

# 基于压缩感知与矩阵补全技术的 WSN 数据收集算法

张策<sup>1</sup>, 李鸥<sup>1</sup>, 童昕<sup>2</sup>, 杨延平<sup>3</sup>

(1. 信息工程大学信息工程学院, 河南 郑州 450001; 2. 61377 部队, 广东 深圳 518000; 3. 清华大学电子系, 北京 100084)

**摘 要:** WSN 无线链路不可靠, 分组丢失现象普遍存在, 且基于压缩感知 (CS) 数据收集算法对分组丢失十分敏感。首先, 通过实验对分组丢失率和基于 CS 数据重构精度关系进行定量研究, 提出极稀疏块观测矩阵, 在降低每轮数据采集能耗的同时, 也保持观测矩阵的近似低秩性质。其次, 结合矩阵补全 (MC) 技术与 CS 技术, 提出基于极稀疏块观测矩阵的压缩感知数据收集算法, 在一个采集周期内进行数据收集, 利用 MC 技术恢复丢失数据, 减少分组丢失对数据收集的影响; 利用 CS 技术重构全网数据, 减少数据收集量, 降低节点在数据收集时所需能耗, 延长网络寿命。仿真分析表明, 所提算法在分组丢失率小于 50% 的情况下能够保证高精度重构全网数据, 抵抗不可靠链路。

**关键词:** 无线传感网; 数据收集; 压缩感知; 不可靠链路; 矩阵补全技术

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2018034

## WSN data gathering algorithm based on compressive sensing and matrix completion technique

ZHANG Ce<sup>1</sup>, LI Ou<sup>1</sup>, TONG Xin<sup>2</sup>, YANG Yanping<sup>3</sup>

1. School of Information Systems Engineering, PLA Information Engineering University, Zhengzhou 450001, China

2. 61377 Unit, Shenzhen 518000, China

3. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

**Abstract:** The unreliable links and packet losing are ubiquitous in WSN. The performance of data collection algorithm based on compressive sensing is sensitive to packet losing. Firstly, the relationship between packet loss rate and CS-based reconstruction precision was analyzed, and the sparsest block measurement (SBM) matrix was formulated to keep the data gathering consumption smallest and make sure the low-rank property of measurements. Then, combined with the matrix completion (MC) and compressive sensing (CS), the CS data gathering algorithm based on sparsest block measurement matrix (CS-SBM) algorithm was proposed. CS-SBM gathered data in a period and recovered the loss data based on MC to weaken the impact of packet loss on data gathering. CS-SBM reconstructed data based on CS to reduce measurement number and energy consumption and prolong the network lifetime. Simulation analysis indicates that the proposed algorithm reconstruct the whole data with high-accuracy under 50% packet loss rate, resisting unreliable links effectively.

**Key words:** WSN, data gathering, compressive sensing, unreliable link, matrix completion technique

### 1 引言

WSN 因传感器的低成本、小型化和低功耗等特点, 在森林火灾监测、战场侦察等应用场景中具有巨大的应用前景。但由于传感网监测环境复杂、

无线链路质量低、分组丢失率较高且单个节点能源受限等原因, 使 WSN 仍受到诸多限制。

近年来, 面向不可靠链路的传感网数据收集方法受到研究者的关注。文献[1]提出经典丢失数据补全算法——KNN (K-nearest-neighbor) 算法, KNN

收稿日期: 2017-10-12; 修回日期: 2018-01-17

通信作者: 童昕, tongrou@foxmail.com

基金项目: 国家科技重大专项基金资助项目 (No.2016zx03001010)

Foundation Item: The National Science and Technology Major Project of China (No.2016zx03001010)

简单地利用丢失数据的  $K$  个最近邻居数据, 估计丢失数据, 算法复杂度低但精度差。文献[2]中提出的 MSSA 算法利用嵌入式自协方差矩阵, 实现无参数和自适应数据恢复, 被广泛应用于恢复地理数据。这些算法在仅有少量数据丢失的情况下有效, 但在链路十分不可靠的传感网中, 分组丢失率较高, 此时, 这些传统的数据恢复算法性能较差。

随着压缩感知 (CS, compressive sensing) 理论<sup>[3,4]</sup>的提出, 一些研究者开始将 CS 应用于 WSN 数据收集, 利用原始数据的空间相关性, 有效地减少数据传输能耗, 且具有天然的能耗均衡特性, 能克服能量洞问题, 延长网络寿命; 基于压缩感知的数据收集算法具有数据重构过程复杂但压缩过程简单的特点、适合传感器节点低信息处理能力和能源受限的特点。文献[5]基于压缩感知, 提出了 CDG (compressive data gathering) 算法, 该算法利用数据的稀疏性和空间相关性, 通过减少数据收集量有效地降低数据传输能耗。文献[6]将 CS 技术与最小路径树路由协议相结合, 以降低整个数据传输路径上的能耗。文献[7]为了减少叶节点和距离叶节点较近的中间节点的通信量, 提出了混合压缩感知 (hybrid-CS) 数据收集方法, 仅对一部分通信量高的父节点使用压缩感知技术, 减少网络数据通信量。文献[8]结合 CS 技术与 Leach 分簇路由, 计算全网最优簇首个数, 使簇首均匀分布于网络中, 以减少网络能耗。

以上基于 CS 数据收集算法在链路可靠的情况下, 均可有效降低网络能耗, 且重构精度较高。但在实际传感网中, 由于节点部署环境复杂和节点收发数据功耗限制等因素, 导致无线链路质量低、网内分组丢失率高。在分布式压缩感知数据收集算法中, 一个分组的丢失, 会影响参与本次数据收集的所有节点, 导致重构数据精度较低, 重构数据无法使用。高分组丢失率使传统的基于 CS 的数据收集算法失效。文献[9]考虑了树型路由中不可靠链路对压缩感知的影响, 所提算法设定参与数据收集的节点比例, 在有分组丢失率的链路下, 基于收到的数据分组构造观测矩阵, 不观测发生分组丢失的节点, 抵抗不可靠链路。但在该算法中, 数据分组仍通过最短路由传输, 能耗不均衡现象仍然存在。文献[10]在簇型路由中,

针对轻负载下的链路不可靠, 建立随机分组丢失模型, 提出 CS-NTSC (neighbor topology spatial correlation prediction based CS data gathering) 算法, 利用邻居拓扑矩阵预测丢失数据, 减小错传的影响; 针对重负载下的链路不可靠, 建立节点伪失效模型, 并提出 CS-SSDG (sparse schedule for CS data gathering) 算法, 通过改变观测矩阵稀疏度, 避免观测出错数据, 弱化不可靠链路的影响。

