E-mail: csa@iscas.ac.cn http://www.c-s-a.org.cn Tel: +86-10-62661041

拓扑感知的无线传感网络数据聚合方法^①

熊 英¹, 李 芸²

¹(江门开放大学 信息技术部, 江门 529234) ²(东华理工大学 水资源与环境工程学院, 南昌 330013) 通讯作者: 熊 英, E-mail: ve5l30@163.com

摘 要: 针对无线传感网络中数据聚合能耗和重建误差问题, 提出了一种拓扑感知的数据聚合方法 (TADA). 首先, 构建了一个包含网络初始化、数据分帧和数据预处理的数据流, 形成无线传感网络的通信过程; 然后, 构造测量矩 阵将数据分解为多个路径转发, 从而进行全网络矢量分配, 并提出了基于平衡最小生成树是数据聚合算法. 通过实 验表明: 所提方法在数据聚合能耗和数据重建错误率要求上低于其它压缩感知的方法.

关键词: 拓扑感知; 无线传感网络; 数据聚合; 平衡最小生成树

引用格式: 熊英,李芸.拓扑感知的无线传感网络数据聚合方法.计算机系统应用,2021,30(9):256-261. http://www.c-s-a.org.cn/1003-3254/8129.html

Topology Aware Data Aggregation Method for Wireless Sensor Networks

XIONG Ying¹, LI Yun²

¹(Information Technology Department, Jiangmen Open University, Jiangmen 529234, China) ²(School of Water Resources and Environment Engineering, East China University of Technology, Nanchang 330013, China)

Abstract: We propose a Topology Aware Data Aggregation (TADA) method to address the problems of energy consumption and reconstruction errors in the data aggregation of a wireless sensor network. Firstly, a data stream including network initialization, data framing, and data preprocessing is constructed to form the communication process of the wireless sensor network. Secondly, a measurement matrix is built to decompose the data aggregation algorithm based on a balanced minimum spanning tree is proposed. Experiments show that the proposed method is lower than other compressed sensing methods in the energy consumption of data aggregation and the error rate of data reconstruction. **Key words**: topology awareness; wireless sensor network; data aggregation; balanced minimum spanning tree

随着无线通信技术的不断发展,对传输数据速率和频谱效率的要求不断提高^[1,2],在这些技术中,无线传感器网络 (Wireless Sensor Networks, WSN) 作为数据监测和数据获取的主要应用方式^[1],通过多跳路由从传感器发送到一个或多个数据接收器.同时,随着移动技术的进步,提出了在数据聚合过程中应用无人机的思

想^[2,3],由于传感器节点通常具有有限的计算能力和功 率储备,在数据聚合过程中利用无人机的主要目标是 以所需的精度收集数据,同时降低传感器的消耗^[4,5]. 压缩感知技术 (Compressive Sensing, CS)^[6] 对于有限资 源的传感节点,可以实现采集和压缩的功能,这是资源 有限的传感器节点的理想特性^[7],文献 [8] 在 CS 数据



① 基金项目: 江西省重点研发计划 (20202BBGL73086); 广东省自然科学基金 (2018A030313063); 江门市基础与理论科学研究项目 (2021C26); 广东远程 开放教育科研基金 (YJ2109)

Foundation item: Key Research and Development Program of Jiangxi Province (20202BBGL73086); Natural Science Foundation of Guangdong Province (2018A030313063); Basic and Theoretical Science Research Project of Jiangmen City (2021C26); Research Fund for Distance and Open Education of Guangdong Province (YJ2109)

收稿时间: 2021-01-16; 修改时间: 2021-02-07; 采用时间: 2021-02-23; csa 在线出版时间: 2021-09-02

²⁵⁶ 软件技术•算法 Software Technique•Algorithm

采集方案中,将 WSN 的 N 个传感器节点生成的原始 数据向量表示为 $x = \{x_1, x_2, \dots, x_N\}^T$,聚合过程建模为:

$$\mathbf{y} = \sum_{j=1}^{N} x_j \phi_j = \boldsymbol{\Phi} \mathbf{x} \tag{1}$$

