

无线传感器网络中基于 Top- $|K|$ 查询的分布式数据重构方法^{*}

李国瑞

(东北大学 秦皇岛分校 计算机与通信工程学院, 河北 秦皇岛 066000)

通讯作者: 李国瑞, E-mail: lgr@mail.neuq.edu.cn

摘要: 针对分簇结构或多 Sink 节点的无线传感器网络应用场景, 提出了一种基于 Top- $|K|$ 查询的分布式数据重构方法. 该方法包括分布式迭代硬阈值算法和基于双阈值的分布式 Top- $|K|$ 查询算法两个部分. 其中, 管理节点和成员节点同时运行分布式迭代硬阈值算法, 以分布式方式实现迭代硬阈值计算. 同时, 管理节点和成员节点运行基于双阈值的分布式 Top- $|K|$ 查询算法, 以分布式方式实现前一算法中查询绝对值最大的前 K 项元素和操作. 实验结果表明, 该方法的数据重构性能与现有方法无明显差异, 同时能够有效地减少管理节点和成员节点之间的交互次数, 并且降低网络中传输的数据量.

关键词: 无线传感器网络; 分布式算法; 压缩感知; 数据重构

中文引用格式: 李国瑞. 无线传感器网络中基于 Top- $|K|$ 查询的分布式数据重构方法. 软件学报, 2014, 25(Suppl. (1)): 139-148. <http://www.jos.org.cn/1000-9825/14016.htm>

英文引用格式: Li GR. Top- $|K|$ query based distributed data reconstruction method in wireless sensor networks. Ruan Jian Xue Bao/Journal of Software, 2014, 25(Suppl. (1)): 139-148 (in Chinese). <http://www.jos.org.cn/1000-9825/14016.htm>

Top- $|K|$ Query Based Distributed Data Reconstruction Method in Wireless Sensor Networks

LI Guo-Rui

(School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066000, China)

Corresponding author: LI Guo-Rui, E-mail: lgr@mail.neuq.edu.cn

Abstract: To address the application scenarios of cluster based or multiple sink based wireless sensor networks, this paper proposes a Top- $|K|$ query based distributed data reconstruction method which includes the distributed iterative hard thresholding algorithm and the dual thresholds based distributed Top- $|K|$ query algorithm. The administrative node and member nodes run the distributed iterative hard thresholding algorithm, realizing the iterative hard thresholding algorithm in a distributed manner at the same time. Meanwhile, the administrative node and member nodes run the dual thresholds based distributed Top- $|K|$ query algorithm which can query the Top- $|K|$ sum of the elements in the result of the previous algorithm in a distributed manner as well. Experimental result show that the proposed method can achieve the same data reconstruction performance as the existing methods. Furthermore, it can also decrease the transmitted data number and interactive time between the administrative node and member nodes.

Key words: wireless sensor networks; distributed algorithm; compressive sampling; data reconstruction

无线传感器网络由大量低成本、资源受限的微型传感器节点组成, 通过利用近距离无线通信方式组成一个多跳的自组织网络系统, 从而实现对所部署区域的协作感知与监测. 由于无线传感器节点自身的计算、存储、通信和能量资源有限, 如何充分利用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, 是无线传感器网络中亟待解决的关键问题.

时空关联性分析预测^[1-3]和分布式压缩编码^[4,5]是无线传感网中普遍采用的两类低能耗数据收集方式. 其

* 基金项目: 国家自然科学基金(61402094); 国家教育部博士点基金(20120042120009); 国家教育部中央高校基本科研基金(N120423005); 河北省自然科学基金(F2012501014); 东北大学秦皇岛分校科技支撑项目(XNK201401)

收稿时间: 2014-05-10; 定稿时间: 2014-08-26

中,时空关联性分析预测方法主要基于无线传感器节点采集数据的时空关联特性,通过构建时空预测模型对网内节点采集的数据进行分析和预测,从而降低网络整体的数据传输量,延长无线传感网的生存时间.分布式压缩编码方法主要基于 Slepian-Wolf 编码理论,通过设计分布式编码算法,从而减少传感器节点间传输的数据量,进而提高无线传感器网络的数据收集效率.以上两类方法都需要在传感器节点中执行复杂的运算,即构建时空数据关联预测模型或对数据进行压缩编码,同时还需要在网络中传输大量的参数和数据.因此,现有的数据收集方式对传感器节点的各种资源消耗过高,单纯地通过修改模型或调整参数并不能从本质上克服方式本身的缺陷.

