sensor networks are usually densely deployed, so the data sensed from neighboring sensor nodes is highly correlated. For redundancy removal in wireless sensor networks, this paper presents an algorithm named entropy evaluation for correlation dominating set construction (EECDs). The algorithm first determines the correlation degree between sensor nodes by evaluating the entropy of Gaussian random variables, and then distributively generates a correlation graph. Based on the correlation graph, the EECDs algorithm finally constructs a connected correlation dominating set by removing redundant sensor nodes. Data gathering policies with the help of connected correlation dominating sets will greatly reduce data redundancy of dense sensor networks, and therefore result in decrease of energy consumption and prolong lifetime of wireless sensor networks.

Keywords wireless sensor networks; information entropy; differential entropy; correlation graph; correlation dominating set; redundancy removal

收稿日期:2008-11-20;最终修改稿收到日期:2010-12-20.本课题得到国家自然科学基金(61070162,71071028,60802023,70931001)、高等学校博士学科点专项科研基金(20100042110025,20070145017)、中央高校基本科研业务费专项资金(N090504003, N090504006, N100417001)和辽宁省博士科研启动基金项目(20101040)资助.于瑞云,男,1974年生,博士,讲师,主要研究方向为无线自组通信、无线传感器网络、多传感器数据融合等. E-mail: yury@mail.neu.edu.cn.王兴伟,男,1968年生,博士,教授,博士生导师,主要研究领域为下一代互联网、移动 Internet 和 IP/DWDM 光 Internet 等.刘永和,男,1974年生,博士,副教授,主要研究方向为无线组网、传感器网络、安全、系统集成等.

1 引 言

无线传感器网络是由大量具有感知、数据处理和通信能力的传感器节点组成的网络,并且通过传感器节点之间的协作,实现对各种现象的监测.由于具有低功耗、低成本、智能化、分布式、自组织等特点,无线传感器网络在军事、环保、医疗、商业、灾害预测及救援等领域有着广阔的应用前景.

无线传感器网络通常是密集分布的,因此网络节点感知的数据之间通常具有很大的冗余性.由于传感器网络节点是能量受限的,支持冗余缩减的技术能在很大程度上减少网络的能量消耗,从而能够延长无线传感器网络的生命周期.

很多文献提出通过冗余缩减的方法减少在网络中传输的数据量,进行高效数据获取.

文献[1-5]采用编码策略来消除网络中冗余数据,减少向汇聚节点传输的数据量;文献[6-7]提出随机树构造算法为单汇聚节点的数据采集应用构造数据融合树,网络节点感知的数据在向汇聚节点传输的过程中可以在中间网络节点上进行数据融合;文献[8-9]提出的方法都是结合无线传感器网络节点的空间相关度,研究从网络中选取一部分节点向汇聚节点传输感知的数据,以此消除无线传感器网络中的冗余数据.

本文在关联图和关联支配集概念的基础上设计了一个基于熵评判的关联支配集构造算法(Entropy Evaluation for Correlation Dominating Set Construction,EECDs),构造一个连通关联支配集对剩余节点进行关联支配,并利用连通关联支配集中的节点进行数据感知、采集和传输,实现数据冗余缩减的目的.

文献[8-9]提出的方法与本文类似,但文献[8]只考虑直接邻居节点之间的相关性.与本文提出的算法一样,文献[9]也考虑了 k 跳邻近节点的相关性,本文提出了一个熵评判(entropy evaluation)方法,与文献[9]中的方法相比,熵评判方法能更加准确地度量网络节点之间的数据相关度,另外本文在构造连通关联支配集时采用了一个能量感知的优先级策略,能有效地均衡无线传感器网络的能量消耗.

EECDs算法首先通过评价随机变量的熵值来判断 g 跳网络节点间的数据相关性,然后在网络中分布式地构造一个关联图,最后借助关联图的信息构造一个连通关联支配集. EECDs算法能有效缩减无线传感器网络的数据冗余,同时具有很好的能量

均衡特性和可扩展性.

2 网络模型

将无线传感器网络模型化为一个单位圆图 $G(V,E)$,其中 V 表示网络节点集, E 为在通信半径范围内的节点所形成的边的集合.假定网络是密集连通的,即集合 E 具有较大的势,因此相邻网络节点采集的数据具有较大的相关性.

无线传感器网络所监测的现象大都是空间上分散的,且符合高斯分布,这样的物理环境通常称为高斯随机场.因此,我们假设 n 个网络节点采集的空间数据向量 \mathbf{X} 具有高斯随机过程的特征,并且符合 n 维正态分布 $G_n(\mu,\mathbf{K})$,则随机向量 \mathbf{X} 的概率密度函数为

$$f(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^n |\mathbf{K}|^{1/2}} e^{-\frac{1}{2}(\mathbf{X}-\mu)^T \mathbf{K}^{-1}(\mathbf{X}-\mu)} \quad (1)$$

其中参数 μ 为正态分布的均值, $\mathbf{K}=(k_{ij})_{n \times n}$ 为 \mathbf{X} 的协方差矩阵, $k_{ij}=\sigma^2 e^{-cd_{ij}^\beta}$, $\beta \in \{1,2\}$, d_{ij} 为网络节点 i 和 j 之间的距离, c 为调节系数.

协方差矩阵 \mathbf{K} 刻画了网络空间数据的相关度,节点之间的数据相关度会随着距离的增加而缩减.

为表示无线传感器网络中的数据相关性,本文借鉴了文献[9]中提出的关联图(correlation graph)的概念.