矩阵补全 (MC, matrix completion) 技术是近几年新兴的一种技术, 利用矩阵的低秩特性, 恢复矩阵中丢失的数据, 将 MC 技术应用于 WSN 数据收集, 可以减少数据的采集量。文献[11]单一地利用 MC 技术压缩恢复数据, 减少网络通信量, 延长网络寿命; 将 MC 数据恢复过程的 NP-hard 问题转化为可解的、复杂度低的最优化问题。但该算法没有分析真实数据的近似低秩性, 且在树型路由下, 网络能耗不均情况仍然存在。文献[12]结合 CS 和 MC, 利用 MC 补全数据传输时丢失的数据, 利用数据的时间相关性进行 CS 数据压缩, 达到抵抗不可靠链路、减少节点传输能耗的目的。但当网络数据采集周期长、网内受同一事件影响时, 节点采集数据的时间相关性会很弱、空间相关性强, 在时间上进行 CS 数据压缩, 其恢复精度较低, 且算法时间复杂度高。每个节点在采集完所有时隙的数据后进行数据处理传输, 网络时延大。

为此, 本文在无线链路不可靠条件下, 对一个采集周期的数据, 利用 MC 技术恢复因链路不可靠丢失的数据; 利用节点的强空间相关性, 对节点数据进行 CS 压缩, 减少网络的数据传输量。本文主要贡献有以下 3 点。

1) 对一个时隙内的真实数据进行分析, 实验证明数据具有近似低秩的性质, 且收集一个时隙内数据, 以降低网络时延。

2) 利用数据的低秩性质与空间相关性, 将 MC 技术与 CS 技术相结合, 提出极稀疏块观测矩阵, 减小不可靠链路对数据重构的影响, 在保证重构精度的前提下降低网络能耗。

3) 采用 GreenOrbs 系统真实数据, 利用仿真软件 Matlab 对所提算法进行实验验证, 实验结果表明所提算法在链路具有 50% 分组丢失率情况下, 仍能保证高重构精度。

## 2 问题描述

### 2.1 基于压缩感知的分布式数据收集方法

网络采用簇型路由结构，整个网络划分为若干个簇，在每个簇内使用压缩感知技术进行数据收集，整个网络的拓扑结构示意图如图 1 所示。

假设簇内部署了  $N_1$  个节点，其观测数据为  $\mathbf{X}_{N_1 \times 1} = (x_1, \dots, x_{N_1})^T$ ， $\mathbf{X}$  在稀疏基  $\Psi_{N_1 \times N_1}$  下是稀疏的，即  $\mathbf{X} = \Psi \mathbf{d}$ ， $\mathbf{d}$  为  $\mathbf{X}$  在  $\Psi$  下的投影。随机观测矩阵为  $\Phi = (\phi_{ij})_{M \times N_1}$ ，其中， $M \ll N_1$ ，经过  $M$  次观测，观测向量  $\mathbf{Y} = (y_1, \dots, y_M)^T$  可以表示为

$$\mathbf{Y} = \Phi \mathbf{X} = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1N_1} \\ \vdots & \ddots & \vdots \\ \phi_{M1} & \cdots & \phi_{MN_1} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{N_1} \end{pmatrix} \quad (1)$$

sink 节点在收到观测向量  $\mathbf{Y}$  后，通过求解凸优化问题，求解全簇的观测数据。

$$\hat{\mathbf{X}} = \min \|\mathbf{X}\| \quad (2)$$

$$\text{s.t. } \mathbf{Y} = \Phi \mathbf{X} = \Phi \Psi \mathbf{d}$$

在数据传输过程中，由于复杂的监测环境、链路拥塞或冲突造成无线链路不可靠，存在一定的误码率与分组丢失率，对于误码率，当节点发生错传时，若接收节点通过信道译码无法恢复发生的错传位，接收节点将丢弃错传数据分组。本文假设节点一旦接收含错数据分组即丢弃，不进行纠错，因此，本文所假设的分组丢失率包含发生错传和丢失 2 种情况，节点未成功接收完整无错的数据分组即分组丢失。在簇型网络中，若节点  $i$  的数据分组  $x_i$  丢失，则簇首得到的观测向量为

$$\mathbf{Y}' = \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_M \end{pmatrix} = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1N_1} \\ \vdots & \ddots & \vdots \\ \phi_{M1} & \cdots & \phi_{MN_1} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ 0 \\ \vdots \\ x_{N_1} \end{pmatrix} \quad (3)$$

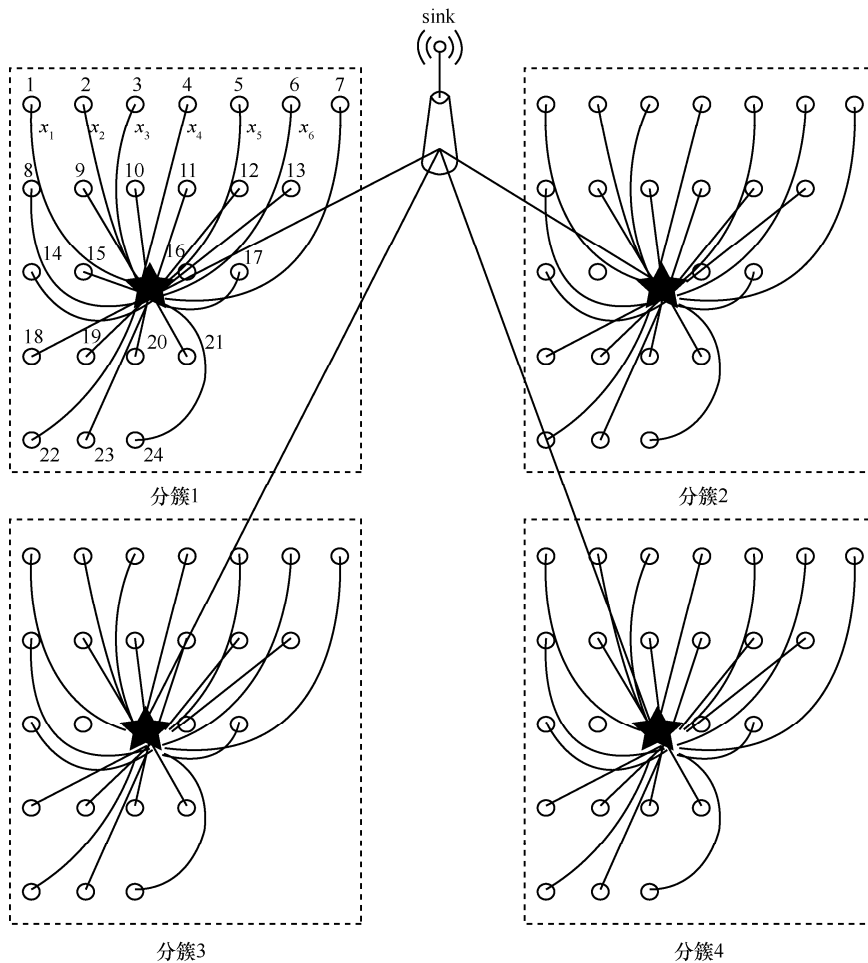


图 1 网络拓扑结构示意图

每一个观测向量  $(y_i)_{M \times 1}$  都会受到影响，这是由于 CS 过程中叠加各节点的采样数据造成的，分组丢失节点数量越多，对观测向量的影响就越大，且  $M$  个观测值都存在不同的误差，若使用观测向量  $\mathbf{Y}'$  进行数据重构，则会产生较大的误差，导致重构数据精度低，甚至无法使用。