近年来,压缩感知理论作为信号处理领域中备受关注的前沿热点研究理论之一,为无线传感器网络中数据收集问题的研究提供了新的思路 and 方向.经典的香农采样定理仅利用了信号的有限带宽假设,没有充分利用其它的有用先验信息,因此采集到的数据具有较高的冗余性.压缩感知理论突破了香农采样定理的理论极限,其本质是稀疏或可压缩信号的少量随机线性投影即包含了信号重构和处理的足够信息,也就是仅仅利用信号稀疏或可压缩的先验知识和少量的全局线性测量即可以进行原始信号的精确重建^[6].由于压缩感知理论同时具有通用性、鲁棒性、压缩性、计算非对称性和可渐进性等优异特性^[7],因此非常适用于无线传感器网络的应用环境.目前,基于压缩感知理论的无线传感器网络应用研究刚刚起步,主要集中在数据收集^[8]、目标定位^[9]、信道估计^[10]、路由协议^[11]等方面的研究.

在无线传感器网络的数据收集问题中,文献[12]首次基于压缩感知理论提出了 CDG(compressive data gathering)方案,通过采用链式拓扑结构和分布式随机系数投影,实现了无线传感器节点采集数据的压缩感知和 Sink 节点的稀疏数据重构.文献[13]基于无线传感器节点采集数据的稀疏度具有时空不一致的特性,通过在数据重构过程中引入自回归模型,设计了一种具有自适应特性的基于压缩感知的数据收集方案.文献[14]利用离散余弦变换函数的对称和正交特性构造了一组压缩稀疏函数,基于部分传感器节点的采集数据值重构出压缩稀疏函数的系数,进而对全网节点采集的数据值进行重构.文献[15]通过利用混沌压缩感知矩阵设计了传感器节点的调度规则与采集数据的投影压缩方法,从而实现了一种兼具安全和高效特性的数据收集方案.以上方案均利用压缩感知理论实现了无线传感器网络中高效、节能的数据收集机制.然而,这些方案都采用集中式的数据重构算法,即网内传感器节点的投影数据必须全部传输至 Sink 节点,统一进行数据重构操作.在分簇结构或多 Sink 节点的无线传感器网络应用场景中,各个簇头节点或成员 Sink 节点均无法同时获取管理 Sink 节点的数据重构结果,即以上方案无法满足分布式数据重构的需求.

本文针对分簇结构或多 Sink 节点的无线传感器网络应用场景,提出了一种基于 Top-|K|查询的分布式数据重构方法,通过利用分布式迭代硬阈值算法和基于双阈值的分布式 Top-|K|查询算法实现了分布式数据重构,并通过仿真实验验证了所设计方法的有效性和算法性能.

1 压缩感知理论

1.1 压缩感知理论

在压缩感知理论中,定义向量 $x \in \mathbb{R}^n$ 为原始信号,向量 $y \in \mathbb{R}^m$ 为观测信号,测量矩阵 $\Phi \in \mathbb{R}^{m \times n}$ 用于将原始信号投影为观测信号,其中 $m < n$.整个信号采集过程可以表示为

$$y = \Phi x \quad (1)$$

由于 $m < n$,因此欠定系统(1)有无穷多个解,无法由观测信号 y 对原始信号 x 进行精确数据重构.然而,常见的自然信号虽然在时域内是非稀疏的,但是在某个变化基 $\Psi \in \mathbb{R}^{n \times n}$ 下是稀疏的,因此原始信号 x 可以表示成:

$$x = \Psi s \quad (2)$$

其中, $s \in \mathbb{R}^n$ 为 K 稀疏向量,即 $\|s\|_0 = K$, $\|s\|_0$ 表示向量 s 中非零元素的个数.将式(2)代入式(1)可得:

$$y = \Phi x = \Phi \Psi s = A s, \quad (3)$$

其中,矩阵 $A = \Phi \Psi \in \mathbb{R}^{m \times n}$ 称为感知矩阵.现有理论已经证明,如果感知矩阵 A 满足约束等距性质^[16],并且观测信号 y 的维数 $m = O(K \ln(n))$,则原始信号 x 可通过求解优化问题

$$\begin{aligned} \min \|x\|_0 \\ \text{s.t. } Ax = y \end{aligned} \tag{4}$$

进行精确重构.然而,优化问题(4)是 NP 困难的,在实际求解过程中一般将其松弛为优化问题

$$\begin{aligned} \min \|x\|_1 \\ \text{s.t. } Ax = y \end{aligned} \tag{5}$$

进行求解,其中 $\|x\|_1$ 表示向量 x 中非零元素的绝对值之和.针对优化问题(5),常用的求解方法包括凸优化方法、贪婪算法、迭代算法、统计优化方法等.

1.2 迭代硬阈值算法

迭代硬阈值算法^[17]属于迭代类算法,其基本思想是利用感知矩阵 A 的转置来近似感知矩阵的逆,即令 $A^T \approx A^{-1}$,通过多次迭代运算来求解优化问题(5).