定义 1. 给定一个无线传感器网络,传感器节点集合为 R ,则关联图定义为以 R 为顶点集合、以 $P(R) \times R$ 的一个子集为有向超边(hyperedge)集合的有向超图(hypergraph),其中 $P(R)$ 是集合 R 的幂集(power set),该关联图用 $G^*(V=R, E \subseteq P(R) \times R)$ 表示.

关联图 G^* 中的一条边 (W,u) , $u \notin W$ 表示节点 u 感知的数据与集合 W 中的节点感知的数据高度相关,因此节点 u 的数据可以在一定的误差范围之内由集合 W 中的节点推算出,换言之,在给定集合 W 中节点感知的数据的基础上,节点 u 的数据是冗余的.将集合 W 定义为关联子集(correlated subset),节点 u 定义为关关节点(correlated vertex).

下面在关联图的基础上,给出关联支配集(correlation dominating set)的概念.

定义 2. 关联支配集是图 $G(V,E)$ 中的顶点集合 $S \subseteq V$,满足对于任何顶点 $u \in (V-S)$,至少存在一个节点子集 $W \subseteq S$,且 (W,u) 为关联图 G^* 中的一个有向超边.如果集合 S 的诱导子图是图 G 的一个连通子图,则称 S 为连通关联支配集(connected

correlation dominating set).

图论中的支配集 (dominating set) 概念如定义 3 所述.

定义 3. 支配集是图 $G(V, E)$ 中的顶点集合 $S \subset V$, 满足对于任何顶点 $v \in V$, 或者 $v \in S$, 或者 v 与集合 S 中至少一个顶点相邻.

与支配集的概念不同, 在关联支配集中, 非支配集节点不一定要与支配集中的节点相邻, 但是要确保非支配集中的任何节点都与支配集中的某些节点具有较高的相关度. 相关度的度量方法将在第 3 节中进行详细定义.

3 问题定义

基于条件熵评判的关联支配集构造算法 (EECDs) 是在给定的无线传感器网络中根据关联图结构来构造一个最小的连通关联支配集 C , 通过收集集合 C 中节点感知的数据能够以较小的误差复原网络中所有节点产生的数据. EECDs 是通过减少参与通信的节点数量来消除网络中冗余数据的传输, 能在保证网络数据完整性的基础上, 有效地降低网络节点的能量消耗. 另外, EECDs 算法还可以对网络节点进行有效的休眠调度, 均衡网络的能量消耗.

在计算给定图 $G(V, E)$ 的关联图 G^* 时, 我们采用熵评判机制来判断超边的存在, 具体地说, 如果节点 u 相对于集合 W 中节点的条件熵小于一个应用程序定义的阈值时, 则关联图 G^* 中就存在一个从集合 W 至节点 u 的有向超边 (W, u) .

如第 2 节中所述, n 个网络节点采集的数据 \mathbf{X} 具有高斯随机过程的特征, 并且符合 n 维正态分布 $G_n(\mu, \mathbf{K})$, \mathbf{K} 为 \mathbf{X} 的协方差矩阵. 集合 W 为集合 \mathbf{X} 的一个子集, 假设 $W = (X_{i_1}, X_{i_2}, \dots, X_{i_k})$, 其中 $i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}$, $k \leq n$ 为任意节点 ID 组合, 集合 W 符合 k 维正态分布 $G_k(\mu, \mathbf{K}_W)$, 其协方差矩阵为 \mathbf{K}_W , \mathbf{K}_W 为协方差矩阵 \mathbf{K} 中与 ID 值 i_1, i_2, \dots, i_k 相关的子矩阵.

假设无线传感器网络节点感知的数据都以同样的步长独立地进行量化分级, 因此在这里将采用条件微分熵进行节点数据相关度的评判. 假设 $W = (X_{i_1}, X_{i_2}, \dots, X_{i_k})$ 是在较小的步长 Δ 下进行独立量化分级得到的 k 个离散的随机值, 根据文献[10], k 维正态分布 $G_k(\mu, \mathbf{K}_W)$ 的微分熵为

$$h(G_k(\mu, \mathbf{K}_W)) = \frac{1}{2} \log(2\pi e)^k |\mathbf{K}_W| \quad (2)$$

该正态分布的联合熵近似为

$$\begin{aligned} H(G_k(\mu, \mathbf{K}_W)) &\approx h(G_k(\mu, \mathbf{K}_W)) - k \log \Delta \\ &= \frac{1}{2} \log(2\pi e)^k |\mathbf{K}_W| - k \log \Delta \quad (3) \end{aligned}$$

则在给定集合 W 节点的数据的基础上, 节点 u 需要传输的数据率 R_u 即为节点 u 相对于集合 W 的条件熵, 其值如式(4)所示, 其中 $W' = W \cup \{u\}$:

$$\begin{aligned} R_u &= H(u, W) - H(W) \\ &= \frac{1}{2} \log(2\pi e)^{k+1} |\mathbf{K}_{W'}| - (k+1) \log \Delta - \\ &\quad \left(\frac{1}{2} \log(2\pi e)^k |\mathbf{K}_W| - k \log \Delta \right) \\ &= \frac{1}{2} (\log(2\pi e)^{k+1} |\mathbf{K}_{W'}| - \log(2\pi e)^k |\mathbf{K}_W|) - \\ &\quad \log \Delta \quad (4) \end{aligned}$$

根据文献[10], 一个符合正态分布的连续随机变量的微分熵为

$$h(G(\mu, \sigma^2)) = \frac{1}{2} \log(2\pi e \sigma^2) \quad (5)$$

则节点 u 的离散熵的值为

$$H_u = h_u - \log \Delta = \frac{1}{2} \log(2\pi e \sigma^2) - \log \Delta \quad (6)$$