本文用归一化平均绝对误差 (NMAE, normalized mean absolute error) 衡量数据重构误差，即

$$NMAE = \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_2}{\|\mathbf{X}\|_2} = \frac{\sqrt{\sum_{n=1}^{N_1} (\hat{x}_n - x_n)^2}}{\sqrt{\sum_{n=1}^{N_1} x_n^2}} \quad (4)$$

其中， $\hat{\mathbf{X}}$  表示重构数据， $\mathbf{X}$  表示原始数据。NMAE 越小表示重构误差越小，算法性能越好。

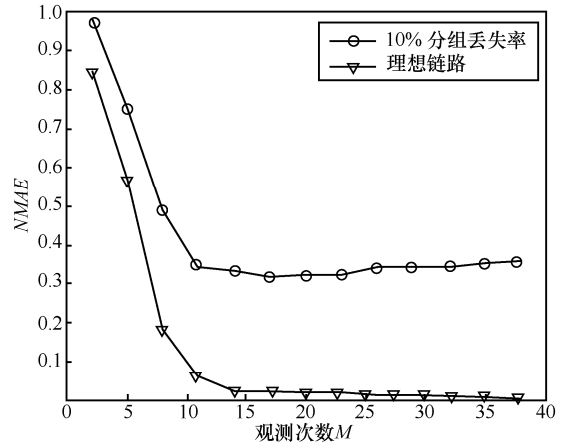
图 2 评估了理想链路和 10% 分组丢失率的有损链路下，经典 CDG 在簇型路由中的性能。由图 2(a) 可知，随着观测次数  $M$  的增加，理想链路下 NMAE 急剧下降，最终趋于稳定且小于 0.01；而在 10% 分组丢失率的有损链路下，随着  $M$  增加，虽然 NMAE 有所减小，但始终大于 0.3。图 2(b) 为  $M = 50$  时，分组丢失率和重构精度之间的关系。由图 2(b) 可知，随着分组丢失率的增加，NMAE 迅速增大，当分组丢失率为 10% 时，NMAE=0.36，基于 CS 的数据收集算法对分组丢失十分敏感，即使分组丢失率很低，也会造成基于 CS 数据收集算法精度变差，导致重构数据无法使用。因此，如何在高分组丢失率链路下，仍保持基于 CS 的数据收集算法精度亟待解决，同时也是本文的研究重点。

### 2.2 矩阵补全技术

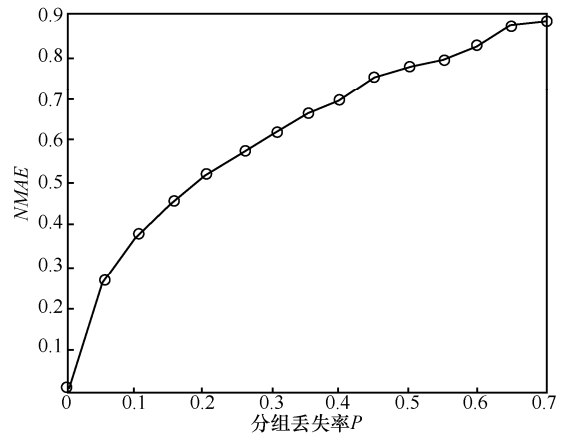
假设低秩矩阵  $\mathbf{B} \in \mathbb{R}^{n_1 \times n_2}$ ，可以通过它的  $s$  个部分元素恢复整个矩阵，且  $s \geq Cd^{\frac{6}{5}} r \lg(d)$ ，其中， $C$  为正常数， $d = \max(n_1, n_2)$ ， $r$  为矩阵  $\mathbf{B}$  的秩，其恢复过程可以表示为

$$\begin{aligned} \min \text{rank}(\mathbf{S}) \\ \text{s.t. } S_{ij} = B_{ij} \end{aligned} \quad (5)$$

其中， $\mathbf{S} \in \mathbb{R}^{n_1 \times n_2}$ ， $(i, j) \in \Omega$ ，指标集  $\Omega \subset \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ 。由于矩阵的秩函数是非连续、非凸的，求解式(5)是一个 NP-hard 问题，其求解复杂度呈指数级增长，最广泛的做法是将秩函数凸松弛到核



(a) 理想链路与有损链路下 CDG 性能对比



(b) 分组丢失率与 NMAE 关系

图 2 分组丢失率对 CS 数据重构精度影响

范数，转换为凸优化问题，即

$$\begin{aligned} \min \|\mathbf{S}\|_F \\ \text{s.t. } S_{ij} = B_{ij} \end{aligned} \quad (6)$$

其中， $\|\cdot\|_F$  表示 Frobenious 范数， $(i, j) \in \Omega$ 。

矩阵补全技术要求矩阵  $\mathbf{B}$  具有低秩的特性<sup>[13,14]</sup>，才能通过求解凸优化问题(6)被恢复。对任意的矩阵  $\mathbf{S} \in \mathbb{R}^{n_1 \times n_2}$  进行奇异值分解，即

$$\mathbf{S} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \sum_{i=1}^{\min(n_1, n_2)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (7)$$

其中， $\mathbf{U} \in \mathbb{R}^{n_1 \times n_1}$ ， $\mathbf{U} \mathbf{U}^T = \mathbf{U}^T \mathbf{U} = \mathbf{I}_{n_1 \times n_1}$ ， $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2}$ ， $\mathbf{V} \mathbf{V}^T = \mathbf{V}^T \mathbf{V} = \mathbf{I}_{n_2 \times n_2}$ 。 $\boldsymbol{\Sigma}$  是由奇异值组成的  $n_1 \times n_2$  阶对角矩阵，其元素按由大至小的顺序排列，即  $\sigma_i \geq \sigma_{i+1}$ ， $i = 1, \dots, \min(n_1, n_2)$ ，其中， $\min(n_1, n_2)$  表示矩阵  $\mathbf{S}$  的奇异值个数。若矩阵  $\mathbf{S}$  的秩  $r \ll \min(n_1, n_2)$ ，则称矩阵为低秩矩阵。由主成分分析 (PCA, principal component analysis) 理论可知<sup>[15]</sup>，