由约束条件 $Ax = y$ 可知 $A^T Ax = A^T y$,进而可得:

$$x = (I - A^T A)x + A^T y \tag{6}$$

以式(6)为基础构造迭代公式

$$x^{(t+1)} = x^{(t)} + A^T (y - Ax^{(t)}) \tag{7}$$

因为原始信号是 K 稀疏的,所以在每次迭代过程中通过硬阈值操作 $H_K(\cdot)$ 选取当前信号中绝对值最大的前 K 项元素进入下一次迭代,即将式(7)修改为

$$x^{(t+1)} = H_K(x^{(t)} + A^T (y - Ax^{(t)})) \tag{8}$$

硬阈值操作

$$H_K(x)_i = \begin{cases} x_i, & x_i \in \{x_{Top-|K|}\} \\ 0, & x_i \notin \{x_{Top-|K|}\} \end{cases}, \tag{9}$$

其中 $x_{Top-|K|}$ 表示 x 中绝对值最大的前 K 项元素.经过若干次迭代后,迭代硬阈值算法将收敛至优化问题(5)的最优解.

2 系统模型

本文中提出的分布式数据重构方法适用于分簇结构的无线传感器网络或具有多 Sink 节点的无线传感器网络.系统模型如图 1 所示,图中的代理节点既可以是簇头节点,也可以是 Sink 节点.

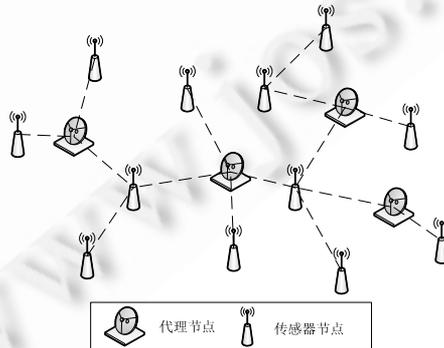


图 1 系统模型

设无线传感器网络中共有 p 个代理节点,令 1 号代理节点为管理节点,其他代理节点为成员节点.由于无线传感器网络所监测的物理数据具有空间关联性,因此在某一时刻内所有传感器节点采集到的数据向量 $x \in \mathbb{R}^n$ 具有稀疏度 K .设每个代理节点 $i \in \{1, \dots, p\}$ 分别收集了经过其感知矩阵 $A_i \in \mathbb{R}^{m_i \times n}$ 投影后的测量信号 y_i ,因此网内

所有代理节点共收集到了 $m = m_1 + m_2 + \dots + m_p$ 个测量数据.令整个传感器网络系统的感知矩阵为 A ,收集到的测量信号为 y ,则:

$$A = \begin{bmatrix} A_1 \\ \dots \\ A_p \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \dots \\ y_p \end{bmatrix} \quad (10)$$

因此,所有代理节点的测量数据重构问题可以表示为优化问题(5),通过利用分布式数据重构算法可以对其求解,从而保证所有代理节点均可同时获得原始信号 x .

3 基于 Top- $|K|$ 查询的分布式数据重构方法

基于 Top- $|K|$ 查询的分布式数据重构方法由两个算法组成:分布式迭代硬阈值算法和基于双阈值的分布式 Top- $|K|$ 查询算法.其中,分布式迭代硬阈值算法以分布式方式实现迭代硬阈值计算,基于双阈值的分布式 Top- $|K|$ 查询算法用于分布式实现前一算法中的查询绝对值最大的前 K 项元素和的操作,即以分布式方式实现式(9)中的硬阈值操作 $H_K()$.

3.1 分布式迭代硬阈值算法

每个代理节点首先令原始信号的初值 $x^{(0)}$ 为零向量,迭代次数 $t=0$.然后分别利用自身已知的感知矩阵 A_i 和收集到的测量信号 y_i 计算迭代运算的中间结果,执行计算:

$$z_i^{(t)} = \begin{cases} x^{(t)} + A_i^T (y_i - A_i x^{(t)}), & i=1 \\ A_i^T (y_i - A_i x^{(t)}), & i \neq 1 \end{cases} \quad (11)$$

其中,管理节点计算 $x^{(t)} + A_i^T (y_i - A_i x^{(t)})$,成员节点计算 $A_i^T (y_i - A_i x^{(t)})$.随后,管理节点执行基于双阈值的分布式 Top- $|K|$ 查询算法,以分布式方式实现硬阈值操作 $H_K()$,即管理节点计算:

$$x^{(t+1)} = H_K \left(\sum_{i=1}^p z_i^{(t)} \right) \quad (12)$$

最后,管理节点将本轮迭代所得计算结果 $x^{(t+1)}$ 发送至各个成员节点,所有代理节点继续执行下一轮迭代硬阈值计算.

