

# 一种基于矩阵补全的 无线传感网数据收集方案

李国瑞<sup>1</sup>, 王 颖<sup>2</sup>, 王 聪<sup>1</sup>

(1. 东北大学计算机科学与工程学院, 辽宁沈阳 110819; 2. 秦皇岛职业技术学院信息工程系, 河北秦皇岛 066100)

**摘 要:** 针对无线传感器网络中的数据收集问题, 设计了一种基于矩阵补全的数据收集方案. 首先利用低秩矩阵分解模型将传感器节点的数据收集问题转换为两个相互关联的凸优化子问题, 然后利用基于改进 Kaczmarz 迭代的矩阵补全算法循环交替求解凸优化子问题. 在 Kaczmarz 迭代过程中, 依据 Johnson-Lindenstrauss 引理对投影操作降维, 从而提高了矩阵补全算法的计算速度. 实验结果表明与现有矩阵补全算法相比, 文中所设计的矩阵补全算法在重构精度、成功重构概率以及重构时间等方面具有较优的重构性能.

**关键词:** 无线传感器网络; 数据收集; 矩阵补全; 优化; 数据重构

**中图分类号:** TP311 **文献标识码:** A **文章编号:** 0372-2112 (2018)12-2950-07

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.12.018

## A Matrix Completion Based Data Collection Scheme in Wireless Sensor Networks

LI Guo-rui<sup>1</sup>, WANG Ying<sup>2</sup>, WANG Cong<sup>1</sup>

(1. School of Computer Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China;

2. Department of Information Engineering, Qinhuangdao Institute of Technology, Qinhuangdao, Hebei 066100, China)

**Abstract:** Considering the data collection problem in wireless sensor networks, a matrix completion based data collection scheme is proposed. By transforming the data collection problem into two correlated convex sub-optimal problems according to the low rank matrix factorization model, an improved Kaczmarz iteration based matrix completion algorithm is proposed to solve these two sub-optimal problems alternatively and iteratively. Furthermore, dimensional reduction operations are executed before projections in the Kaczmarz iteration according to the Johnson-Lindenstrauss lemma with the goal of accelerating the speed of the algorithm. Experiment results show that the proposed algorithm presents superior reconstruction performance in terms of reconstruction accuracy, the probability of successful reconstruction, and reconstruction time than the existed algorithms.

**Key words:** wireless sensor networks; data collection; matrix completion; optimization; data reconstruction

## 1 引言

无线传感器网络由大量计算、存储、通信和能量等各种资源受限的无线传感器节点组成, 广泛应用于军事侦察、环境监测、智能家居、工农业自动化等各个领域. 传感器节点采集的监测数据需要利用无线多跳通信的方式, 借助其它传感器节点传输至汇聚(sink)节点, 从而实现对所部署区域的感知与监测. 由于传感器

节点的各种资源严格受限, 因此如何结合无线传感器网络自身的特点, 实现高效节能的数据收集机制是无线传感器网络研究领域中的关键问题<sup>[1]</sup>.

目前, 无线传感器网络中的数据收集方案根据采用的核心技术不同可以分为以下四类: 基于信号处理的数据收集方案、基于路由的数据收集方案、基于信息论的数据收集方案以及基于压缩感知的数据收集方案<sup>[2]</sup>. 其中, 基于信号处理的数据收集方案主要采用多

收稿日期: 2017-09-15; 修回日期: 2018-03-21; 责任编辑: 梅志强

项目基金: 国家自然科学基金(No. 61402094); 河北省自然科学基金(No. F2016501076); 辽宁省自然科学基金(No. 201602254); 教育部中央高校基本科研业务费项目(No. N172304022)

种信号变换与压缩编码技术,对传感器节点采集的感知数据进行压缩编码. 该类方案根据变换与压缩编码技术的还原特性不同又可细分为无损压缩方案和有损压缩方案,其缺点是压缩率低、资源消耗较大. 基于路由的数据收集方案以最小化网络的总体能耗为设计目标,通过构建不同结构的数据收集路径,从而实现传感器节点感知数据的高效收集. 代表性的方案包括基于数据收集树的方案和基于数据收集森林的方案等,其缺点是没有充分利用感知数据的时空相关性. 在基于信息论的数据收集方案中,传感器节点基于同组内其它节点提供的辅助信息,利用 Slepian-Wolf 编码或 Wyner-Ziv 编码技术对其感知数据进行压缩编码. 然而,由于网络中感知数据的统计分布特性不断变化以及辅助信息容易丢失等原因,导致该类方案在实际应用中具有较强的局限性. 基于压缩感知理论的数据收集方案以压缩感知理论为基础,充分利用传感器网络感知数据集的稀疏特性,通过采用多种优化算法在服务器端重构原始数据,从而实现无线传感器网络中高效节能的数据收集操作. 该类数据收集方案即适用于平面结构的无线传感器网络,也适用于分簇或树形等层次结构的无线传感器网络,目前已成为该领域的研究热点<sup>[3]</sup>.

压缩感知理论由于同时兼具压缩性、鲁棒性、计算非对称性和可渐进性等优异特性,非常适用于无线传感器网络的数据收集操作<sup>[4]</sup>. 然而,在实际应用中基于压缩感知理论的数据收集方案也存在以下一些不足之处. 首先,该类方案需要针对不同类型的感知数据预先指定一组稀疏基或者字典,从而确保感知数据集在指定稀疏基或字典下具有稀疏表示. 其次,该类方案中的测量矩阵通常是由服从独立同分布的高斯随机变量构成,其中绝大部分元素非零,因此每个传感器节点需要采集监测数据并执行大量的测量操作. 最后,压缩感知理论要求测量数据的数量需要超过指定的阈值才能保证精确重构,然而无线传感器网络中的噪声数据和丢包事件将会影响有效的测量数据数量,进而导致数据重构精度的降低.