矩阵奇异值的大小反映了原始数据阵的主要内容，对于低秩矩阵<sup>[16]</sup>，前  $r$  个奇异值能反映矩阵的主要成分，即  $\sum_{i=1}^r \sigma_i \approx \sum_{i=1}^{\min(n_1, n_2)} \sigma_i$ 。本文定义函数  $g(i)$  来衡量矩阵是否为近似低秩，计算式为

$$g(i) = \frac{\sum_{j=1}^i \sigma_j}{\sum_{j=1}^{\min(n_1, n_2)} \sigma_j} \quad (8)$$

其中， $\sum_{j=1}^i \sigma_j$  为矩阵中前  $i$  个最大奇异值。本文选取 GreenOrbs 系统 5 000 个传感器采集的真实数据，将数据表示为  $100 \times 50$  矩阵形式，分析数据矩阵的低秩性。图 3 反映了该真实数据矩阵的奇异值分布情况， $x$  轴表示前  $i$  个奇异值， $y$  轴表示函数  $g(i)$ 。由图 3 可知，对于温度数据，前 4% 的奇异值占奇异值总和的 91.35%，即  $\sum_{i=1}^2 \sigma_i \approx \sum_{i=1}^{50} \sigma_i$ ；对于湿度数据，前 6% 的奇异值占奇异值总和的 93.64%，即  $\sum_{i=1}^3 \sigma_i \approx \sum_{i=1}^{50} \sigma_i$ 。

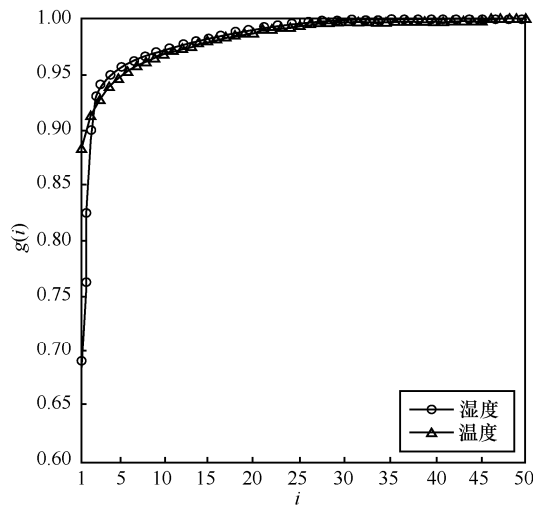


图 3 矩阵低秩特性

实验结果表明，真实环境中传感器感知原始数据具有近似低秩的特征，若节点在数据传输的过程中丢失有一定数量的分组，即无线链路存在分组丢失率，可以利用原始数据矩阵的低秩性质，通过 MC 技术，恢复出丢失的数据分组，减少分组丢失对数据传输的影响。

由于传感网采集的数据阵为近似低秩矩阵，令

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{L}\mathbf{R}^T \quad (9)$$

其中， $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}$ ， $\mathbf{R} = \mathbf{V}\mathbf{\Sigma}^{\frac{1}{2}}$ ， $\mathbf{L} \in \mathbb{R}^{n_1 \times k}$ ， $\mathbf{R} \in \mathbb{R}^{n_2 \times k}$ ，则可通过求解如下最优化问题代替式(6)，以获取更高的数据恢复精度。

$$\min \|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2 + \|\mathbf{L}\mathbf{R}^T - \mathbf{B}\|_F^2 \quad (10)$$

本文采用 EDCA (efficient data collection approach) 算法<sup>[11]</sup>求解式(10)。

### 3 算法设计

针对无线链路有分组丢失情况，提出基于极稀疏块观测矩阵的压缩感知数据收集算法 (CS-SBM, CS data gathering algorithm based on sparsest block measurement matrix)，结合 MC 技术与 CS 技术，在不重传的前提下，对一个采集周期数据进行数据收集，利用 MC 技术恢复丢失数据，减少分组丢失对数据收集的影响，利用 CS 技术重构全网数据，减少数据收集量，降低节点在数据收集时所需能耗，延长网络寿命。

#### 3.1 极稀疏块观测矩阵

在簇内，成员节点根据存储的稀疏观测向量元素是否为 0 来判断是否参与本轮数据收集。在数据传输过程中若有分组丢失，则簇首收到的数据并不完整，此时可以利用 MC 技术恢复丢失数据。

簇首收到的数据矩阵具有低秩或近似低秩的性质，是 MC 技术恢复丢失数据矩阵的充分条件。由 2.2 节可知，全网数据  $\bar{\mathbf{X}}$  具有近似低秩的性质，若使用观测向量  $\Phi$  对  $\bar{\mathbf{X}}$  进行随机观测后传输，观测向量  $\mathbf{Y} = \Phi\bar{\mathbf{X}}$  表示为矩阵形式后，不具备低秩或近似低秩的性质，则无法通过 MC 技术对丢失后的观测值  $\mathbf{Y}'$  进行恢复。本文构造极稀疏块观测 (SBM, sparsest block measurement) 矩阵，对原始数据向量  $\bar{\mathbf{X}}$  进行观测，该矩阵中每一行有且仅有一个非零值 1，即每轮数据传输只有一个节点需要发送数据，此时每轮数据收集能耗最低，且观测向量表示为矩阵形式后，仍具有近似低秩的性质。

假设数据向量  $\bar{\mathbf{X}} = (x_1, \dots, x_{N_i})^T$ ，表示为矩阵形式后为

$$\mathbf{X}_{mn} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ & x_{ij} & \cdots & x_{i(j+q-1)} \\ \vdots & \vdots & \ddots & \vdots \\ & x_{(i+p-1)j} & \cdots & x_{(i+p-1)(j+q-1)} \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}_{m \times n} \quad (11)$$

其块矩阵  $\mathbf{X}' \in \mathbb{R}^{p \times q}$  为

$$\mathbf{X}'_{p \times q} = \begin{pmatrix} x_{ij} & \cdots & x_{i(j+q-1)} \\ \vdots & \ddots & \vdots \\ x_{(i+p-1)j} & \cdots & x_{(i+p-1)(j+q-1)} \end{pmatrix}_{p \times q} \quad (12)$$