由式(11)和式(12)可知,每轮迭代操作所得计算结果 $x^{(t+1)}$ 与式(8)相等,即分布式迭代硬阈值算法与迭代硬阈值算法等价.经过若干次迭代后,两个算法都将收敛至优化问题(5)的最优解.分布式迭代硬阈值算法的伪代码见算法 1.

算法 1. 分布式迭代硬阈值算法.

输入:感知矩阵 A_i ,测量信号 y_i ,稀疏度 K ,误差 ε ;

输出:原始信号 x .

1. 初始化: $t \leftarrow 0, x^{(0)} \leftarrow 0$
2. Do
3. 所有代理节点按照式(11)计算 $z_i^{(t)}$
4. 管理节点执行 Top- $|K|$ 查询,得到 $x^{(t+1)}$
5. 管理节点将 $x^{(t+1)}$ 传输至成员节点
6. $t \leftarrow t+1$
7. While $\|x^{(t+1)} - x^{(t)}\|_2 < \varepsilon$

3.2 基于双阈值的分布式 Top- $|K|$ 查询算法

基于双阈值的分布式 Top- $|K|$ 查询算法以分布式方式实现硬阈值操作 $H_K()$.与经典的 Top- K 查询算法仅查询值最大的前 K 项元素和不同,Top- $|K|$ 查询需要在分布式系统中实现对绝对值最大的前 K 项元素和进行查询

操作.该算法的基本思想是通过在管理节点与成员节点之间进行 3 次数据交换,从而实现分布式系统中全部元素集合的有效剪枝操作.整个算法包括双阈值计算阶段、候选集计算阶段、Top-|K|查询阶段.其中,双阈值计算阶段用于估算 Top-|K|项元素和的取值范围,计算正阈值 \bar{T} 与负阈值 \underline{T} .候选集计算阶段用于精确计算 Top-|K|项元素和的上下界,构建候选集 S .Top-|K|查询阶段用于最终确定绝对值最大的前 K 项元素和.

每个代理节点 i 在执行算法前需要将分布式迭代硬阈值算法中计算得到的中间结果 $z_i^{(l)}$ 按降序排列,得到降序列表 $L_i = \{(j, v_i(j)), j=1, \dots, n_i\}$, 其中 j 为元素索引, $v_i(j)$ 为代理节点 i 中与索引 j 相对应的元素值, n_i 为代理节点 i 中包含元素的个数.

算法在执行过程中,需要计算代理节点中候选元素的全部和、部分和、全部和上界、全部和下界.其中,元素全部和为

$$V(j) = \sum_{i=1}^p v_i(j) \quad (13)$$

元素部分和为

$$P(j) = \sum_{i=1}^p v_i'(j), v_i'(j) = \begin{cases} v_i(j), v_i(j) \text{ 已知} \\ 0, v_i(j) \text{ 未知} \end{cases} \quad (14)$$

元素全部和上界为

$$U(j) = \sum_{i=1}^p \bar{v}_i(j), \bar{v}_i(j) = \begin{cases} v_i(j), v_i(j) \text{ 已知} \\ \bar{T}, v_i(j) \text{ 未知} \end{cases} \quad (15)$$

元素全部和下界为

$$L(j) = \sum_{i=1}^p \underline{v}_i(j), \underline{v}_i(j) = \begin{cases} v_i(j), v_i(j) \text{ 已知} \\ \underline{T}, v_i(j) \text{ 未知} \end{cases} \quad (16)$$

元素部分和用于估算 Top-|K|项元素和的正、负阈值,通过两次过滤操作可以排除部分不满足条件的元素集合.元素全部和上界与下界用于与元素部分和比较,从而构建满足条件的候选集 S .元素全部和用于最终确定 Top-|K|项元素和.在计算以上元素和的过程中,管理节点遵循与成员节点相同的元素选取规则.基于双阈值的分布式 Top-|K|查询算法的伪代码见算法 2.管理节点与成员节点之间的 3 次交互过程如图 2 所示.

算法 2. 基于双阈值的分布式 Top-|K|查询算法.

输入:降序列表 $L_i = \{(j, v_i(j)), j=1, \dots, n_i\}$;

输出:绝对值最大的前 K 个元素和.

1. 阶段 1:双阈值计算阶段

1.1. If ($i \neq 1$) Then

1.2. 向管理节点发送前 K 个正元素与后 K 个负元素集合

1.3. Else

1.4. 计算所有接收元素的部分和

1.5. $\bar{\tau}_1 \leftarrow$ 前 K 个正元素部分和的下界

1.6. $\underline{\tau}_1 \leftarrow$ 后 K 个负元素部分和的上界

1.7. 向成员节点发送正阈值 $\bar{T} = \bar{\tau}_1 / p$ 与负阈值 $\underline{T} = \underline{\tau}_1 / p$.