针对上述问题,本文基于矩阵补全算法设计了无线传感器网络的数据收集方案. 矩阵补全理论可以看作是压缩感知理论的推广,即将一维稀疏向量重构问题推广至低秩矩阵的补全问题,目前已被广泛应用于机器学习、计算机视觉、推荐系统和信号处理等领域. 矩阵补全的基本思想是基于矩阵的低秩特性,依据矩阵中已知的部分元素恢复矩阵中所有的未知元素. 由于传感器网络的感知数据集具有较强的时空关联特性,因此时空监测数据矩阵满足低秩特性,可利用矩阵补全算法实现监测数据矩阵中未知元素的重构. 本文设计的方案通过将传感器网络数据收集问题转换为低秩

矩阵分解问题,并依据 Johnson-Lindenstrauss 引理对 Kaczmarz 迭代中的投影操作降维,从而在低维子空间中选取最优的行或列子集,进而加快凸优化子问题的求解. 与现有方案相比,本方案提高了矩阵补全算法的重构性能,实现了无线传感网中高效节能的数据收集操作.

## 2 理论基础

### 2.1 符号表示

向量的下标和上标分别表示向量的某个元素和该向量在某次迭代时的值,如  $x_i$  和  $x^{(j)}$  分别向量  $\mathbf{x}$  的第  $i$  个元素和向量  $\mathbf{x}$  在第  $j$  次迭代时的值. 矩阵的下标和上标分别表示矩阵的行向量和列向量,如  $\mathbf{A}_i$  和  $\mathbf{A}^j$  分别表示矩阵  $\mathbf{A}$  的第  $i$  个行向量和第  $j$  个列向量,而  $A_{ij}$  则表示矩阵  $\mathbf{A}$  的第  $i$  行第  $j$  列元素. 向量  $\mathbf{x}$  和矩阵  $\mathbf{A}$  的转置分别表示为  $\mathbf{x}^T$  和  $\mathbf{A}^T$ . 向量  $\mathbf{x} \in \mathbf{R}^n$  的  $l_2$  范数记为  $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ . 矩阵  $\mathbf{A} \in \mathbf{R}^{m \times n}$  的 Frobenius 范数记为  $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$ ,核范数记为  $\|\mathbf{A}\|_* = \sum_{i=1}^r \sigma_i$ ,其中  $r$  为矩阵  $\mathbf{A}$  的秩, $\sigma_i$  为其第  $i$  个奇异值.

### 2.2 矩阵补全理论

对于低秩矩阵  $\mathbf{X} \in \mathbf{R}^{m \times n}$ ,矩阵补全理论依据部分已知元素来重构所有未知元素. 该问题可以表述为以下优化问题

$$\begin{aligned} \min_{\mathbf{X}} \text{rank}(\mathbf{X}) \\ \text{s. t. } \mathbf{B} \circ \mathbf{X} = \mathbf{Y} \end{aligned} \quad (1)$$

其中  $\mathbf{Y} \in \mathbf{R}^{m \times n}$  为已知观测矩阵, $\mathbf{B} \in \mathbf{R}^{m \times n}$  为投影矩阵,其定义为

$$B_{ij} = \begin{cases} 1, & Y_{ij} \text{ 已知} \\ 0, & Y_{ij} \text{ 未知} \end{cases} \quad (2)$$

运算符  $\circ$  表示矩阵逐元素相乘,即  $Y_{ij} = B_{ij} X_{ij}$ .

由于矩阵的秩函数为非连续、非凸函数,直接求解优化问题式(1)较为困难. 因此一般将秩函数  $\text{rank}(\mathbf{X})$  松弛为其凸包络函数  $\|\mathbf{X}\|_*$ . 此时,优化问题式(1)可松弛为

$$\begin{aligned} \min_{\mathbf{X}} \|\mathbf{X}\|_* \\ \text{s. t. } \mathbf{B} \circ \mathbf{X} = \mathbf{Y} \end{aligned} \quad (3)$$

Candes 和 Tao 等人已证明当已知观测元素数量  $c \geq C\mu^2 nr (\log n)^6$  时,优化问题式(3)可以以不低于  $1 - n^{-3}$  的概率重构矩阵  $\mathbf{X}$ ,其中  $C$  为常数, $\mu$  为不相关系数, $r$  为矩阵的秩.

目前,学术界已涌现出许多不同类型的矩阵补全算法. Liu 等人通过将优化问题式(3)转化为半正定规划问题,提出了一种基于内点法 (interior point method)

的矩阵补全算法<sup>[5]</sup>. 然而, 该算法的计算复杂度较高, 求解速度较慢. Cai 等人通过将优化问题式(1)正则化, 提出了一种奇异值阈值 (Singular Value Thresholding, SVT) 算法<sup>[6]</sup> 求解满足约束条件的低秩矩阵. 与奇异值阈值算法类似, Ma 等人联合使用不动点连续算法和 Bregman 迭代 (Fixed Point Continuation with Bregman iterative method, FPCB) 算法求解优化问题式(1)的另一种正则化表示形式<sup>[7]</sup>. Xu 等人提出了一种通用的块坐标下降 (Block Coordinate Descent, BCD) 方法, 用于求解矩阵或张量的补全问题. 类似的代表性矩阵补全算法还包括分割 Bregman 迭代矩阵恢复 (Matrix recovery algorithm using Split Bregman, MSB) 算法、迭代重加权无约束  $l_q$  (Iterative Reweighted unconstrained  $l_q$ , IRuclq) 算法、增量秩幂分解 (Incremented Rank Power Factorization, IRPF) 算法、随机 Kaczmarz 矩阵分解 (Matrix Factorization algorithm based on Randomized Kaczmarz, MFRK) 算法<sup>[8]</sup> 等.