式(1)中,每个原始数据x;由权重向量 q;编码,该权重 向量也是测量矩阵 $\boldsymbol{\phi}$ 的第 i 列向量, 测量矩阵 $\boldsymbol{\phi}$ 通常为 M×N随机矩阵, M 为权重向量数, N 为网络传感器节点 数,以下均为此表述.因此,聚合向量 y 小于未知数据的 数目, 当x 满足k-稀疏时且矩阵 $\phi^{M \times N}$ 满足限定等矩性^[7] (Restricted Isometry Property, RIP) 条件时, Sink 移动节 点可以从 $y = \{y_1, y_2, \dots, y_N\}^T$ 中恢复原始数据向量 x 的 精度. 文献 [8] 采用局部混合 CS 方法构建簇的大小与 传输次数之间的关系的模型,以找到最优簇的大小,以 此来提高数据聚合的效率,但由于无线传感网络拓扑 结构具有动态性, 当新增传感器时, 局部 CS 方法较难 快速构造出可行测量矩阵,网络向所有传感器节点广 播新的测量矩阵会增加能耗且数据重建误差增大;文 献 [9] 提出了全局 CS 的方法, 将感知数据进行加权计 算而非将单个原始数据传送到每个 Sink 节点, 有利于 数据向量的聚合,但由于全局 CS 方法中点到点传输的 次数一般大于传感器节点的最大传输单元,需要将每 个数据编码的权重向量分割成多个包,这样会使传输 的次数急剧增加,导致更高的能耗.

综上所述,为了解决上述基于 CS 方法的挑战,本 文提出了一种拓扑感知的无线传感网络数据聚合方案 (TADA). 主要是利用拓扑信息以高精度重建原始数据,构造测量矩阵对无线传感网络多路径转发,从而进行 全网络矢量分配,并提出了基于平衡最小生成树的数据 聚合算法,使数据聚合更能适应拓扑结构的动态变化.

1 问题描述

在数据聚合过程中,编码集 D 是一组权重向量,用 于对传感器节点的原始数据进行编码.在接收器端,该 过程可以表示为:

$$y = \boldsymbol{\Phi}^{M \times N} x_J, \ x_J \in R^{N \times 1}, y \in R^{M \times 1}$$
(2)

式(2)中, $x_J = \{x_1, x_2, \dots, x_N\}^T$ 表示路径 J的数据, $\Phi^{M \times N}$ 为映射矩阵, 它由 D 的权重向量构成(即列向量), R 为当前传感器节点. 当数据沿着从叶节点到汇聚节点的路径聚合时,每个节点将它的数据加权向量与从它的子节点接收的向量组合, 然后将组合后的向量转发给

相应的父节点;最后通过多跳方式,在接收端收集数据向量,利用路径 *J* 的路由信息对 x_J 进行解码.在式(2)中,测量信号 y_J = { y_1, y_2, \dots, y_N }^T为在接收器处收集的数据编码矢量.

2 通信过程

2.1 TADA 数据流

(1) 网络初始化: N个传感器节点在有限区域内随机均匀分布.设置一个平衡最小生成树 (Balanced Minimum Spanning Tree, BMST) 的拓扑构造算法 (将在第4小节中进一步阐述),在 BMST 形成拓扑结构后,每个节点沿着 BMST 给出的路由路径向 Sink 发送帧跳信号, Sink 收集所有的帧跳并相应地生成一个哈希表.如图 1 所示,首先,哈希表记录每个路径的路径索引号和路由序列,通过哈希表接收器获取最长的路由路径 l_{max} , Sink 生成一个权重向量数 $M \ge l_{\text{max}}$ 的向量编码集 $D = \{d_i\}$,其中 $d_i \in \mathbb{R}^{M \times 1}$;然后,按照此权重向量分配策略,从 D 中为每个节点分配一个权重向量,即为矢量分配的策略,将在 2.2 节中阐述.



(2)数据分帧:初始化完成后,每个节点通过将其 与指定的权重向量x_i×**Φ**^{M×1}(*i*)相乘,对其原始传感基 准x_i进行编码,并将其发送到其父节点.如果节点J从 节点*i*接收到包x_i×**Φ**^{M×1}(*i*),则将其编码的分组与接收 的分组线性结合,即x_i×**Φ**^{M×1}(*i*)+x_j×**Φ**^{M×1}(*j*).同时,如 图 2 所示,帧头记录已将其向量添加到线性组合中的 节点的索引号,如果节点上没有需要上传的数据,那么它 会转发接收到的数据包,帧头也不会记录其节点的索引号.