H_u 实际上为节点 u 独立传送感知数据的数据率, 所以当 R_u 的值小于一个给定的阈值 μH_u , $\mu \rightarrow 0$ 时, 则可以判定节点 u 与节点集合 W 高度相关, 即在超图 G^* 中存在有向超边 (W, u) , 其中参数 μ 为误差系数, 由 $R_u < \mu H_u$ 可得

$$\begin{aligned} &\frac{1}{2} (\log(2\pi e)^{k+1} |\mathbf{K}_{W'}| - \log(2\pi e)^k |\mathbf{K}_W|) - \log \Delta < \\ &\mu \left(\frac{1}{2} \log(2\pi e \sigma^2) - \log \Delta \right) \quad (7) \end{aligned}$$

EECDs 算法首先通过熵评判机制构造一个关联图, 然后在此基础上构造一个连通关联支配集. 由于在无线传感器网络中, 节点间数据的相关度会随着节点距离的增加而呈现指数衰减, 所以在构造关联图的时候, 每个网络节点只判定与其 g 跳范围内节点集的幂集元素的相关性, g 通常取 $1 \sim 3$, 这能够在很大程度上减少计算的开销.

4 算法描述

4.1 m 跳邻居发现

为均衡无线传感器网络的能量消耗, 算法引入了一个能量感知的优先级策略, 节点用二元组 $(energy, id)$ 进行优先级分级, 其中 $energy$ 代表节点的剩余能量值, id 代表节点的 ID 值.

定义 4. 设节点 u 的优先级 $p_u = (energy_u, id_u)$,

节点 v 的优先级 $p_v = (energy_v, id_v)$, 则 $p_u > p_v$, 当其仅当:

- (1) $energy_u < energy_v$ 或
- (2) $energy_u = energy_v$ 且 $id_u < id_v$

显而易见, 剩余能量值较低的节点将具有较高的优先级, 在 EECDS 算法中, 此类节点将以较高的优先级从连通关联支配集中移除, 被移除的节点在数据采集的时候不参与通信, 因而减少了能量的消耗, 延长了节点的生存时间。

在数据采集过程中消耗能量过多的节点会在下一轮连通关联支配集构造的过程中获得较高的优先级, 因此能够均衡网络的能量消耗。

EECDS 算法要求传感器网络节点能通过接收 GPS 信号或者其它定位方法确定各自的位置信息。

邻居发现阶段通过 m 轮 HELLO 消息交换过程让网络中的每个节点得到其 m 跳之内的所有邻居节点的状态信息, 包含节点的 ID、剩余能量、位置、着色(见 4.3 节)等信息。

在第 1 轮消息交换过程中, 每个节点广播一个 HELLO 消息, 收到 HELLO 消息的节点将其一跳邻居节点的信息保存在邻居列表里, 直至它从所有的一跳邻居处都收到 HELLO 消息。

在第 $i \in [2, m]$ 轮消息交换中, 每个节点在广播的 HELLO 消息中捎带它的 $i-1$ 跳邻居列表信息, 让该节点的直接邻居节点获取各自的 i 跳邻居节点的信息。

经过 m 轮消息交换后, 每个节点都会完成其 m 跳邻居节点信息的收集工作, 这些邻居节点的信息存储在邻居列表 NBList 中。

如第 3 节中所述, 每个网络节点只判定与其 g 跳范围内节点集的幂集元素的相关性, 通常令 $g \in \{1, 2, 3\}$, g 的取值要保证在生成关联图时让节点获得足够的关联邻居节点的信息(见 4.2 节)。

在邻居发现阶段, 通常设定跳数 $m \in \{2, 3\}$ 且 $m \geq g$, m 的取值要保证在构造连通关联支配集时, 节点要有足够的邻居节点信息判断移除该节点后本地网络的连通性(见 4.3 节)。

4.2 生成关联图

EECDS 算法通过本地算法构造一个关联图, 在关联图构造完成之后, 每个节点成为关联图中的一个关联节点, 并且维护与该节点相邻接的关联子集的信息以及该节点所在的关联子集的信息。

具体过程如下:

1. 假设集合 $N_g(u)$ 为节点 u 的 g 跳开放邻居节点集, 对于 $N_g(u)$ 的幂集 $P(N_g(u))$ 中的每个集合元素 $P_j, j \in$

$\{1, 2, \dots, 2^{|N_g(u)|}\}$, 令 $W = P_j$, 计算节点 u 与节点集合 W 的关联关系, 如果满足式(7)的约束关系, 则将集合 W 加入到节点 u 的关联子集列表 CSList 中;

2. 每个关联节点计算完与其邻接的关联子集信息之后, 向其 g 跳邻居节点广播 CSNOTIFY 消息, 在消息中包含该节点的 ID 和 CSList 信息;

3. 当一个节点收到 CSNOTIFY 消息时, 如果该节点 ID 出现在 CSNOTIFY 消息的 CSList 中的记录中, 则将这些记录保存在关联节点列表 CVList 中, 并在 CVList 中保存相应关联节点的信息。

经过上述过程, EECDS 算法通过节点的分布式计算和有限 g 跳广播构造了一个关联图, 每个节点只掌握其与 g 跳范围之内的邻居节点构成的局部关联图信息, 其中, 与该节点邻接的关联子集的信息存储在关联子集列表 CSList 中, 包含该节点的关联子集的信息保存在关联节点列表 CVList 中。

4.3 构造连通关联支配集

连通关联支配集的构造过程是在连通的网络中分布式移除一些关联节点, 在移除这些关联节点的同时保证网络的正常连通, 直至构造一个较小的连通关联支配集对整个网络进行关联支配。