2. 阶段 2:候选集计算阶段

2.1. If ($i \neq 1$) Then

2.2. 向管理节点发送 $\geq \bar{T}$ 的正元素或 $\leq \underline{T}$ 的负元素集合

2.3. Else

2.4. 计算所有接收元素的部分和

2.5. $\bar{\tau}_2 \leftarrow$ 前 K 个正元素部分和的下界

2.6. $\underline{\tau}_2 \leftarrow$ 后 K 个负元素部分和的上界

- 2.7. 计算所有接收元素的全部和上界与下界
- 2.8. 构建候选集 $S = \{ \text{全部和上界} \geq \bar{\tau}_2 \text{ 或全部和下界} \leq \underline{\tau}_2 \text{ 的元素} \}$
- 2.9. 向成员节点发送候选集 S
3. 阶段 3: Top- $|K|$ 查询阶段
 - 3.1. If ($i \neq 1$) Then
 - 3.2. 向管理节点发送候选集 S 中的元素集合
 - 3.3. Else
 - 3.4. 计算候选集 S 中所有元素的全部和并选取绝对值最大的前 K 项元素和

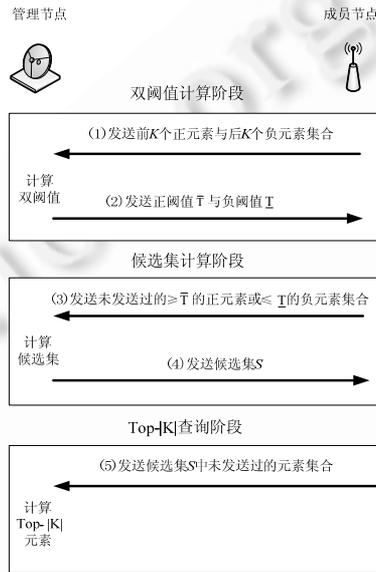


图 2 交互过程

3.3 基于双阈值的分布式 Top- $|K|$ 查询算法实例

本节通过实例简要说明基于双阈值的分布式 Top- $|K|$ 查询算法的计算过程.实例中管理节点与成员节点所包含的元素集合见表 1,各节点中的元素已按照值降序排列.本实例中 $p=3, K=2$,即查询 Top- $|2|$ 项元素.Top- $|K|$ 算法的计算结果见表 2.

表 1 Top- $|K|$ 算法实例

管理节点	成员节点 1	成员节点 2
(1,10)	(3,10)	(6,12)
(4,9)	(1,8)	(1,8)
(3,3)	(5,3)	(5,7)
(5,1)	(7,1)	(7,2)
(6,-1)	(6,-1)	(4,1)
(9,-2)	(8,-3)	(9,-1)
(8,-6)	(4,-5)	(2,-6)
(2,-9)	(2,-10)	(3,-7)

表 2 Top- $|K|$ 算法计算结果

阶段	$\bar{\tau}_1$	$\underline{\tau}_1$	$\bar{\tau}_2$	$\underline{\tau}_2$	\bar{T}	\underline{T}	Top- $ K $ 集合
1	12	-6			4	-2	
2			12	-9			
3							{(1,26), {2,-25}}

在双阈值计算阶段,由于 $v_1(1)=10, v_2(1)=8, v_3(1)=8$ 均已知,因此部分和 $P(1)=10+8+8=26$.在计算 $P(3)$ 时,由于管理节点中的(3,3)不属于前 2 个正元素与后 2 个负元素集合,因此根据式(14)令 $v_1(3)=0$.由于 $v_2(3)=10, v_3(3)=-7$,因此部分和 $P(3)=0+10-7=3$.其余元素部分和的计算方式与此类似.计算可得正元素部分和上界 $\bar{\tau}_1=12$,负元素部分和上界 $\underline{\tau}_1=-6$,正阈值 $\bar{T}=12/3=4$,负阈值 $\underline{T}=-6/3=-2$.

在候选集计算阶段,元素部分和的计算方式与前一阶段类似.在计算元素全部和上界时,由于 $v_1(1)=10, v_2(1)=8, v_3(1)=8$ 均已知,因此全部和上界 $U(1)=10+8+8=26$.在计算 $U(3)$ 时,由于管理节点中的(3,3)不符合 $\geq \bar{T}$ 或 $\leq \underline{T}$ 的条件,因此根据式(15)令 $\bar{v}_1(3)=4$.由于 $v_2(3)=10, v_3(3)=-7$,因此全部和上界 $U(3)=4+10-7=7$.其余元素的全部和上界与下界的计算方式与此类似.计算可得 $\bar{\tau}_2=12, \underline{\tau}_2=-9$,候选集 $S=\{1,6,5,8,2\}$.