(3)数据预处理:当接收器接收到一个数据包时, 就会检查来自帧头的索引号,然后形成未知的原始数 据向量,最后运行数据检测算法.

2.2 数据检测

上所述过程中,每个节点在发送到其父节点之前

Software Technique•Algorithm 软件技术•算法 257

用权重向量 ϕ_j 对其自身的数据 x_j 进行编码, Sink 接收的数据包是所有数据编码向量沿路径的线性组合, 如式 (3) 所示:

$$y = \sum_{j=1}^{N} x_j \phi_j \tag{3}$$



图 2 描述了 TADA 中数据转发和收集处理的示例, 节点 *R*₁ 将其数据编码向量发送到其父节点 *R*₃. 然后, 节点 *R*₃ 将其自身的数据向量与从节点 *R*₁ 接收到的向量线性地组合成一个分组, 并将其转发到节点 *R*₄. 这些进程不断重复, 直到组合的数据到达接收器为止. 在 Sink 节点收集的数据向量为:

$$y = x_1\phi_1 + x_3\phi_3 + x_4\phi_4 + x_7\phi_7 + x_{10}\phi_{10}$$
(4)

原始数据向量的维数为 *N*, 即为网络中的传感器 节点数, 任何路径的原始数据向量只在路径上与这些 节点对应的位置具有非零值. 对于上述示例中的 *x*₁, 这 些节点是节点 *R*₁、*R*₃、*R*₄、*R*₇ 和 *R*₁₀. 根据矩阵, 收集 过程表示为:



- 3 向量分配策略
- 3.1 编码集 D 构造

从一个编码集 D 构造映射为矩阵 **Φ**^{M×N}, 然后进一

258 软件技术•算法 Software Technique•Algorithm

步解释网络中加入更多节点时如何扩展映射矩阵.对 编码集建模为:

$$\boldsymbol{\Phi}^{M \times N} \boldsymbol{x}^{N} = \boldsymbol{y}^{M}, \text{ s.t. } N \ge M$$
(6)

式中, N 为网络中的传感器节点数或原始数据向量维数, $\boldsymbol{\phi}^{M \times N}$ 是存储在 Sink 节点中用于数据恢复的映射矩阵. 当节点数 N 增加时, 映射矩阵也随之改变, 使得映射矩阵的构造成为关键问题. 本文所提出的一种利用编码集 D 构造 $\boldsymbol{\phi}^{M \times N}$ 的方法, 首先将数据聚合过程并行分解为多条路由路径的数据转发过程. 如图 3 所示, 描述了路径 1 由 { $x_1, x_3, x_4, x_7, x_{10}$ }组成, 借鉴文献 [10]的方法可分析为: 如果相应的权重向量 { $\varphi_1, \varphi_2, \varphi_3, \varphi_7, \varphi_{10}$ }为正交, 则 Sink是具有路径 1 的路由表知识, Sink 可以成功地检索 { $x_1, x_3, x_4, x_7, x_{10}$ }, 从而求解检测方程并得到唯一解.



基于上述分析,设计了一个正交向量集 $D = \{d_1, d_2, \dots, d_{l_{max}}\}, 其中<math>l_{max} = \max_i\{l_i\}, l_i$ 为路径 *i* 的路径长度, $d_j \in R^{M \times 1}, j \in \{1, 2, \dots, l_{max}\}$ 是相互正交.很明显,任何一 个路径,只要 $M > l_{max}$ 且所有 l_i 节点从集合中选择它们 的权重向量 D,接收器 Sink 可以成功检索原始数据.

3.2 全网矢量分配策略

在构造编码集 D 之后,下一步就需要将 D 扩散到 整个网络,利用哈希表进行矢量分配,由于 BMST 总会 生成一个树结构,因此从接收器到任何节点的路径都 是唯一的,向量分配过程由 Sink 节点执行.