EECDS 算法采用一个着色过程 (coloring process) 完成连通关联支配集的构造, 具体实现过程描述如下:

1. 在起始阶段, 每个节点都标记为灰色 (GRAY);

2. 符合以下条件的灰色节点 u 将自己标记为白色 (WHITE):

2.1. 查看邻居列表 NBList, 假设节点 u 所有 m 跳范围内的非白色节点集合为 $Z \subseteq N_m(u)$, 节点 u 的优先级 $p_u > p_{x_i}, \forall x_i \in Z, x_i$ 为灰色节点, 并且集合 Z 的诱导子图是连通的;

2.2. 检查关联子集列表 CSList, 在关联图 G^* 中存在一个超边 (W, u) , 对于节点 $v \in W$, 或者节点 v 是黑色 (BLACK) 的, 或者节点 v 是灰色的且 $p_v < p_u$, 如果存在多个满足条件的集合 W , 选取其中具有最小势的集合;

2.3. 检查关联节点列表 CVList, 对于关联图 G^* 中的每一个超边 $(U, v), u \in U$, 节点 v 或者为白色, 或者为黑色, 或者是灰色的且 $p_v < p_u$ 。

3. 当灰色节点 u 将自己标记为白色后, 向其 g 跳范围内的邻居节点发送一个 WHITEN 消息;

4. 集合 W 中的灰色节点收到节点 u 发来的 WHITEN 消息后, 将自己标记为黑色, 然后发送一个 BLACKEN 消息;

5. 其它灰色节点收到一个 WHITEN 消息或者一个 BLACKEN 消息后, 更新邻居列表 NBList 中发送消息节点的着色信息, 重新测试步 2 中的条件, 如果满足条件, 开始另一轮着色过程;

6. 当网络中的灰色节点全部被标记成黑色或白色, 或

者网络中没有符合步 2 中条件的灰色节点, 算法结束.

EECDs 算法执行完毕后, 所有黑色节点和灰色节点(如果存在)构成一个连通关联支配集. 需要说明的是, EECDs 算法是采用分布式的节点退出机制构造连通关联支配集, 在算法执行的任何时刻, 所得到的关联支配集都是连通的.

5 性能评价

在本节中, 我们将通过仿真来分析 EECDs 算法的性能, 为进行比较, 同时对文献[9]中的算法(为方便表述, 用 EGCD 标记该算法)进行仿真分析.

5.1 协方差矩阵参数分析

为构造关联图, 节点需要判断与其邻接的关联子集是否满足式(7)的条件, 式(7)中的协方差矩阵元素 $k_{ij} = \sigma^2 e^{-cd_{ij}^2}$ 表示节点之间数据的相关度, σ^2 为正态分布的方差, d_{ij} 为网络节点 i 和 j 之间的距离, 节点之间的数据相关度会随着距离的增加而缩减, β 的值通常取 1 或 2.

当 $\sigma=1, \beta=2$ 时, 在系数 c 取不同的值时, 两个节点的数据相关度随距离的变化情况如图 1 所示, c 值越大, 在给定节点距离的情况下, 两个节点之间的数据相关度越低.

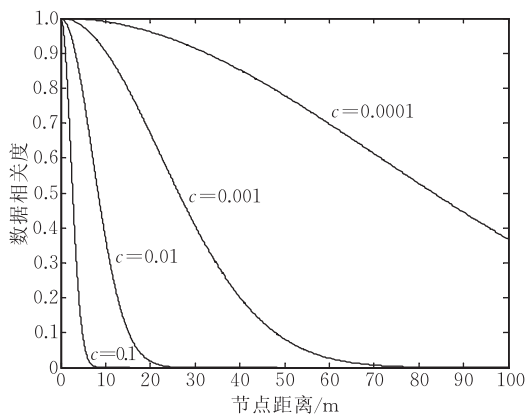


图 1 节点数据相关度随距离的变化

5.2 连通关联支配集的大小

在本节中, 仿真环境配置为: 网络节点的通信半径 $\gamma=30$ m, 网络节点随机布置在 $100\text{ m} \times 100\text{ m}$ 的网络区域内, 节点数量分别取 60、80、100、120、140 和 160.

式(7)中的误差系数 $\mu=0.2$, 微分熵步长 $\Delta=2^{-4}$, 协方差矩阵参数 $\sigma=1, \beta=2$. 关联邻居跳数 $g=1$, 通信邻居跳数 $m=2$.

图 2 显示了在协方差矩阵参数 $c=0.0001$ 和 $c=0.0005$ 时 EECDs 算法获得的连通支配集大小以及 EGCD 算法得到的连通支配集的大小. 在网络

密集的情况下, EECDs 算法具有较好的性能. 当 $c=0.0001$ 时, EECDs 算法获得较好的性能, 但当 $c=0.0005$ 时, 在节点数为 60 和 80 时, 网络中只有极少数节点被缩减掉.

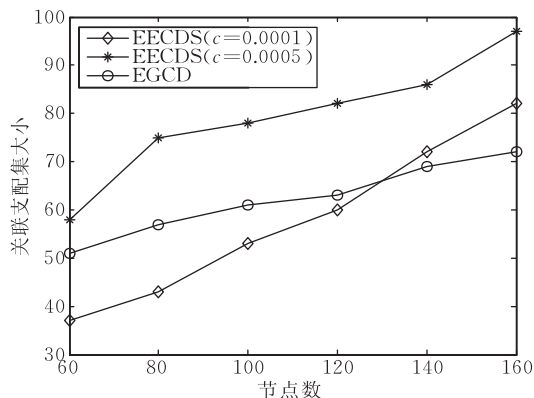


图 2 连通关联支配集的大小

EGCD 算法采用一个简单的概率方法确定节点的关联子集, 在仿真中取关联百分比 $P=0.1$, 关联距离 $d=20$ m, 当网络密度增大时, EGCD 算法能获得比 EECDs 更好的性能, 但在网络稀疏时性能相对较差.