### 2.3 Kaczmarz 迭代算法

Kaczmarz 迭代算法是一种针对过采样线性等式系统设计的迭代型算法, 适用于求解大规模线性等式系统. 由于其具有使用简单、速度快、内存占用率低等优点, 已被广泛应用于数字信号处理、医学成像等应用领域. 在求解线性一致等式系统  $\mathbf{Ax} = \mathbf{b}$  时, Kaczmarz 迭代算法循环遍历矩阵  $\mathbf{A}$  的所有行, 并将当前迭代  $\mathbf{x}$  投影至由矩阵  $\mathbf{A}$  中当前选择行所对应的超平面上, 即

$$\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} + \frac{b_i - \langle \mathbf{A}_i, \mathbf{x}^{(j)} \rangle}{\|\mathbf{A}_i\|_2^2} \mathbf{A}_i^T \mathbf{e}_i \quad (4)$$

直至迭代结果收敛为止, 其中,  $\mathbf{A} \in \mathbf{R}^{m \times n}$ ,  $\mathbf{x} \in \mathbf{R}^n$ ,  $\mathbf{b} \in \mathbf{R}^m$ ,  $i = j \bmod m$ . Kaczmarz 迭代算法的几何解释如图 1 所示, 图中  $\mathbf{x}^{(0)}$  为初始解, 算法根据式(4) 逐次循环迭代, 最终将收敛至线性一致等式系统  $\mathbf{Ax} = \mathbf{b}$  的精确解.

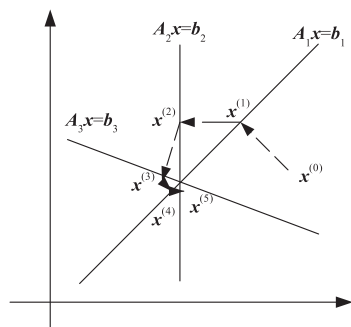


图1 Kaczmarz算法的几何解释

最近, Strohmer 等人提出了一种基于随机遍历的 Kaczmarz 迭代算法<sup>[9]</sup>. 该算法在选择投影的超平面时, 根据指定的概率分布进行选择, 即将矩阵  $\mathbf{A}$  的第  $i$  行被选中的概率设置为  $p(\mathbf{A}_i) = \|\mathbf{A}_i\|_2^2 / \|\mathbf{A}\|_F^2$ . Chen 等人

证明了该算法具有指数收敛率<sup>[10]</sup>, 即

$$E(\|\mathbf{x}^{(j)} - \mathbf{x}^*\|^2) \leq (1 - \kappa(\mathbf{A})^{-2})^j \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 \quad (5)$$

其中  $\kappa(\mathbf{A}) = \|\mathbf{A}\|_F \|\mathbf{A}^\dagger\|_2$  是矩阵  $\mathbf{A}$  的条件数,  $\mathbf{x}^*$  是线性一致等式系统  $\mathbf{Ax} = \mathbf{b}$  的精确解.

针对线性不一致等式系统  $\mathbf{Ax} = \mathbf{b} + \mathbf{w}$ , Eldar 等人同样证明了基于随机遍历的 Kaczmarz 迭代算法具有指数收敛率<sup>[11]</sup>, 即

$$E(\|\mathbf{x}^{(j)} - \mathbf{x}^*\|^2) \leq (1 - \kappa(\mathbf{A})^{-2})^j \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 + \kappa(\mathbf{A}) \max_i \left( \frac{w_i}{\|\mathbf{A}_i\|_2} \right) \quad (6)$$

其中  $\mathbf{w}$  为误差向量 (或噪声向量).

## 3 系统模型

假设整个无线传感器网络由  $n$  个无线传感器节点和 1 个汇聚节点构成, 每个无线传感器节点的采样周期为  $\Delta t$ , 每个数据重构周期由  $m$  个采样周期构成. 在每个数据重构周期内, 传感器网络中第  $i$  个无线传感器节点的感知数据构成向量  $\mathbf{x}_i = [x_{i,\Delta t}, x_{i,2\Delta t}, \dots, x_{i,m\Delta t}]^T$ , 其中  $x_{i,j\Delta t}$  为第  $i$  个传感器节点在第  $j$  个采样周期所感知的监测数据. 因此, 整个无线传感器网络在该数据重构周期内的所有感知数据构成一个  $m \times n$  的感知矩阵  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ .

为了降低传感器节点的能量消耗, 每个传感器节点在一个数据重构周期内仅需采集部分感知数据, 并将其传输至汇聚节点. 与文献[12]类似, 汇聚节点可以通过广播或者预设的方式将采样率分发给网络中的所有传感器节点. 每个传感器节点即可根据该采样率随机选择在当前采样周期内是否需要采集感知数据并将其传输至汇聚节点. 在每个数据重构周期结束后, 汇聚节点根据接收到的感知数据集构造观测矩阵  $\mathbf{Y}$  和投影矩阵  $\mathbf{B}$ , 首先将接收到的  $x_{i,j\Delta t}$  填充至观测矩阵  $\mathbf{Y}$  的第  $i$  行第  $j$  列, 然后将投影矩阵  $\mathbf{B}$  中相同位置上的元素置 1. 最后, 运行第 3 节中介绍的矩阵补全算法重构感知矩阵  $\mathbf{X}$ , 补全所有未传输的感知数据.