由图 3 可知,  $\mathbf{X}$  具有近似低秩的性质, 其块矩阵  $\mathbf{X}'$  也具有低秩性, 构造极稀疏块观测矩阵, 使  $\mathbf{X}'$  为观测向量  $\mathbf{Y}$  的矩阵形式, 即  $\mathbf{Y} = \bar{\mathbf{X}}' \in \mathbb{R}^{pq \times 1}$ , 保持了观测矩阵的低秩性质, 则可利用 MC 技术对矩阵  $\mathbf{X}'$  进行恢复。因此, 定义 SBM 矩阵  $\Phi_E \in \mathbb{R}^{M \times N_1}$  为

$$\Phi_E(h, g) = \begin{cases} 1, & g = \Omega_h \\ 0, & \text{其他} \end{cases} \quad (13)$$

其中,  $\Omega_h$  为块矩阵序号索引, 其定义为

$$\Omega_h = (i-1+k) \times n + (j+l) \quad (14)$$

其中,  $\Omega_h < \Omega_{h+1}$ ,  $p, q$  为块矩阵  $\mathbf{X}'$  的维数,  $i, j$  为其第一个元素位置, 观测次数  $M = p \times q$ ,  $k=0, \dots, p-1, l=0, \dots, q-1$ , 且  $j+q-1 \leq n, i+p-1 \leq m$ 。由定义可知,  $\Phi_E$  每一行只有一个非零元素, 且非零元素的位置由序号索引  $\Omega_h$  确定。若已知  $p, q$  与  $i, j$ , 则  $\Omega_h$  被唯一确定。一轮数据收集过程如式(15)所示。

$$\bar{\mathbf{Y}} = \bar{\mathbf{X}}' = \begin{pmatrix} x_{ij} \\ x_{i(j+1)} \\ \vdots \\ x_{i(j+q-1)} \\ x_{(i+1)j} \\ x_{(i+1)(j+1)} \\ \vdots \\ x_{(i+1)(j+q-1)} \\ \vdots \\ x_{(i+p-1)(j+q-1)} \end{pmatrix} = (\phi^E, \dots, \phi_M^E) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_1} \end{pmatrix} \quad (15)$$

其中,  $\phi^E$  表示  $\Phi_E$  的第  $i$  列。SBM 矩阵使每轮数据收集仅有一个节点参与, 降低了数据采集能耗, 且观测矩阵为全网原始矩阵的块矩阵, 保持了矩阵的低秩性。

在数据传输的过程中, 簇内子节点根据自身存储的观测向量, 判断是否发送感知数据至簇首。此时, 若无线链路可靠, 则 sink 收集到本簇观测向量为  $\bar{\mathbf{X}}'_{pq \times 1}$ , 矩阵形式为  $\mathbf{X}'_{pq \times q}$ ; 若链路不可靠, 存在一定的分组丢失率, 则 sink 收集到本簇观测向量为  $\bar{\mathbf{X}}'_{pq \times 1}$  分组丢失后的向量, 其矩阵形式为

$$\mathbf{Y}_R = \mathbf{X}'_L = \begin{pmatrix} x_{ij} & ? & \cdots & x_{i(j+q-1)} \\ ? & x_{(i+1)(j+1)} & \cdots & x_{(i+1)(j+q-1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(i+p-1)j} & x_{(i+p-1)(j+1)} & ? & x_{(i+p-1)(j+q-1)} \end{pmatrix}_{p \times q} \quad (16)$$

其中, ? 表示传输丢失的数据。由于  $\mathbf{X}'$  具有近似低秩的性质, sink 可以利用 MC 技术恢复出完整的  $\hat{\mathbf{Y}}_R$ , 并通过求解式(2)最优化问题, 重构出全网数据  $\hat{\mathbf{X}}$ , 最大程度减小分组丢失对数据重构的影响。算法示意如图 4 所示。

假设一个簇内有  $N_1 = 12$  个成员节点, 其数据向量  $\bar{\mathbf{X}}_{12 \times 1} = (x_1, \dots, x_{12})^T$ , 可表示为  $3 \times 4$  矩阵形式, 即

$$\mathbf{X}_{3 \times 4} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{pmatrix} \quad (17)$$

由 2.2 节可知, 矩阵  $\mathbf{X}_{3 \times 4}$  具有近似低秩性质。若选取其块矩阵  $\mathbf{X}'_{2 \times 2}$  为观测矩阵, 则有

$$\mathbf{X}'_{2 \times 2} = \begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} \quad (18)$$

即  $p=2, q=2, M=p \times q=4, m=3, n=4, i=2, j=2, N_1=12, \mathbf{X}'_{2 \times 2}$  同样具有近似低秩的性质。由式(14)可得, 块矩阵检索索引  $\Omega_4 = \{6, 7, 10, 11\}$ , 基

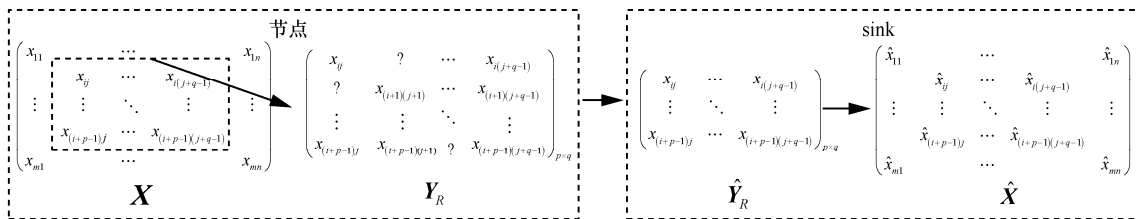


图 4 算法示意

于  $\Omega_4$  构造极稀疏块观测矩阵  $\Phi_E$  为

$$\Phi_E = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (19)$$

则一轮数据收集过后观测向量  $\bar{Y}$  为

$$\bar{Y} = \Phi_E \bar{X}_{12 \times 1} = \bar{X}'_{2 \times 2} = (x_{22} \quad x_{23} \quad x_{32} \quad x_{33})^T \quad (20)$$

其簇首收到的观测向量的矩阵为

$$Y = X'_{2 \times 2} = \begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} \quad (21)$$

若链路存在一定的分组丢失率,  $x_{23}$  数据丢失, 则簇首收到的观测矩阵为

$$Y_R = \begin{pmatrix} x_{22} & ? \\ x_{32} & x_{33} \end{pmatrix} \quad (22)$$

簇首根据  $Y_R$  与  $\Omega_4$  建立分组丢失索引  $\Theta$ , 发送至 sink, sink 可根据 CS-SBM 算法重构出全网数据。