在 Top-|K|查询阶段,管理节点计算候选集 S 中所有元素的全部和,并选取 Top-|2|项元素,可得 $\{(1,26),(2,-25)\}$.

4 实验结果与分析

本文仿真实验主要考虑基于 Top-|K|查询的分布式数据重构方法中重构误差、数据传输量、数据计算总量与网络中传感器节点的个数、代理节点的个数、原始信号的采样率以及监测环境状态之间的关系.仿真使用 Matlab 软件,利用多个不相关的二维高斯分布叠加模拟无线传感器网络中具有空间关联性的数据源.在 $400m \times 400m$ 的监测区域内随机部署 50 个~200 个无线传感器节点采集所监测的数据,代理节点随机选取 4 个~6 个,压缩感知矩阵采用随机高斯矩阵.每次实验重复运行 500 次后取平均值,作为该项实验的计算结果.

4.1 方案性能比较

本小节比较迭代硬阈值算法、基于 ATA(absolute threshold algorithm)查询的分布式数据重构方法^[18]以及基于 Top-|K|查询的分布式数据重构方法的各项性能.其中,基于 ATA 查询的分布式数据重构方法采用分布式计算绝对阈值的方法查询绝对值最大的前 K 项元素和,代理节点之间需要进行多次交互.每次交互过程中多个代理节点之间传递与同一索引相对应的元素值,传输效率较低,并且无法提前确定代理节点之间所需交互的次数.

图 3 展示了 3 种方法的数据重构性能,其中代理节点个数为 5,数据源个数为 3,原始信号采样率为 50%.图中利用重构信号的信噪比

$$SNR = 20 \cdot \log_{10} \left(x / |x - \tilde{x}| \right) \tag{17}$$

来衡量方法的重构误差,式(17)中 $x \in \mathbb{R}^n$ 为原始信号, $\tilde{x} \in \mathbb{R}^n$ 为重构信号.从图 3 中可以看出,3 种方法的数据重构性能没有明显差异,即分布式迭代硬阈值算法与集中式迭代硬阈值算法的数据重构性能相当.

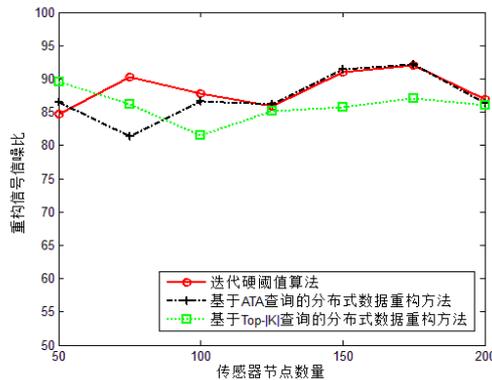


图 3 3 种方法的重构性能比较

图 4 和图 5 分别展示了与上图相同条件下 3 种方法中管理节点与成员节点之间的平均交互次数以及传输数据量.其中,迭代硬阈值算法中成员节点需要一次性将所有数据传输至管理节点,即迭代硬阈值算法中交互

次数为 1,传输数据量为全部数据.从图 4 和图 5 中可以看出,基于 Top-|K|查询的分布式数据重构方法中管理节点与成员节点之间需要进行 3 次数据交互,远低于基于 ATA 查询的分布式数据重构方法,因此产生的延迟较低.同时,该方法在数据重构过程中传输的数据量最少,因此可以有效地减少网络中代理节点的能量消耗.