## 4 改进的矩阵补全算法

### 4.1 矩阵分解模型

根据低秩矩阵分解模型<sup>[8]</sup>, 优化问题式(3)中的矩阵  $\mathbf{X}$  可以分解成两个低秩矩阵  $\mathbf{U} \in \mathbf{R}^{m \times r}$  和  $\mathbf{V} \in \mathbf{R}^{r \times n}$  的乘积, 即  $\mathbf{X} = \mathbf{UV}$ , 其中  $r$  为矩阵  $\mathbf{X}$  的秩. 因此, 优化问题式(3)可以转换成

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{Y} - \mathbf{B} \circ (\mathbf{UV})\|_F^2 \quad (7)$$

通过分离优化变量  $\mathbf{U}$  和  $\mathbf{V}$ , 优化问题式(7)可进一步转换成如下两个凸优化子问题

$$\min_{\mathbf{U}} \sum_{i,j} (\mathbf{Y}_{ij} - \mathbf{B}_{ij} \langle \mathbf{U}_i, \mathbf{V}^j \rangle)^2 \quad (8)$$

$$\min_{\mathbf{V}} \sum_{i,j} (Y_{ij} - \mathbf{B}_{ij} \langle \mathbf{U}_i, \mathbf{V}^j \rangle)^2 \quad (9)$$

此时,利用 Kaczmarz 迭代算法循环遍历矩阵  $\mathbf{U}$  的所有行,并将当前迭代  $\mathbf{U}_i$  投影至与矩阵  $\mathbf{V}$  中当前选择列所对应的超平面上,即可求解子优化问题式(8). 与之类似,利用 Kaczmarz 迭代算法循环遍历矩阵  $\mathbf{V}$  的所有列,并将当前迭代  $\mathbf{V}^j$  投影至与矩阵  $\mathbf{U}$  中当前选择行所对应的超平面上,即可求解子优化问题式(9). 本文通过循环交替求解子凸优化问题式(8)和式(9),即可由观测矩阵  $\mathbf{Y}$  的部分已知元素重构感知矩阵  $\mathbf{X}$  的所有未知元素.

#### 4.2 矩阵补全算法

为了加快重构感知矩阵  $\mathbf{X}$  中未知元素的计算速度,本文设计了一种基于改进 Kaczmarz 迭代的矩阵补全算法. 在利用 Kaczmarz 迭代求解凸优化子问题式(8)和式(9)的计算过程中,依据 Johnson-Lindenstrauss 引理可以对 Kaczmarz 迭代式(4)中的投影操作降维,从而提高矩阵补全算法的计算速度.

**引理 1 (Johnson-Lindenstrauss 引理<sup>[13]</sup>)** 令  $0 < \delta < 1$ ,  $S$  为  $n$  维欧几里得空间  $\mathbf{R}^n$  中的一个有限点集,则对于任意的整数  $d \geq C \log |S| / \delta^2$ , 存在 Lipschitz 映射  $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}^d$  满足

$$(1 - \delta) \leq \frac{\|\Phi(s_i) - \Phi(s_j)\|_2^2}{\|s_i - s_j\|_2^2} \leq (1 + \delta) \quad (10)$$

其中  $s_i, s_j \in S$ ,  $C$  是一个常数.

Johnson-Lindenstrauss 引理指出,  $n$  维空间中的  $|S|$  个点可以嵌入到  $d$  维子空间中,同时保持原始空间中任意两点之间的几何关系近似不变. 该引理的几何解释如图 2 所示. 现有理论已证明高斯随机矩阵映射具有较高的概率满足式(10).

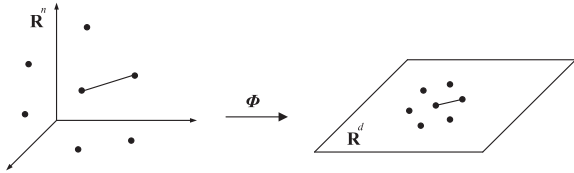


图2 Johnson-Lindenstrauss引理的几何解释

本文设计的基于改进 Kaczmarz 迭代的矩阵补全算法如算法 1 所示. 算法循环执行 maxCount 次,当连续两次循环的优化目标式(7)之差低于预设阈值  $\varepsilon$  时算法亦可提前结束. 算法交替求解子优化问题式(8)(第 3 行~第 17 行)和子优化问题式(9)(第 18 行~第 33 行).