### 3.2 CS-SBM 算法

CS-SBM 算法流程如图 5 所示, 具体步骤如下。

1) 簇首根据本簇节点数及节点 ID, 确定参数  $p$ 、 $q$ 、 $i$ 、 $j$ , 建立块矩阵序号索引  $\Omega_h$ , 生成 SBM 矩阵  $\Phi_{M \times N} = (\phi^E, \dots, \phi_{N_1}^E)$ , 如算法 1 所示, 该算法复杂度为  $O(M)$  且为线性运算, 适合计算能力受限的普通节点, 具有实用性。在簇内实行簇首轮换, 由簇首选取簇内剩余能量最大的节点作为下一轮簇首, 并将新簇首信息发送至成员节点, 簇首轮换可以有效地均衡网络能耗。

**算法 1** 生成 SBM 矩阵  $\Phi_E$

输入  $i$ 、 $j$ 、 $p$ 、 $q$ 、 $n$ 、 $M = p \times q$

输出  $\Phi_E$

①  $\Phi_E = 0, \Omega_h = 0$ ;

②  $h = 1$ ;

③ for  $k = 0 : 1 : p - 1$

④ for  $l = 0 : 1 : q - 1$

⑤  $\Omega_h = (i - 1 + k) \times n + (j + l)$ ;

⑥  $h = h + 1$ ;

⑦ end for

⑧ end for

⑨ for  $h = 1 : 1 : M$

⑩  $g = \Omega_h$ ;

⑪  $\Phi_E(h, g) = 1$ ;

⑫ end for

⑬ return  $\Phi^E$  并将其发送到  $i^{\text{th}}$  节点

2) 簇首将观测向量  $\phi_i, i \in \{1, \dots, N_1\}$  分发至每个节点, 节点根据收到的观测向量决定本轮数据收集是否发送观测数据  $X'$  至簇首, 若不存在分组丢失率, 则簇首接收到观测数据  $Y_{M \times 1}$  为

$$Y = X' = \Phi X \quad (23)$$

由于存在分组丢失率  $P$ , 簇首收到  $M_1$  个原始数据  $Y_R$ , 且  $M_1 \leq M$ 。簇首根据收到的  $M_1$  个数据建立分组丢失索引  $\Theta$ , 用来存储分组丢失节点的节点号, 并将  $Y_R$ 、 $\Theta$  和  $\Omega_h$  发送至 sink 节点。

3) sink 根据 MC 理论, 求解式(10)凸优化问题, 恢复出无丢失观测数据  $\hat{Y}_R$ 。根据  $\Omega_h$  重新生成  $\Phi_E$ , 并利用 CS 技术求解式(2)凸优化问题, 重构出全网原始数据  $\hat{X}$ , 具体算法如算法 2 所示, 算法 2 需要求解 2 个最优化问题, 复杂度高, 但该算法是在有固定能源且计算能力强的 sink 节点处实行; 对于能源与计算能力均受限的普通节点, 仅运行复杂度为  $O(M)$  的线性算法 1 即可, 满足 WSN 的结构特点, 具有实用性与有效性。

**算法 2** 重构全网数据

输入 SBM 矩阵  $\Phi_E$ , 发送数据  $X'$ , 接收数据  $Y_R$ , 分组丢失索引  $\Theta$ , 块矩阵序号索引  $\Omega_h$

输出 全网原数据  $\hat{X}$

① 簇首将  $\Phi_E$  发送至成员节点;

② 成员节点将  $X'$  发送至簇首, 由于分组丢失, 簇首收到数据为  $Y_R$ ;

③ 簇首基于  $Y_R$  和  $\Omega_h$  构造分组丢失索引  $\Theta$ ;

④  $\hat{Y}_R = \min \|Y_R\|_F, \text{ s.t. } Y_R = X'$ ;

⑤  $\hat{V} = \arg \min |V|_0, \text{ s.t. } \hat{Y}_R = \Phi_E X = \Phi_E \Psi V$ ;

⑥  $\hat{X} = \Psi \hat{V}$ ;

⑦ return  $\hat{X}$

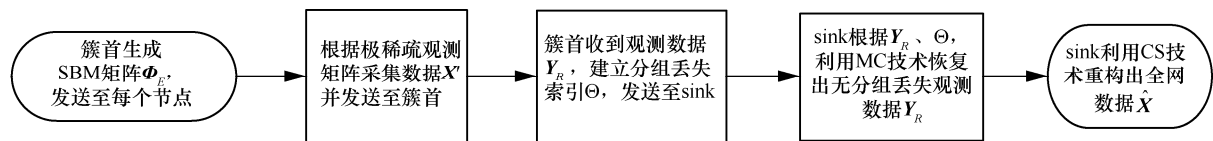


图 5 CS-SBM 算法流程

### 4 仿真与分析

为了验证算法的有效性，在 Matlab 平台下进行仿真分析，仿真环境设置如下：在一个簇内，簇成员节点通过单跳将数据发送至簇首，由簇首进行数据压缩，将观测矩阵发送至 sink 节点，在 sink 处进行数据恢复与重构，发送数据分组长度为 10 B；设簇内有 900 个节点分布在  $30 \times 30$  的范围内，采用 GreenOrbs 系统真实数据，以验证算法的实用性。本文分别采用 EDCA 算法和正交匹配追踪 (OMP, orthogonal matching pursuit) 算法作为数据恢复算法和数据重构算法，采用 FFT 基作为压缩采样稀疏基。

本文假设无线链路不可靠，存在一定分组丢失率  $P$ ，且簇首未收到成员节点的数据分组时不重传，将最终数据重构精度作为算法性能指标，如式(4)所示。本文选取 CDG 算法<sup>[5]</sup>、SRS (sparsest random scheduling based CDG scheme) 算法<sup>[9]</sup>、CS-NTSC 算法和 CS-SSDG 算法<sup>[10]</sup>作为对比算法。传统 CS 数据收集算法 CDG 利用压缩感知技术收集并重构数据，且不考虑分组丢失率对算法的影响；SRS 算法考虑到链路中存在分组丢失，利用实际收到的数据分组，构造极稀疏矩阵观测无分组丢失数据，减小分组丢失率对数据重构的影响；CS-NTSC 算法构造邻居拓扑矩阵，利用数据间的空间相关性，估计丢失数据，减小错误幅度，抵抗不可靠链路；CS-SSDG 根据丢失节点，改变观测矩阵，避免收集发生失效的节点数据，以减小错传或分组丢失对整个数据收集和重构的影响。

图 6 为观测次数  $M = 625$  时，CS-SBM、SRS、CDG、CS-SSDG 和 CS-NTSC 这 5 种算法在不同分组丢失率时的性能对比，参数设置如表 1 所示。