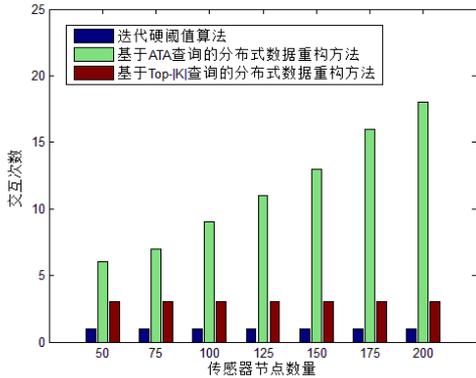


图 4 3 种方法的交互次数比较

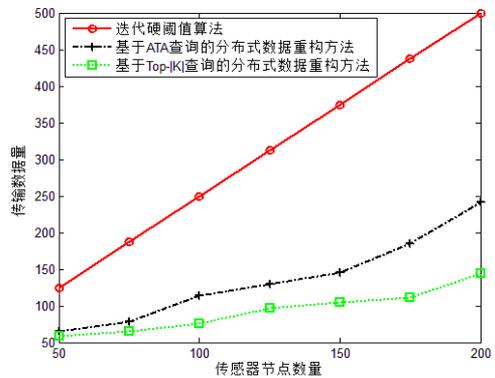


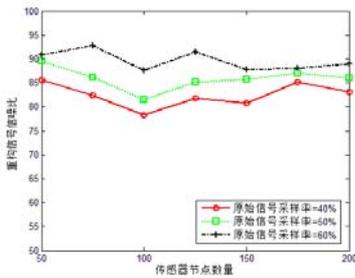
图 5 3 种方法的传输数据量比较

由以上性能比较可知,本文提出的基于 Top-|K|查询的分布式数据重构方法在提供与现有方法相同的数据重构性能前提下,通过代理节点之间的有限次交互,能够有效地降低网络中传输的数据量.

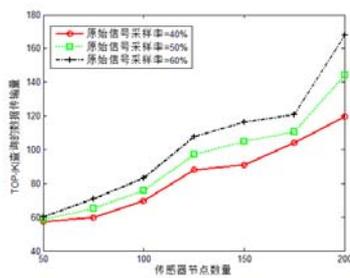
4.2 方案性能分析

本节分析基于 Top-|K|查询的分布式数据重构方法的各项性能,主要包括数据重构性能、传输数据量以及数据计算总量 3 项指标.其中,数据计算总量是指在基于双阈值的分布式 Top-|K|查询算法中,管理节点在 3 个阶段中累计计算的元素总量,用于衡量算法的计算复杂度.

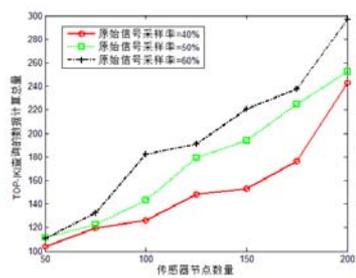
图 6(a)~图 6(c)分别展示了在原始信号采样率分别为 40%、50%和 60%的条件下,利用基于 Top-|K|查询的分布式数据重构方法重构时产生的重构信号信噪比、数据传输量以及数据计算总量.从该图中可以看出,随着原始信号采样率的增加,本文所提方法的数据重构性能逐渐提升,同时,代理节点之间的数据传输量和数据计算总量也逐渐增加.因为信号采样率的增加势必会造成传输和计算数据量的增加,进而可以达到更好的数据重构效果.



(a) 重构性能比较



(b) 数据传输量比较



(c) 数据计算总量比较

图 6 不同信号采样率条件下的性能比较

图 7(a)~图 7(c)分别展示了在随机数据源个数分别为 2、3、4 时,即不同监测环境状态下,利用基于 Top-|K|查询的分布式数据重构方法重构时产生的重构信号信噪比、数据传输量以及数据计算总量.从该图中可以看出,随着随机数据源个数的增加,本文所提方法的数据重构性能逐渐降低,同时,代理节点之间的数据传输量和数据计算总量也逐渐增加.因为数据源个数的增加意味着监测数据集的稀疏度增加,进而会影响数据的重构性能,同时也需要传输和计算更多的数据以实现原始信号的重构.

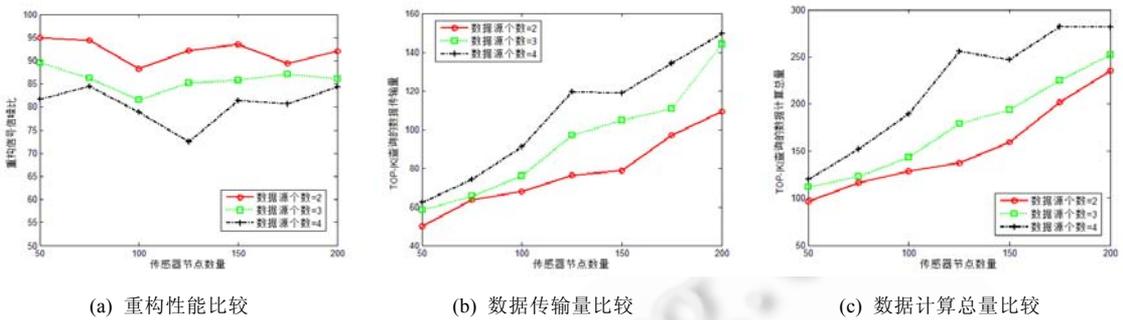


图 7 不同监测环境状态下的性能比较

图 8(a)~图 8(c)分别展示了在代理节点个数分别为 4、5、6 的条件下,利用基于 Top-|K|查询的分布式数据重构方法重构时产生的重构信号信噪比、数据传输量以及数据计算总量.从该图中可以看出,随着代理节点个数的增加,本文所提方法的数据重构性能没有显著差异,同时,代理节点之间的数据传输量和数据计算总量逐渐增加.因为代理节点的个数与方案的重构性能无关,但是更多的代理节点意味着数据重构时需要传输和计算更多的数据.