#### 算法 1 基于改进 Kaczmarz 迭代的矩阵补全算法

输入  $\mathbf{Y} \in \mathbf{R}^{m \times n}$ ,  $\mathbf{U} \in \mathbf{R}^{m \times r}$ ,  $\mathbf{V} \in \mathbf{R}^{r \times n}$ , maxCount,  $\varepsilon$

1  $k = 1$   
2 While  $k \leq \text{maxCount}$  or  $\text{obj}(k-1) - \text{obj}(k) \geq \varepsilon$  do

3 创建  $s \times r$  高斯随机矩阵  $\Phi$   
4 依据概率  $\|\mathbf{V}^j\|_2^2 / \|\mathbf{V}\|_F^2$  随机选取  $\mathbf{V}$  中的  $r$  列构造子矩阵  $\hat{\mathbf{V}}$   
5  $\mathbf{A} = \Phi \hat{\mathbf{V}}$   
6 For  $i = 1$  to  $m$  do  
7     For  $j = 1$  to  $r$  do  
8          $\gamma_j = |Y_{ij} - \langle \mathbf{U}_i \Phi^T, \mathbf{A}^j \rangle| / \|\mathbf{A}^j\|_2$   
9     End For  
10      $p = \arg \max_{j \in \{1, \dots, r\}} \gamma_j$   
11      $\gamma_p = |Y_{ip} - \langle \mathbf{U}_i, \hat{\mathbf{V}}^p \rangle| / \|\hat{\mathbf{V}}^p\|_2$   
12     随机选择  $\mathbf{V}$  中的列  $\mathbf{V}^l$ , 计算  $\gamma_l = |Y_{il} - \langle \mathbf{U}_i, \mathbf{V}^l \rangle| / \|\mathbf{V}^l\|_2$   
13     If  $\gamma_l > \gamma_p$   
14          $\mathbf{U}_i = \mathbf{U}_i + ((Y_{il} - \langle \mathbf{U}_i, \mathbf{V}^l \rangle) / \|\mathbf{V}^l\|_2^2) (\mathbf{V}^l)^T$   
15     Else  
16          $\mathbf{U}_i = \mathbf{U}_i + ((Y_{ip} - \langle \mathbf{U}_i, \hat{\mathbf{V}}^p \rangle) / \|\hat{\mathbf{V}}^p\|_2^2) (\hat{\mathbf{V}}^p)^T$   
17     End If  
18 End For  
19 创建  $t \times r$  高斯随机矩阵  $\Psi$   
20 依据概率  $\|\mathbf{U}_i\|_2^2 / \|\mathbf{U}\|_F^2$  随机选取  $\mathbf{U}$  中的  $r$  行构造子矩阵  $\hat{\mathbf{U}}$   
21  $\mathbf{B} = \Psi \hat{\mathbf{U}}^T$   
22 For  $j = 1$  to  $n$  do  
23     For  $i = 1$  to  $r$  do  
24          $\gamma_i = |Y_{ij} - \langle \mathbf{B}^i, \Psi \mathbf{V}^j \rangle| / \|\mathbf{B}^i\|_2$   
25     End For  
26      $q = \arg \max_{i \in \{1, \dots, r\}} \gamma_i$   
27      $\gamma_q = |Y_{qj} - \langle \hat{\mathbf{U}}_q, \mathbf{V}^j \rangle| / \|\hat{\mathbf{U}}_q\|_2$   
28     随机选择  $\mathbf{U}$  中的行  $\mathbf{U}_l$ , 计算  $\gamma_l = |Y_{lj} - \langle \mathbf{U}_l, \mathbf{V}^j \rangle| / \|\mathbf{U}_l\|_2$   
29     If  $\gamma_l > \gamma_q$   
30          $\mathbf{V}^j = \mathbf{V}^j + ((Y_{lj} - \langle \mathbf{U}_l, \mathbf{V}^j \rangle) / \|\mathbf{U}_l\|_2^2) (\mathbf{U}_l)^T$   
31     Else  
32          $\mathbf{V}^j = \mathbf{V}^j + ((Y_{qj} - \langle \hat{\mathbf{U}}_q, \mathbf{V}^j \rangle) / \|\hat{\mathbf{U}}_q\|_2^2) (\hat{\mathbf{U}}_q)^T$   
33     End If  
34     End For  
35      $k = k + 1$   
36 End While  
输出  $\hat{\mathbf{X}} = \mathbf{U}\mathbf{V}$

求解子优化问题式(8)时,首先创建  $s \times r$  的高斯随机矩阵  $\Phi$ ,用于实现后续 Kaczmarz 迭代中投影操作的降维运算,其中  $s = \log(r)$ ,并依据矩阵  $\mathbf{V}$  中每列大小随机选取  $r$  列构造子矩阵  $\hat{\mathbf{V}}$ . 然后对  $\mathbf{U}$  的每行进行循环,选取矩阵  $\mathbf{V}$  中与当前行  $\mathbf{U}_i$  最匹配的列进行投影. 由引理 1 可知,  $(\Phi \mathbf{U}_i^T)^T$  与  $\Phi \hat{\mathbf{V}}^p$  的内积和  $\mathbf{U}_i$  与  $\hat{\mathbf{V}}^p$  的内积近似,故算法可在降维后的子空间中选取与  $\mathbf{U}_i$  最匹配的列  $\hat{\mathbf{V}}^p$ , 其中  $p = \arg \max_j |Y_{ij} - \langle \mathbf{U}_i \Phi^T, \Phi \hat{\mathbf{V}}^j \rangle| / \|\Phi \hat{\mathbf{V}}^j\|_2$ . 为了确保列选取的最优性,算法随机选取矩阵  $\mathbf{V}$  中的一列  $\mathbf{V}^l$  对  $\hat{\mathbf{V}}^p$  进行验证,并保留  $\hat{\mathbf{V}}^p$  和  $\mathbf{V}^l$  两者中与当前行  $\mathbf{U}_i$  最匹配的列进行投影运算.