由图 6 可知，1) 当观测次数一定时，随着分组丢失率的增加，CS-SBM 算法与 CDG 算法性能变差，SRS、CS-SSDG 和 CS-NTSC 算法性能变化平缓；2) 当分组丢失率  $P \leq 0.5$  时，CS-SBM 算法优于其他算法；当  $P > 0.5$  时，CS-SSDG 算法性能最优。由于当无线链路质量较好时，SRS 算法基于接收到的数据构造极稀疏观测矩阵，其观测次数为  $M \times P < 625$ ，CS-SSDG 算法改变观测矩阵，减少观测量，CS-NTSC 算法利用空间相关性预测丢失数据，这些算法在一定程度上虽然都

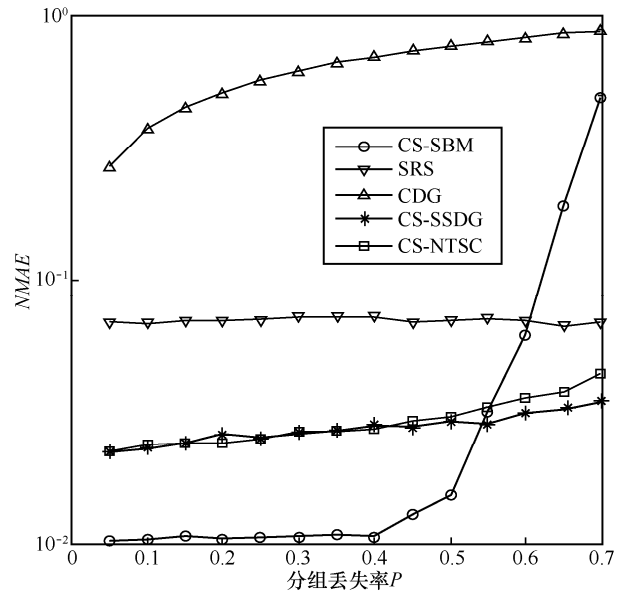


图 6 CS-SBM 算法性能 ( $M=625$ )

表 1 参数设置

参数	描述	值
$i$	块矩阵首元素所在行	2
$j$	块矩阵首元素所在列	2
$p$	块矩阵行数	25
$q$	块矩阵列数	25
$M$	观测次数	625
$n$	全网数据矩阵列数	30

能抵抗不可靠链路，但在压缩感知数据重构时精度低；而 CS-SBM 算法利用 MC 技术恢复丢失数据，在不同的分组丢失率下，其观测次数均为 625，数据重构精度较高。当无线链路质量较差时，对 CS-SBM 算法，信息丢失严重，sink 接收到的数据分组数量，不足以利用 MC 技术高精度恢复出丢失数据，使恢复矩阵  $\hat{Y}$  精度差，导致最终 CS 数据重构精度差。

图 7 为恢复矩阵  $\hat{Y}$  的恢复精度与分组丢失率的关系。由图 7 可知，随着分组丢失率的增加， $\hat{Y}$  的恢复精度逐渐降低。当分组丢失率  $P = 0.3$  时，CS-SBM 算法  $NMAE=0.0104$ ，而 SRS、CDG、CS-SSDG 和 CS-NTSC 算法的  $NMAE$  分别为 0.073、0.621、0.027 和 0.026，此时 MC 数据恢复  $NMAE = 0.026$ ；当分组丢失率  $P = 0.65$  时，CS-SBM、SRS、CDG、CS-SSDG 和 CS-NTSC 这 5 种算法的  $NMAE$  分别为 0.188、0.066、0.861、0.032 和 0.038，此时，MC 数据恢复  $NMAE = 0.252$ 。因此，当  $P < 0.5$  时，CS-SBM 算

法能够以高精度重构出全网数据，具有有效性；当  $P > 0.5$  时，链路质量很差，CS-SBM 算法不再适用。

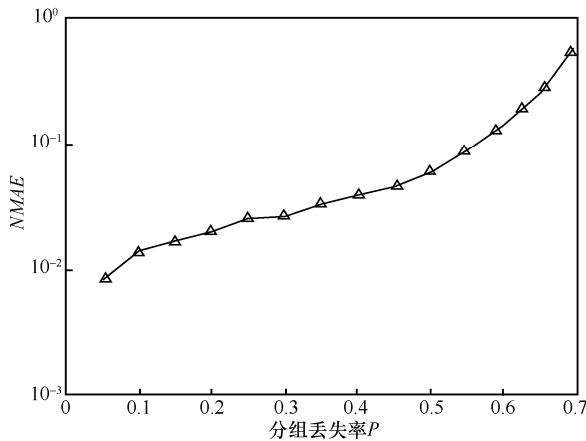


图 7 MC 恢复精度与分组丢失率关系

不同观测次数下的算法性能如图 8 所示。由图 8 可知，1)对不同的观测次数，当分组丢失率  $P \leq 0.5$  时， $NMAE$  变化平缓，CS-SBM 算法可以有效恢复丢失数据，减少分组丢失对数据重构精度的影响；当  $P > 0.5$  时， $NMAE$  迅速增大，算法性能变差；2)当分组丢失率一定时，随着观测次数  $M$  的增加，CS-SBM 算法的重构误差  $NMAE$  减小。

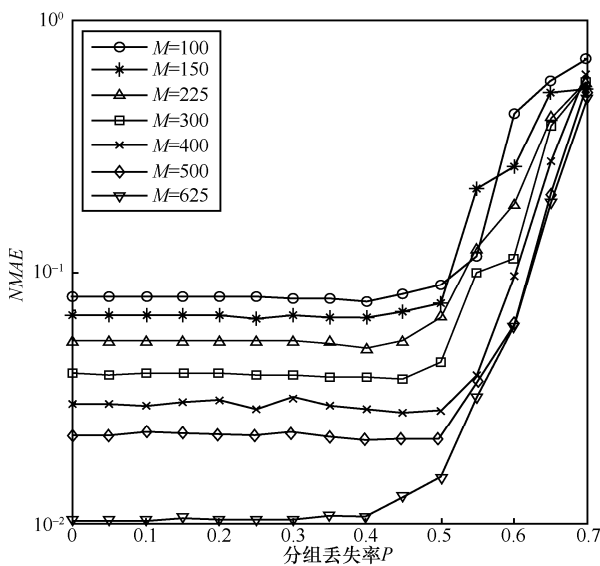


图 8 不同观测次数下算法性能

图 9 为  $P = 0.4$  时， $M$  与重构精度  $NMAE$  的关系，当  $M$  分别为 225、400 和 625 时， $NMAE$  分别为 0.495、0.283 和 0.014。即使在高分组丢失率下，CS-SBM 算法依然能够通过增加观测次数，提高重构精度，抵抗不可靠链路。