图 3 显示了在上述配置情况下, 节点数量为 80 时连通关联支配集的构造情况, 图 3(a)为网络的初始连接, 图 3(b)为得到的连通关联支配集, 其中 *

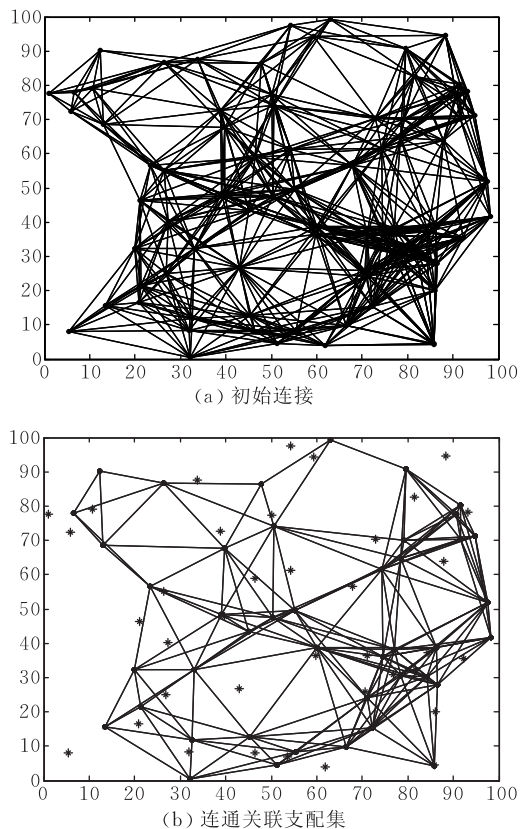


图 3 节点数为 80 时连通关联支配集的构造

号表示 EECDs 算法缩减掉的网络节点。

5.3 各项参数对关联支配集大小的影响

5.3.1 误差系数 μ 的影响

在本小节中,仿真环境配置为:网络节点的通信半径 $\gamma = 30$ m,网络节点随机布置在 $100\text{ m} \times 100\text{ m}$ 的网络区域内,节点数量分别取 80 和 100。

微分熵步长 $\Delta = 2^{-4}$,协方差矩阵参数 $\sigma = 1$, $\beta = 2$, $c = 0.0002$ 。关联邻居跳数 $g = 1$,通信邻居跳数 $m = 2$ 。

根据式(7),调整误差系数 μ 将影响熵评判的结果,增大 μ 将构造一个更大的关联图,因而在构造连通关联支配集的时候,节点将有更多的关联子集可供选择,所以将缩减连通支配集的大小。如图 4 所示,当误差系数 μ 从 0.1 调整到 0.6 时,在 80 个节点的情况下,连通关联支配集的大小从 50 降到 36,当网络密度增加时,连通关联支配集大小占网络节点数量的比率会随之降低,如图 4 中 100 个网络节点的情况,当 μ 从 0.1 调整到 0.6 时,其连通支配集大小从 61 降到 39,当 $\mu = 0.6$ 时,比率值为 39%。

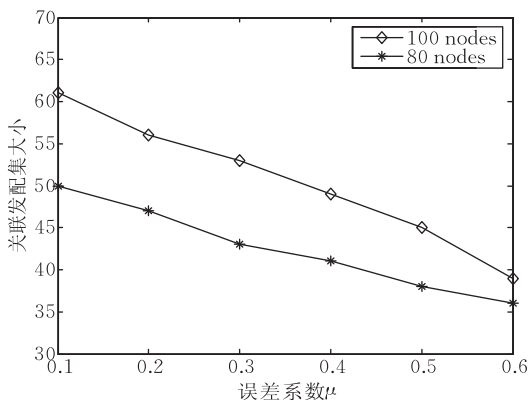


图 4 误差系数 μ 对关联支配集大小的影响

5.3.2 微分熵步长值 Δ 的影响

在本小节中,仿真环境配置为:网络节点的通信半径 $\gamma = 30$ m,网络节点随机布置在 $100\text{ m} \times 100\text{ m}$ 的网络区域内,网络节点数分别取 80 和 100。

协方差矩阵参数 $\sigma = 1$, $\beta = 2$, $c = 0.0002$,误差系数 $\mu = 0.2$,关联邻居跳数 $g = 1$,通信邻居跳数 $m = 2$ 。

微分熵步长 Δ 分别取 $1/32$ 、 $1/16$ 、 $1/8$ 、 $1/4$ 、 $1/2$,图 5 显示了步长 Δ 对连通关联支配集大小的影响。

如式(3)和式(6)所示, k 维正态分布的离散熵为其微分熵与 $k \log \Delta$ 的差值,因此,当 Δ 取值越大, k 维正态分布变量的离散熵的值越小。由于 μ 的取值小于 1,当步长 Δ 增大时,式(7)两端的值都减小,