子优化问题式(9)的求解方法与上述过程类似. 首先创建  $t \times r$  的高斯随机矩阵  $\Psi$ , 其中  $t = \log(r)$ , 并依据矩阵  $U$  中每行的大小随机选取  $r$  行构造子矩阵  $\hat{U}$ . 然后, 对矩阵  $V$  的每列进行循环, 选取矩阵  $U$  中与当前列  $V^j$  最匹配的行  $\hat{U}_q$  进行投影, 其中  $q = \operatorname{argmax}_i |Y_{ij} - \langle \hat{U}_i \Psi^T, \Psi V^j \rangle| / \|\hat{U}_i \Psi^T\|_2$ . 同时, 为了确保行选取的最优性, 随机选取  $U$  中的一行  $U_i$  进行验证, 并保留  $\hat{U}_q$  和  $U_i$  两者中与当前列  $V^j$  最匹配的行进行投影运算.

循环结束后, 将迭代结果  $U$  和  $V$  相乘即可得到重构感知矩阵  $\hat{X}$ . 在算法 1 中, 最外层循环最多执行  $k$  次. 在迭代求解子优化问题式(8)时, 需要执行  $mr$  次循环. 在迭代求解子优化问题式(9)时, 需要执行  $nr$  次循环. 因此, 算法 1 的计算复杂度为  $O(kr \max(m, n))$ . 而 MKFC 算法<sup>[8]</sup> 的计算复杂度为  $O(kmn)$ , 因此本算法由于进行了降维操作, 在计算速度上优于 MKFC 算法.

## 5 实验结果与分析

本节分别利用合成数据集和真实数据集测试文中所提出的基于改进 Kaczmarz 迭代的矩阵补全算法的性能. 为了方便描述, 将该算法简称为 IKMC (Improved Kaczmarz iteration based Matrix Completion) 算法. 通过与 SVT 算法<sup>[6]</sup>、FPCB 算法<sup>[7]</sup> 以及 MFRK 算法<sup>[8]</sup> 进行比较, 评测 IKMC 算法在重构精度、成功重构概率、重构时

间等方面的性能. 仿真实验使用 Matlab 软件, 每次实验重复运行 100 次后取平均值作为该项实验的计算结果.

合成数据集的生成方式与文献[8]一致, 首先利用两个低秩高斯随机矩阵  $U$  和  $V$  相乘得到矩阵  $X$ , 然后正则化矩阵  $X$  的行并将每行乘以一个  $(0, 1)$  之间服从均匀分布的随机数. 因此, 矩阵  $X$  的所有行可以看作是  $n$  维空间中一组具有不同方向、不同大小的向量. 在生成观测矩阵  $Y$  时, 根据采样率  $s$  随机选取矩阵  $X$  中的元素并将其填充至矩阵  $Y$  中相应的位置上, 同时将  $Y$  中其余的元素置 0. 合成数据集实验中的默认参数设置为: 感知矩阵  $X$  的行数  $m = 250$ , 列数  $n = 250$ , 秩  $r = 30$ , 采样率  $s = 30\%$ , 成功重构阈值  $\delta = 0.01$ , 算法终止阈值  $\varepsilon = 0.0001$ .

矩阵补全算法的重构精度可用归一化均方误差  $\eta = \|X - \hat{X}\|_F / \|X\|_F$  来衡量. 显然, 该误差与矩阵补全算法的重构精度成反比. 图 3 展示了 IKMC 算法、SVT 算法、FPCB 算法、MFRK 算法在不同秩、采样率以及矩阵大小条件下的重构精度. 从图 3 中可以看出, IKMC 算法具有最高的重构精度, 并且以上四种算法的重构精度随着矩阵秩的增加而降低, 随着采样率和矩阵大小的增加而升高. 产生上述现象的原因在于矩阵补全的复杂程度与矩阵的秩成正比, 与采样率和矩阵大小成反比. 即矩阵的秩越高, 补全感知矩阵中的未知元素越困难; 矩阵的采样率越高、矩阵越大, 补全感知矩阵中的未知元素越容易.

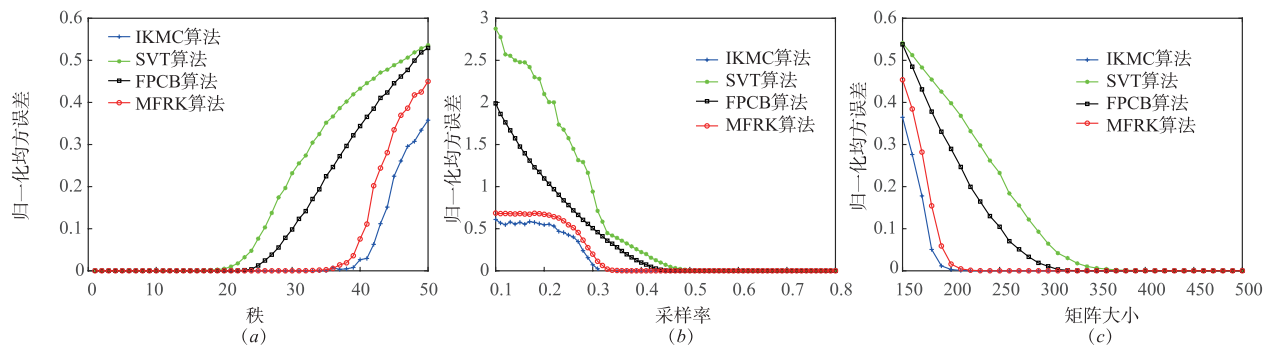


图3 合成数据集实验的重构精度

图 4 展示了 IKMC 算法、SVT 算法、FPCB 算法、MFRK 算法在不同秩、采样率以及矩阵大小条件下的成功重构概率. 当重构误差低于成功重构阈值  $\delta$  时, 即认为本次矩阵补全运算重构成功. 由图 4 可知, IKMC 算法具有最高的成功重构概率, 并且以上四种算法的成功重构概率随着矩阵秩的增加而降低, 随着采样率和矩阵大小的增加而升高. 与上述原因相同, 矩阵补全的复杂程度与矩阵的秩成正比, 与采样率和矩阵大小成反比. 补全运算越复杂, 数据重构的成功概率就越低.