如图 4 所示,前一示例的权重向量分配对于任何 路径 *i*, Sink 首先检查哈希表中的信息,然后 Sink 按照 节点存储在哈希表中的顺序逐个分配 *D* 的向量.路径 2 由 4 个传感器节点组成. Sink 首先从哈希表中查询 信息, 然后依次将权重向量 d₁, d₂, d₃, d₄ 分配给节点 R₁₀, 节点 R₇, 节点 R₆, 节点 R₅. 然后, 按照分配给传感 器节点的权重向量的顺序, 通过对齐来构建整个网络 的矩阵 **Φ**^{M×N}. 从而构造一个了策略矩阵, 得出最长路径 完全消耗了编码集 D 的权值向量, 由于每个路径权值 向量的正交性, 测量矩阵可以高精度地检索任意路径 的原始数据, 从而进行网络扩展.



4 基于 BMST 的数据聚合算法

上述策略矩阵的构造,是通过编码集 D 的基数 lmax 逐步扩展网络,而 lmax 则是整个网络中最长路径的 长度,为使扩展的网络达到平衡状态,采用平衡最小生 成树 (Balanced Minimum Spanning Tree, BMST)^[11]. 网 络中的节点 i 先不发送原始数据 d_i, 而是根据从编码 集 D 分配的权重向量对 d_i 进行编码, 并将长度 l_{max} 的 权重数据发送到其父节点.因此,每次传输的有效载荷 是 lmax, 而本文的主要目的是降低 lmax, 从而降低数据 聚合过程中的能耗. 在所有路径长度均为平衡的条件 下,在所提方法中考虑了树构造的邻居节点选择跳数. 树的构造过程从 Sink 节点开始, Sink 选择一个最优节 点,并在每个结构更新步骤中将其添加到网络中.此进 程终止,直到所有节点都在网络上.这里,最优节点是 指对平衡树贡献值最大的节点. 将树上未连接节点 k与节点 n 连接的贡献评估函数表示为 $C(D'(d_{n_k}))$, $J(h_n)$). 其中, D'是编码集,包含了原始数据 d_i 到 d_k 的 距离, k是节点 k到节点 n 的距离, $J \in h_n$ 的函数, h_n 是 连接节点 n 到 Sink 的跳数. 用 C 来表示 k 到 n 的连接 贡献,定义为:

$$C(d_{n,k},h_n) = \alpha D'(d_{n,k}) - J(h_n) \tag{7}$$

其中, α 是 D'(d) 的正系数, 控制距离对路径选择的影响, 选择附近的节点加入网络可以促进降低能耗. 因此,

函数 D'定义为:

$$D'(d) = 1/d^2$$
 (8)

式中,函数 J(h_n) 表示节点 n 到达接收器的跳数函数, 而跳数越多会导致拓扑不平衡,即对网络贡献的一种 惩罚.因此,为了让 J(h) 随着跳数 h 的增加而增长得更 快,加入系数 λ 为惩罚度,使函数 J(h) 定义为:

$$J(h_n) = e^{\lambda h} \tag{9}$$

BMST 生成方法如算法1所示.

算法 1. BMST 生成
1. 输入: $N / / 节点数量, T_t / / 拓扑表, F_t / / 未连接节点集合, h_j / / 从节点$
j 到 Sink 的单跳数量, p_j //父类节点索引
2. 搜索最近的节点 j 至 Sink 且生成一个路由表
3. $T_i = \{Node_j: \{h_j, p_j\}\};$
$F_t = \{ Node_1, \dots, Node_N \} / Node_j;$
4. Set $C_{\max} \leftarrow -\infty$
5. while <i>i</i> in <i>N</i> -1 do
6. Set $C_{\max} \leftarrow -\infty$
7. while $Node_k$ in F_t do
8. while $Node_n$ in T_t . Node do
9. Count $C(d_{n,k}, h_n)$;
10. if $C(d_{n,k}, h_n) \ge C_{\max}$ then
11. Set $C_{\max} \leftarrow C(d_{n,k}, h_n)$
12. Set $s \leftarrow k$
13. Set $h_s \leftarrow h_n + 1$
14. Set $p_s \leftarrow n$
15. end if
16. end while
17. end while
18. $T_t \leftarrow T_t \cup \{Node_s: \{h_s, p_s\}\}$
19. $F_t \leftarrow F_t / Node_s$
20. end while
21. 输出: <i>T_t</i>

5 实验结果分析

将本文所提方法与 RW 方法^[12] 和 ICS 方法^[13] 进 行比较,并从数据聚合能耗、数据重建恢复率和测量 矩阵存储要求 3 个方面对本文所提方法进行了评估.