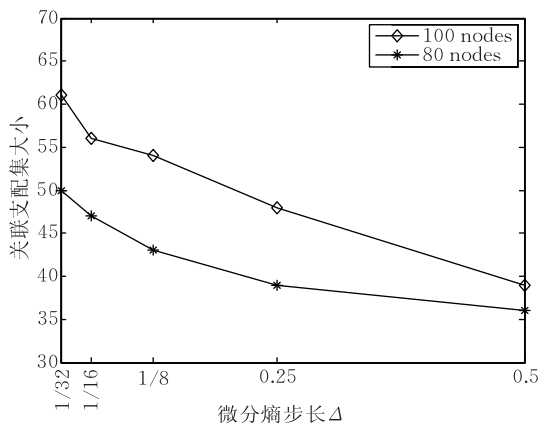


图 5 微分熵步长 Δ 对关联支配集大小的影响

但左端值的变化速度要高于右端,其结果是式(7)得到满足的可能性增加了,也就是说,在构造关联图时,节点的关联子集数量增多了,因此,获得最优连通关联支配集的可能性随之提高,如图 5 所示,当步长 Δ 增大时,组成连通关联支配集的节点数量呈现出降低的趋势。

5.4 算法的能耗特性

5.4.1 平均能量消耗

本节的仿真配置为:网络节点的通信半径 $\gamma = 30$ m,网络节点随机布置在 $100\text{ m} \times 100\text{ m}$ 的网络区域内,节点数量分别取 60、80、100、120、140、160。式(7)中的误差系数 $\mu = 0.2$,微分熵步长 $\Delta = 2^{-4}$,协方差矩阵参数 $\sigma = 1$, $\beta = 2$, $c = 0.0002$ 。关联邻居跳数 $g = 2$,通信邻居跳数 $m = 2$ 。

假设仿真过程中发送和接收的数据包都具有同样大小,能耗情况根据文献[11]中的能耗模型给出,网络节点发送数据包消耗 1 个单位的能量,接收数据包消耗 0.5 个单位的能量,网络节点在空闲状态下的能耗忽略不计。

EECDs 算法的能耗主要包括在邻居发现阶段和连通关联支配集构造阶段发送和接收消息所消耗的能量以及在生成关联图阶段用于计算的能量开销。构造关联图时,每个节点计算并比对 g 跳开放邻居节点集的幂集元素与其的相关度,该过程的算法复杂度为 $O(n^3 \cdot 2^n)$,其中 n 为网络节点的 g 跳开放邻居节点集的势。对于密集网络,这个计算过程是非常耗资源的,因此,在实际运算的过程中,每个节点最多选取与其最相关的 s 个节点(比如 $s = 12$)进行关联图的构造,假设 $n \geq s$ 的网络节点生成关联图所消耗的能量与发送 10 个消息消耗的能量相等,则对于 $n < s$ 的节点,生成关联图的能耗值等同于发送 $10 \cdot (n/s)^3 \cdot 2^{n-s}$ 个消息。

EGCD 算法的关联图构造过程的算法复杂度为 $O(K^2L^6)$, 其中 K 是为了计算相关度而收集的采样个数, L 为 rank 值, 并且 $L < K$, 所以每个节点为计算关联图而消耗的能量基本相同, 假设均与发送 10 个消息所消耗的能量相等 (文献 [9] 中假设与发送 40 个消息的能耗相等)。

此外, EGCD 算法是通过收集邻近节点的采样数据来判断与邻近节点的数据相关性, 在开始构造关联图之前, 每个节点还需要广播收集到 i 个采样数据供其它节点进行数据相关性分析。

在不同的网络拓扑结构下连续运行 EECDS 算法和 EGCD 算法各 10 次, 图 6 显示了两种算法的平均能量消耗情况, 在每种网络配置下, EECDS 算法的节点平均消耗都要优于 EGCD 算法。

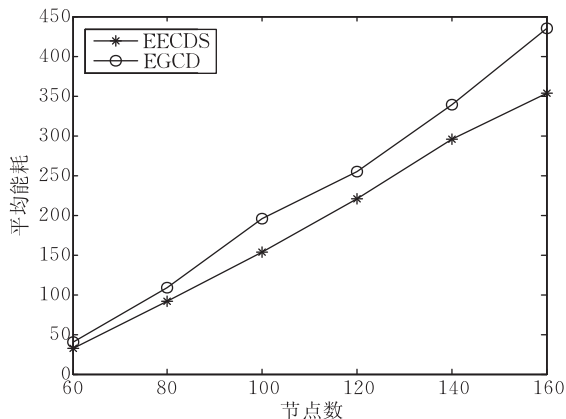


图 6 算法的平均能量消耗

5.4.2 能量均衡特性

本节的仿真配置为: 网络节点的通信半径 $r = 30\text{m}$, 80 个网络节点随机布置在 $100\text{m} \times 100\text{m}$ 的网络区域内。误差系数 $\mu = 0.2$, 微分熵步长 $\Delta = 2^{-4}$, 协方差矩阵参数 $\sigma = 1, \beta = 2, c = 0.0002$ 。关联邻居跳数 $g = 2$, 通信邻居跳数 $m = 2$ 。

每个节点的初始能量值设为 1450~1500 个单位之间的随机值。在相同的网络拓扑结构下连续运行 EECDS 算法和 EGCD 算法各 5 次, 在每次执行过程中, 首先构造一个连通关联支配集, 然后在此支配集的基础上随机选择一个节点为根节点构造一个树状通信结构, 并在此通信树结构上运行一个简单的查询应用程序 10 次。该应用程序首先从根节点洪泛一个 QUERY 请求消息, 然后从每个节点沿着通信树结构向根节点发回一个数据包。