图 5 展示了 IKMC 算法、SVT 算法、FPCB 算法、

MFRK 算法在不同秩、采样率以及矩阵大小条件下的重构时间. 该重构时间在一定程度上反映了不同矩阵补全算法的计算复杂程度. 从图 5 中可以看出, IKMC 算法和 MFRK 算法具有相似的计算时间, 并且都低于 SVT 算法和 FPCB 算法 (除了图 5(c) 中矩阵的大小高于 380 时 SVT 算法具有最低的计算时间). 由图 3、图 4、图 5 可知, 与 SVT 算法和 FPCB 算法相比, IKMC 算法和 MFRK 算法具有较高的重构精度、成功重构概率和较快的重构速度. 本文将继续在真实数据集上比较 IKMC 算法和 MFRK 算法的重构性能.

本文使用 Intel Berkeley 实验室提供的无线传感器网络采集数据集作为真实数据集进行重构性能测试. 具体数据取自 50 个 Mica2Dot 节点(由于损毁等原因未使用其中的 5 号、32 号、50 号、53 号节点所对应的数据)于 2004 年 3 月 1 日采集到的温度、湿度、光照数据. 由于数据集生成矩阵的秩和大小已经固定,因此实验中只测试了不同采样率条件下的数据重构性能. 图 6

展示了 IKMC 算法和 MFRK 算法在不同采样率条件下的平均重构误差. 从图中可以看出,IKMC 算法的重构误差低于 MFRK 算法. 同时,当采样率高于 30% 时,本文提出的 IKMC 算法在重构温度、光照、湿度数据时的误差分别低于 0.3 摄氏度、2.8 流明和 1.0%,数据重构精度能够满足实际需求.