5.1 数据聚合能耗

假设 N 个传感器节点随机均匀地分布在一个正方 形区域中, 准备将其读数发送到 Sink 节点. 这时, 将网 络通信能量消耗表示为 E, 表示与权重向量大小和传 输次数有关. 计算为:

 $E=O\left(\left[1+\frac{VectorSize}{MTU}\right] \times TransmissionNumber
ight)$ (10) 式中, *MTU* 表示由通信协议确定的最大传输单元, 如

Software Technique•Algorithm 软件技术•算法 259

果在所考虑的数据聚合网络中由 M 确定的向量大小 大于 MTU,则将每个加权向量分割成 [1+ VectorSize MTU] 个包. 让每一轮数据表示每个节点完成将其缓冲区的 第一次读取上载到 Sink 节点的时间段,传输数是指每 一轮数据聚合中的点对点传输的总数.

(1) 第一种情况下: 在 WiFi 环境下权重向量 *M* 的 远远小于 *MTU*, *E* 主要由发送的数目 *N_T* 控制. 如图 5 所示,本文所提方法与 ICS 和 RW 相比,本文方法比 RW 能量消耗更低,这是由于本文所提 TADA 方法构 造了一个编码集 *D* 映射为矩阵 *Φ^{M×N}*,进而扩散到整个 网络,相比 RW 方法降低了传输的次数,让每一轮数据 表示在一定正方形区域内. 但是本文方法与 ICS 相比,能量消耗仍然较高,这是由于在形成网络过程中,形成 了一个 BMST 树形结构,每一次通信传输均需要计算 到叶子节点,增加了每轮数据的传输计算次数和时间.



图 5 本文所提方法与 ICS、RW 在节点数的能耗比较

(2) 第二种情况下:考虑传感器节点正在利用轻量级通信协议 (如蓝牙或 ZigBee 网络环境下),其具有较小的 *MTU*,它需要将一个数据编码的权重向量分割为若干个数据包进行发送.考虑包括 300 个节点和 *M*_{ICS}=86 和 *M*_{TADA}=9 的场景,能耗比较如图 6 所示.由于权值向量较小,TADA 在能耗方面具有更好的性能优势.这是由于本文方法采用了最小平衡树,使网络达到平衡状态,用贡献平衡函数使 Sink 跳到最优节点,控制距离对路径选择的影响,达到聚合过程中的能耗.因此,当使用轻量级通信协议时,TADA 是首选的.

5.2 数据重建恢复率

由于数据聚合的目标是收集感测到的原始数据, 因此会检查不同协议的数据恢复率.在 CS 中,通过一 些贪婪的算法从 *M* 维数据向量 *v* 中恢复 *N* 维信号向量.

考虑一个 N=100 节点的 WSN 网络, 假设原始数

260 软件技术•算法 Software Technique•Algorithm

据 x 由 k ∈ {4, 5, 6, 7, 8, 9, 10}稀疏源产生, RW, ICS 和 TADA 的权重向量 *M*=20. 如图 7 所示,显示了具有 3 个协议的网络的错误率曲线.结果表明:无论稀疏度 *k* 有多大, TADA 网络的错误率都可以忽略不计,但 RW 和 ICS 的错误率都随着稀疏度 *k* 的增加而增加. 参考文献 [14] 所得出的证明, CS 随机矩阵成功恢复数 据的概率为 $P = 1 - 2e^{-\delta^2 M/8}$,其中 δ 是限定等距常数 (RIC),换言之, RW 和 ICS 的恢复误差是不可避免的. 这是由于本文方法逐步分配权重向量和策略矩阵,在 贡献平衡函数中选择控制距离,从而使将叶子节点到 父节点的每条路径由相应的权重向量子集构成正交矩 阵的子矩阵,以顽强满足数据恢复的要求.