仿真结束后, 网络节点的能耗情况如图 7 所示。

从图 7 可以看出, EECDS 算法具有很好的能量均衡性, 这是因为 EECDS 算法采用了能量感知的

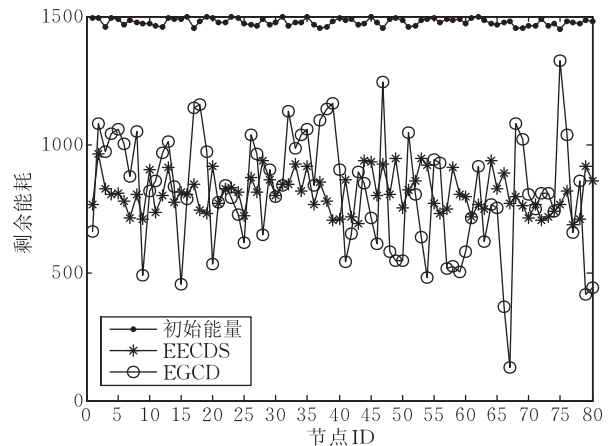


图 7 算法的能量均衡性

优先级机制, 组成连通关联支配集的节点负责网络数据的采集和传送, 要消耗更多的能量, 因而在下一轮关联支配集的构造过程中将获得较高的优先级, 增大了从网络中移除的可能性。

EECDS 算法的能量均衡特性对延长无线传感器网络的生命周期起到了十分显著的作用。

6 相关研究与本文结论

文献 [1] 主要研究单一输入的编码策略, 提出了两种编码方法: 外编码和自编码。针对这两种方法, 分别提出了构造数据采集树的最优算法和近似最优算法, 目的是最小化在网络中传输的数据的加权比特数。

文献 [2] 提出用数据采集节点跟踪网络节点的相关性信息, 当获得足够的节点相关信息, 数据采集节点通知各个网络节点进行数据编码的比特位数, 从而进行有效地冗余抑制。

文献 [3-4] 设计了两种编码策略, 一种基于条件熵, 另一种基于 Slepian-Wolf 编码模型 [12], 算法研究了在两种编码策略下, 进行通信结构和节点数据率分配的联合优化问题, 从而在单汇聚节点的树状通信结构中最小化网络数据传输的能量开销。

文献 [5] 考虑了以数据采集为背景的传感器网络节点布置和通信结构的联合优化问题, 目的是优化布置传感器网络节点, 使网络节点产生的数据能在汇聚节点处以较小的失真率进行数据重建, 同时最小化网络通信的能量消耗。文献研究了在满足网络最大失真和平均失真约束条件下的基于条件熵和 Slepian-Wolf 编码的算法设计。

文献 [6] 提出了一个随机树构造算法为单汇聚节点的数据采集应用构造数据融合树, 网络节点感

知的数据在向汇聚节点传送的过程中,可以在中间网络节点上进行数据融合,从而减少了在网络中传输的数据量.该算法对于所有非降的凹聚合函数都有很好的近似比.

在文献[7]中,为在传感器网络中构造进行高效数据采集的数据融合树,作者提出了一个简单的随机树构造算法,该算法在以栅格图方式布置的网络中具有常数近似比.

文献[8]根据传感器网络节点的空间相关度,研究从网络中选取一部分节点向汇聚节点传输感知的数据,以消除网络中的冗余数据.文献提出的算法采用一个简单的本地算法选取一些簇头节点,然后选取一些中间节点将簇头节点连通成一个最短路径树,在树结构之外的节点将不参与数据采集的通信过程.

文献[9]先从邻近节点收集一定数量的采样数据,然后根据这些采样数据判断与邻近节点的相关度,并基于节点相关程度构造一个关联图.接着在关联图的基础上构造一个连通关联支配集,关联支配集之外的节点将从数据采集应用的通信结构中移除.

文献[13-19]研究了在无线网络中构造连通支配集的算法,但这些算法都是研究通用的连通支配集构造策略,不是专门基于传感器网络背景的,因而也没有考虑在无线传感器网络中进行数据冗余缩减的问题.

文献[20-22]是从路由的角度设计一些分布式算法来鉴别在路由过程中节点的冗余性,也就是在满足网络路由性能要求的前提下,可以在路由通信过程中排除一些网络节点,这些节点可以被调度休眠,从而节省网络的能量消耗.但这些策略只是考虑网络节点的冗余,而没有考虑节点之间数据的相关性.

本文提出了一个基于熵评判的关联支配集构造算法(EECDs),算法首先在网络中分布式构造一个关联图,然后结合关联图的信息构造一个连通关联支配集.通过收集关联支配集中节点感知的数据,能够以较小的误差复原网络中所有节点产生的数据,因此EECDs算法能在保证网络数据完整性的前提下有效地缩减无线传感器网络的数据冗余,减少在网络中传输的数据量,从而降低网络的能量消耗,延长网络的生命周期.