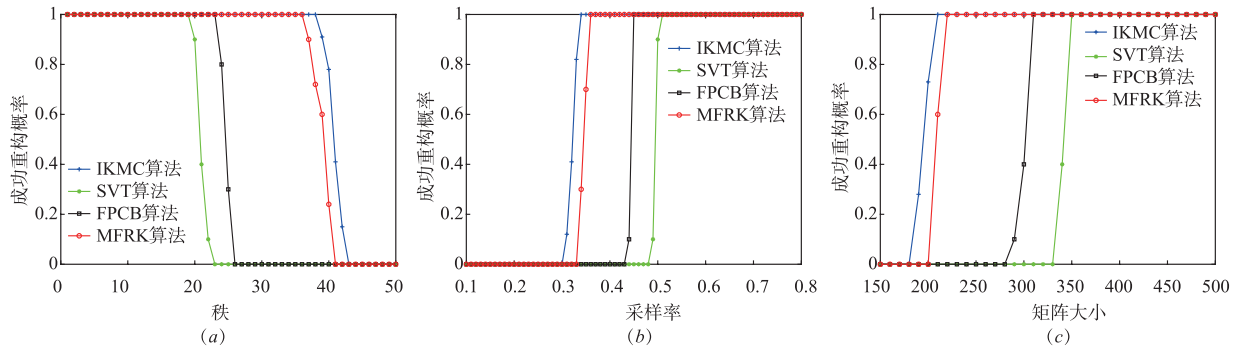


图4 合成数据集实验的成功重构概率

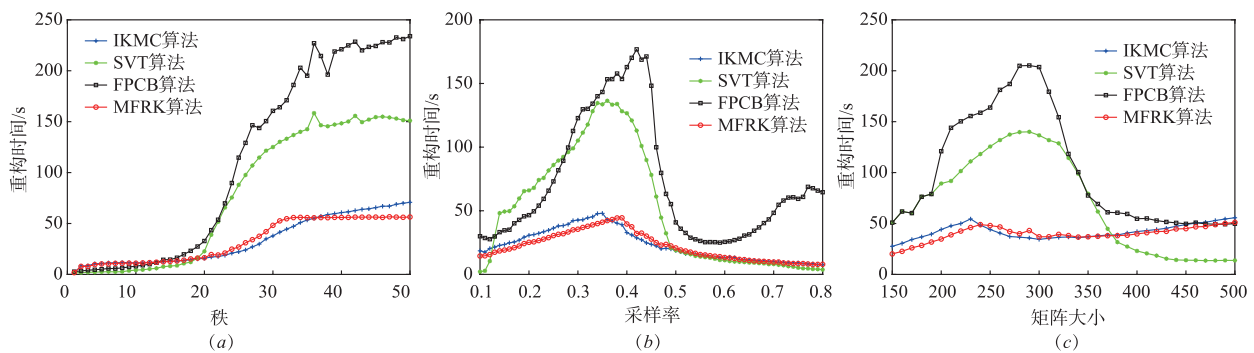


图5 合成数据集实验的重构时间

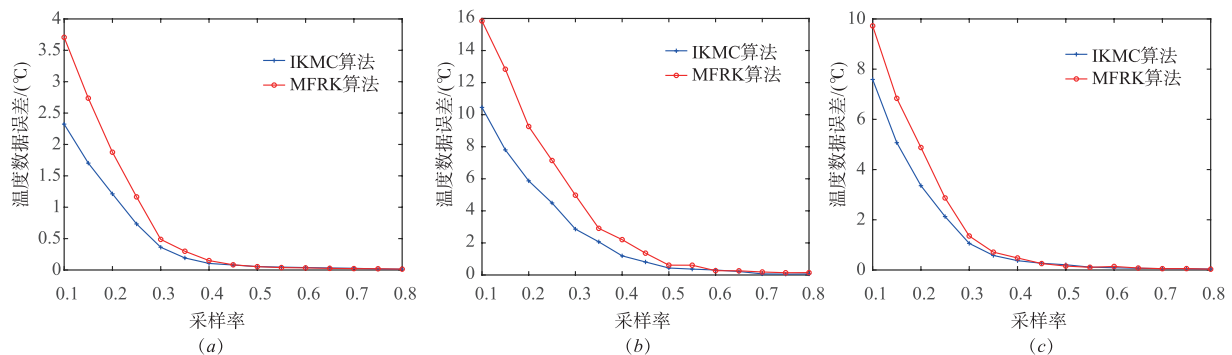


图6 真实数据集实验的重构精度

## 6 结束语

本文提出了一种基于矩阵补全的无线传感器网络数据收集方案. 通过将传感器网络数据收集问题转换为低秩矩阵分解问题,并依据 Johnson-Lindenstrauss 引理对 Kaczmarz 迭代中的投影操作降维,从而提高了矩

阵补全算法的重构性能. 在仿真实验中,分别利用合成数据集和真实数据集测试和分析了所设计的矩阵补全算法在重构精度、成功重构概率以及重构时间等方面的性能指标. 下一步研究方向将着重考虑结合概率近似方法进一步提高矩阵补全算法的重构性能.

## 参考文献

- [1] 徐佳,冯鑫,杨富贵,王传平,王汝传. 最大化最小能耗概率的移动 Sink 无线传感器网络数据收集方法[J]. 电子学报,2015,43(12):2470-2475.  
XU Jia, FENG Xin, YANG Fu-gui, WANG Chuan-ping, WANG Ru-chuan. A data collection method by maximizing minimum probability of energy consumption for mobile sink based WSNs [J]. Acta Electronica Sinica, 2015, 43(12):2470-2475. (in Chinese)
- [2] CAMPOBELLO G, SEGRETO A, SERRANO S. Data gathering techniques for wireless sensor networks: a comparison [J]. International Journal of Distributed Sensor Networks, 2016, 12(3):1-17.
- [3] MIDDYA R, CHAKRAVARTY N, NASKAR M. Compressive sensing in wireless sensor networks- a survey [J]. IETE Technical Review, 2017, 34(6):642-654.
- [4] 焦李成,杨淑媛,刘芳,侯彪. 压缩感知回顾与展望[J]. 电子学报,2011,39(7):1651-1662.  
JIAO Li-cheng, YANG Shu-yuan, LIU Fang, Hou Biao. Development and prospect of compressive sensing [J]. Acta Electronica Sinica, 2011, 39(7):1651-1662 (in Chinese).
- [5] LIU Z, VANDENBERGHE L. Interior-point method for nuclear norm approximation with application to system identification [J]. SIAM Journal on Matrix Analysis and Applications, 2009, 31(3):1235-1256.
- [6] CAI J, CANDÈS E, SHEN Z. A singular value thresholding algorithm for matrix completion [J]. SIAM Journal on Optimization, 2010, 20(4):1956-1982.
- [7] MA S, GOLDFARB D, CHEN L. Fixed point and Bregman iterative methods for matrix rank minimization [J]. Mathematical Programming, 2011, 128(1):321-353.
- [8] GOGNA A, MAJUMDAR A. Matrix factorization model using Kaczmarz algorithm; application in sensor localization [A]. Proceedings of IEEE International Conference on Digital Signal Processing [C]. Piscataway: IEEE, 2015. 219-223.
- [9] STROHMER T, VERSHYNIN R. A randomized Kaczmarz algorithm for linear systems with exponential convergence [J]. Journal of Fourier Analysis and Applications, 2009, 15(2):262-278.
- [10] CHEN X, POWELL A. Almost sure convergence of the Kaczmarz algorithm with random measurements [J]. Journal of Fourier Analysis and Applications, 2012, 18(6):1195-1214.
- [11] NEEDELL D. Randomized Kaczmarz solver for noisy linear systems [J]. BIT Numerical Mathematics, 2010, 50(2):395-403.
- [12] CHENG J, YE Q, JIANG HB, WANG D, WANG CG. 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.
- [13] ELDAR Y, NEEDELL D. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss Lemma [J]. Numerical Algorithms, 2011, 58(2):163-177

## 作者简介



李国瑞(通信作者) 男,1980年7月出生,山西夏县人,博士,副教授,硕士生导师,毕业于北京工业大学,美国弗吉尼亚理工学院暨州立大学访问学者,主要研究方向为物联网、优化算法。

E-mail: lgr@neuq.edu.cn



王颖 女,1979年8月出生,天津人,硕士,讲师,毕业于燕山大学,主要研究方向为无线传感器网络,优化算法。

E-mail: wyqhd@hotmail.com



王聪 男,1981年8月出生,河北抚宁人,博士,副教授,毕业于东北大学,主要研究方向为云计算,智能优化算法。

E-mail: wangcong@neuq.edu.cn