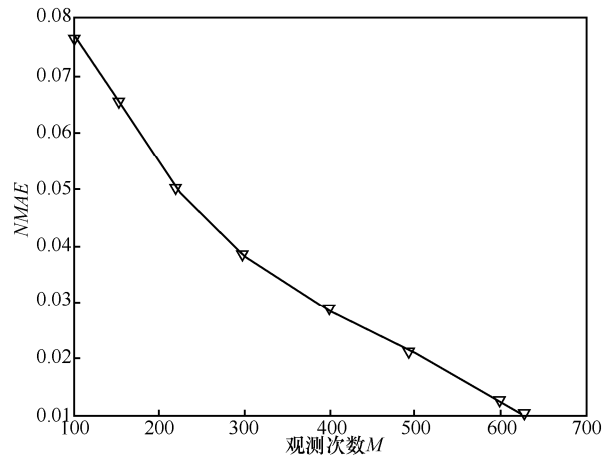


图 9 观测次数与重构精度的关系 ( $P=0.4$ )

### 5 结束语

本文针对无线传感网链路不可靠时分组丢失会使压缩感知数据重构精度变差的问题，提出将 MC 技术与 CS 技术相结合用于数据收集。首先，分析了分组丢失对数据重构精度的影响，链路分组丢失率越高，恢复精度越差。其次，对 GreenOrbs 系统的真实传感器数据进行实验分析，证明其具有近似低秩的性质，满足 MC 技术使用条件。最后，构造 SBM 矩阵，使观测矩阵为原始数据矩阵的块矩阵，仍具有近似低秩的性质，并结合 MC 技术与 CS 技术，提出了 CS-SBM 算法，对有分组丢失的观测矩阵先进行数据恢复，恢复出丢失数据，再进行全网数据重构。仿真结果表明，当  $P \leq 0.5$  时，CS-SBM 算法能够保持高重构精度，抵抗分组丢失对数据重构的影响，且可以通过增加观测次数  $M$  提高重构精度，具有有效性；当  $P > 0.5$  时，CS-SBM 算法性能变差，不再适用。

本文构造了 SBM 矩阵，并采用 FFT 基作为稀疏基，通过仿真实验可得，CS-SBM 算法可以通过 CS 技术重构出全网数据，但并未理论证明 SBM 矩阵和 FFT 基满足 RIP 条件，如何从数学角度严谨地证明所提 SBM 矩阵满足 RIP 条件是下一步工作重点。

### 参考文献:

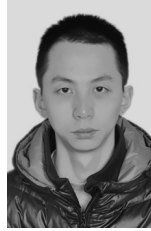
- [1] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [2] ZHU H, ZHU Y, LI M, et al. SEER: metropolitan-scale traffic perception based on lossy sensory data[C]//IEEE INFOCOM 2009. 2009: 217-225.
- [3] DONOHO D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.
- [4] BARANIUK R. Compressive sensing[J]. IEEE Signal Processing

Magazine, 2007, 56(4): 4-5.

- [5] LUO C, WU F, SUN J, et al. Compressive data gathering for large-scale wireless sensor networks[C]//The 15th Annual Int Conf on Mobile Computing and Networking. 2009: 145-156.
- [6] LUO C, WU F, SUN J, et al. Efficient measurement generation and pervasive sparsity for compressive data gathering[J]. IEEE Transactions on Wireless Communications, 2010, 9(12): 3728-3738.
- [7] LUO J, XIANG L, ROSENBERG C. Does compressed sensing improve the throughput of wireless sensor networks?[C]//IEEE International Conference on Communications (ICC 2010). 2010: 1-6.
- [8] WU X, XIONG Y, HUANG W, et al. An efficient compressive data gathering routing scheme for large-scale wireless sensor networks[J]. Computers and Electrical Engineering, 2013, 39(6): 1935-1946.
- [9] WU X, YANG P, JUNG T, et al. Compressive sensing meets unreliable link: sparsest random scheduling for compressive data gathering in lossy WSN[C]//The 15th ACM Int Symposium on Mobile Ad Hoc Networking and Computing. 2014: 13-22.
- [10] 张策, 张霞, 李鸥, 等. 不可靠链路下基于压缩感知的 WSN 数据收集算法[J]. 通信学报, 2016, 37(9): 131-141.  
ZHANG C, ZHANG X, LI O, et al. Compressive sensing based data gathering algorithm over unreliable links in WSN[J]. Journal on Communications, 2016, 37(9): 131-141.
- [11] CHENG J, JIANG H, MA X, et al. Efficient data collection with sampling in WSNs: making use of matrix completion techniques[C]//Global Telecommunications Conference. 2010: 1-5.
- [12] FRAGKIADAKIS A, ASKOXYLAKIS I, TRAGOS E. Joint compressed-sensing and matrix-completion for efficient data collection in WSNs[C]//International Workshop on Computer Aided Modeling and Design of Communication Links and Networks. 2014: 84-88.
- [13] CANDES E, RWCHT B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.
- [14] CHENG J, YE Q, JIANG H, et al. STCDG: an efficient data gathering algorithm based on matrix completion for wireless sensor networks[J]. IEEE Transactions on Wireless Communications, 2013, 12(2): 850-861.
- [15] LAKHINA A, PAPAGIANNAKI K, CROVELLA M, et al. Structural analysis of network traffic flows[J]. ACM SIGMETRICS Performance Evaluation Review, 2004, 32(1): 61-72.
- [16] ZHANG Y, ROUGHAN M, WILLINGER W, et al. Spatio-temporal compressive sensing and internet traffic matrices[C]//ACM SIG

COMM 2009 Conference on Data Communication. 2009: 267-278.

#### [作者简介]



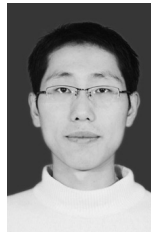
张策 (1991-), 男, 四川南充人, 信息工程大学博士生, 主要研究方向为无线自组织网络、无线传感网与路由协议。



李鸥 (1961-), 男, 陕西宝鸡人, 博士, 信息工程大学教授、博士生导师, 主要研究方向为无线传感网、认知无线网络与无线自组织网络。



童昕 (1990-), 女, 湖北黄冈人, 61377 部队助理工程师, 主要研究方向为多天线信号联合处理。



杨延平 (1986-), 男, 河南平顶山人, 清华大学博士后, 主要研究方向为无线认知通信、网络编码、自适应调制。