图 6 本文所提方法与 ICS、RW 在 MUT 大小的能耗比较



图 7 本文所提方法与 ICS、RW 数据重建错误率比较

6 结论与展望

本文构造测量矩阵将数据分解为多个路径转发, 从而进行全网络矢量分配,并提出了基于平衡最小生 成树的数据聚合算法,缓解了无线传感网络中数据聚 合能耗和重建误差问题.但本文方法对数据的构造还 不够精确,在聚合过程中较难把握通信数据的语义方 向,在下步工作中需要进一步完善编码集 D 的构造和 矢量分配,以优化整个算法性能,进一步适应动态拓扑的变化.

参考文献

- 1 Maheshwari P, Sharma AK, Verma K. Energy efficient cluster based routing protocol for WSN using butterfly optimization algorithm and ant colony optimization. Ad Hoc Networks, 2021, 110: 102317. [doi: 10.1016/j.adhoc.2020. 102317]
- 2 Lin XS, Xia JJ, Wang Z. Probabilistic caching placement in UAV-assisted heterogeneous wireless networks. Physical Communication, 2019, 33: 54–61. [doi: 10.1016/j.phycom. 2019.01.004]
- 3 Ramli MR, Lee JM, Kim DS. Hybrid MAC protocol for UAV-assisted data gathering in a wireless sensor network. Internet of Things, 2021, 14: 100088. [doi: 10.1016/j. iot.2019.100088]
- 4 Luo CW, Chen WP, Li DY, et al. Optimizing flight trajectory of UAV for efficient data collection in wireless sensor networks. Theoretical Computer Science, 2021, 853: 25–42. [doi: 10.1016/j.tcs.2020.05.019]
- 5 Wang G, Lee B, Ahn J, *et al.* A UAV-assisted CH election framework for secure data collection in wireless sensor networks. Future Generation Computer Systems, 2020, 102: 152–162. [doi: 10.1016/j.future.2019.07.076]
- 6 Lv CC, Wang Q, Yan WJ, *et al.* Compressive sensing-based sequential data gathering in WSNs. Computer Networks, 2019, 154: 47–59. [doi: 10.1016/j.comnet.2019.03.004]
- 7 Candès EJ. The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique, 2008, 346(9–10): 589–592. [doi: 10.1016/

j.crma.2008.03.014]

- 8 Xie RT, Jia XH. Transmission-efficient clustering method for wireless sensor networks using compressive sensing. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(3): 806–815. [doi: 10.1109/TPDS.2013.90]
- 9 Lan KC, Wei MZ. A compressibility-based clustering algorithm for hierarchical compressive data gathering. IEEE Sensors Journal, 2017, 17(8): 2550–2562. [doi: 10.1109/JSEN. 2017.2669081]
- 10 Zhao N, Cheng F, Yu FR, et al. Caching UAV assisted secure transmission in hyper-dense networks based on interference alignment. IEEE Transactions on Communications, 2018, 66(5): 2281–2294. [doi: 10.1109/TCO MM.2018.2792014]
- 11 Mishra G, Mohanty SK. Efficient construction of an approximate similarity graph for minimum spanning tree based clustering. Applied Soft Computing, 2020, 97: 106676. [doi: 10.1016/j.asoc.2020.106676]
 - 12 Zheng HF, Yang F, Tian XH, *et al.* Data gathering with compressive sensing in wireless sensor networks: A random walk based approach. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(1): 35–44. [doi: 10.1109/TPDS. 2014.2308212]
 - 13 Wang J, Tang SJ, Yin BC, *et al.* Data gathering in wireless sensor networks through intelligent compressive sensing. Proceedings of IEEE INFOCOM. Orlando, FL, USA. 2012. 603–611.
 - 14 Laska JN, Davenport MA, Baraniuk RG. Exact signal recovery from sparsely corrupted measurements through the pursuit of justice. Proceedings of the 43rd Asilomar Conference on Signals, Systems and Computers. Pacific Grove, CA, USA. 2009. 1556–1560.

Software Technique•Algorithm 软件技术•算法 261