参 考 文 献

- [1] Von Rickenbach P, Wattenhofer R. Gathering correlated data in sensor networks//Proceedings of the 2004 Joint Workshop on Foundations of Mobile Computing (DIALM-POMC'04). Philadelphia, PA, United States, 2004: 60-66
- [2] Chou J, Petrovic D, Kannan R. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks//Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03). San Francisco, CA, USA, 2003: 1054-1062
- [3] Cristescu R, Vetterli M. Power efficient gathering of correlated data: Optimization, np-completeness and heuristics. SIGMOBILE Mobile Computing and Communications Review, 2003, 7(3): 31-32
- [4] Cristescu R, Beferull-Lozano B, Vetterli M. On network correlated data gathering//Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'04). Hong Kong, China, 2004: 2571-2582
- [5] Ganesan D, Cristescu R, Beferull-Lozano B. Power-efficient sensor placement and transmission structure for data gathering under distortion constraints//Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN'04). Berkeley, CA, USA, 2004: 142-150
- [6] Goel A, Estrin D. Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk. Algorithmica, 2005, 43(1-2): 5-15
- [7] Enachescu M, Goel A, Govindan R, Motwani R. Scale free aggregation in sensor networks//Proceedings of the 1st International Workshop on Algorithmic Aspects of Wireless Sensor Networks (Algosensors'04). Turku, Finland, 2004: 71-84
- [8] Yoon SunHee, Shahabi Cyrus. Exploiting spatial correlation towards an energy efficient clustered aggregation technique (CAG)//Proceedings of the International Conference on Communications (ICC'05). Seoul, South Korea, 2005: 3307-3313
- [9] Gupta H, Navda V, Das S, Chowdhary V. Efficient gathering of correlated data in sensor networks. ACM Transactions on Sensor Networks, 2008, 4(1): 1-31
- [10] Cover T M, Thomas J A. Elements of Information Theory. New York: John Wiley and Sons, Inc., 1991
- [11] Kubisch M, Karl H, Wolisz A, Zhong L C et al. Distributed algorithms for transmission power control in wireless sensor networks//Proceedings of the 2003 IEEE Wireless Communications and Networking Conference (WCNC'03). New Orleans, LA, USA, 2003: 558-563
- [12] Slepian D, Wolf J K. Noiseless coding of correlated information sources. IEEE Transactions on Information Theory, 1973, IT-19(4): 471-480
- [13] Stojmenovic I, Seddigh M, Zunic J. Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks. IEEE Transactions on Parallel and Distributed Systems, 2002, 13(1): 14-25
- [14] Adjih C, Jacquet P, Viennot L. Computing connected dominated sets with multipoint relays. Ad Hoc and Sensor Networks, 2005, 1(1): 27-39

- [15] Wu J, Li H. On calculating connected dominating set for efficient routing in ad hoc wireless networks//Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications. Seattle, Washington, United States, 1999
- [16] Dai F, Wu J. An extended localized algorithm for connected dominating set formation in ad hoc wireless networks. *IEEE Transactions on Parallel and Distributed Systems*, 2004, 15(10): 908-920
- [17] Wan Peng-Jun, Alzoubi K M, Frieder O. Distributed construction of connected dominating set in wireless ad hoc networks//Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'02). New York, NY, USA, 2002: 1597-1604
- [18] Dai F, Wu J. On constructing k -connected k -dominating set in wireless networks//Proceedings of the 19th IEEE International Symposium on Parallel and Distributed Processing. Denver, CO, USA, 2005: 10
- [19] Thai M T, Zhang N, Ravi T, Xu X. On approximation algorithms of k -connected m -dominating sets in disk graphs. *Theoretical Computer Science*, 2007, 385(1-3): 49-59
- [20] Chen B, Jamieson K, Balakrishnan H, Morris R. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks//Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom'01). Rome, Italy, 2001
- [21] Xu Y, Heidemann J, Estrin D. Geography-informed energy conservation for ad hoc routing//Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom'01). Rome, 2001: 70-84
- [22] Ye F, Zhong G, Cheng J, Lu S et al. Peas: A robust energy conserving protocol for long-lived sensor networks//Proceedings of the International Conference on Distributed Computing Systems (ICDCS'03). Providence, RI, United States, 2003: 28-37



YU Rui-Yun, born in 1974, Ph. D., lecturer. His research interests include wireless ad-hoc communication, wireless sensor networks, multi-sensor data fusion, etc.

WANG Xing-Wei, born in 1968, Ph. D., professor, Ph. D. supervisor. His research interests include NGI, mobile wireless Internet, IP/DWDM optical Internet, etc.

LIU Yong-He, born in 1974, Ph. D., associate professor. His research interests are wireless networking, sensor networks, security, and system integration.

Background

Wireless sensor networks are data-oriented, whose main task is data processing, including sensing, data gathering, transmission, compress, aggregation, and so on. Wireless sensor networks are usually densely deployed in a network area, and generate a great deal of data. Moreover, the data sensed by neighboring nodes is highly correlated. These all impose great challenges on the design of data gathering algorithms. Since the sensor nodes are power-constrained, transferring large amounts of data definitely will shorten the lifetime of sensor networks. Therefore, the network performance on energy consumption and stability could be improved through data redundancy removal.

In this paper, an algorithm named the entropy evaluation for correlation dominating set construction (EECDs) is presented. The EECDs algorithm exploits the concept of the correlation dominating set. It first determines the correlation between sensor nodes by evaluating the entropy of random variables, and then distributively generates a correlation graph. At last, the algorithm constructs a correlation dominating set using the information of the correlation graph. The EECDs algorithm is very efficient on reducing data redundancy in wireless sensor networks, and performs well on energy balance and scalability.

This work is supported by the National Natural Science Foundation of China under grant Nos. 61070162, 71071028, 60802023 and 70931001; the Specialized Research Fund for the Doctoral Program of Higher Education under grant Nos. 20100042110025 and 20070145017; the Fundamental Research Funds for the Central Universities under grant Nos. N090504003, N090504006 and N100417001; the Doctoral Scientific Research Fund of Liaoning Province under grant No. 20101040.

One of main tasks of the above mentioned projects is virtual backbone construction in wireless networks. Some published papers were focusing on connected dominating set construction, and generating connected dominating sets as backbones for routing, data collection, etc.

This paper highly addresses the correlation fact of wireless sensor networks, and proposes a connected correlation dominating set construction algorithm. The correlation dominating set is used for data redundancy removal, and data gathered from the set could recover the network data within a predefined error bound. The entropy evaluation method exploited in the algorithm can leverage data correlation well, and hence guarantee the efficiency of data redundancy removal in wireless sensor networks.