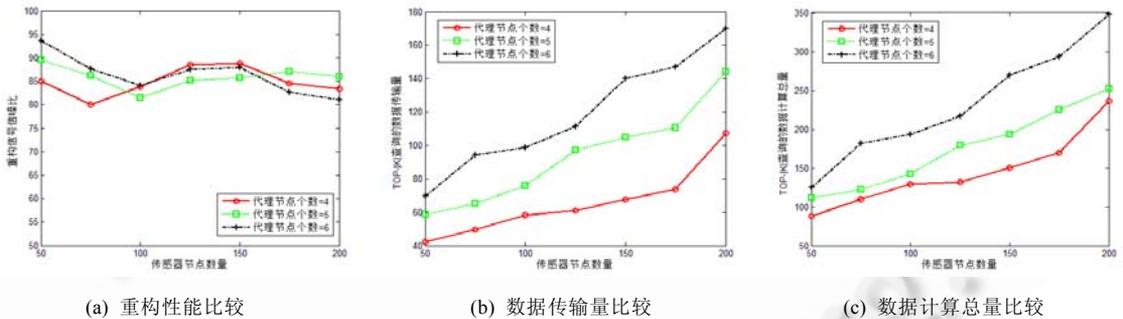


图 8 不同代理节点个数条件下的性能比较

5 结束语

本文针对分簇结构或多 Sink 节点的无线传感器网络中的数据收集问题,基于压缩感知理论和迭代硬阈值算法,设计了一种分布式的采集数据重构方法.其中,管理节点和成员节点以分布式方式执行迭代硬阈值计算,并通过构建正、负阈值和候选集,从而排除部分不满足条件的元素集合,以分布式方式实现对绝对值最大的前 K 项元素和的查询操作.在仿真实验中,通过与迭代硬阈值算法和基于 ATA 查询的分布式数据重构方法比较,验证了文中所提方法的重构性能和通信开销.同时,实验部分还对所提方法的数据重构性能、传输数据量和数据计算总量进行了详细的分析.下一步研究方向将着重考虑结合传感器节点采集数据的时间关联性,实现分布式时空数据重构.

References:

- [1] Jiang HB, Jin SD, Wang CG. Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks. *IEEE Trans. on Parallel and Distributed Systems*, 2011,22(6):1064–1071.
- [2] Villas LA, Boukerche A, Guildoni DL, de Oliveira HABF, de Araujo RB, Loureiro AAF. An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks. *Computer Communications*, 2013,36(9): 1054–1066.
- [3] Li GR, Wang Y. Automatic ARIMA modeling-based data aggregation scheme in wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2013,85.

- [4] Cheng ST, Shih JS, Chou CL, Horng GJ, Wang CH. Hierarchical distributed source coding scheme and optimal transmission scheduling for wireless sensor networks. *Wireless Personal Communications*, 2013,70(2):847–868.
- [5] Qin ZC, Zhou Z, Zhao XC. A distributed source coding algorithm for clustering wireless sensor networks. *Journal of Electronics & Information Technology*, 2013,35(2):328–334 (in Chinese with English Abstract).
- [6] Dai QH, Fu CJ, Ji XY. Research on compressed sensing. *Chinese Journal of Computers*, 2011,34(3):425–434 (in Chinese with English Abstract).
- [7] Eldar Y, Kutyniok G. *Compressed sensing*. Cambridge: Cambridge University Press, 2012. 348–394.
- [8] Xiang L, Luo J, Rosenberg C. Compressed data aggregation: Energy efficient and high fidelity data collection. *IEEE Trans. on Networking*, 2013,21(6):1722–1735.
- [9] He FX, Yu ZJ, Liu HT. Multiple target localization via compressed sensing in wireless sensor networks. *Journal of Electronics & Information Technology*, 2012,34(3):716–721 (in Chinese with English Abstract).
- [10] Pudlewski S, Melodia T, Prasanna A. Compressed-Sensing-Enabled video streaming for wireless multimedia sensor networks. *IEEE Trans. on Mobile Computing*, 2012,11(6):1060–1072.
- [11] Zhao XC, Zhou Z, Qin ZC. Multi-Hop routing protocol based on double cluster head alternation and compressed sensing for wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2012,23:17–24 (in Chinese with English abstract).
- [12] Luo C, Wu F, Sun J, Chen CW. Efficient measurement generation and pervasive sparsity for compressive data gathering. *IEEE Trans. on Wireless Communications*, 2010,9(12):3728–3738.
- [13] Wang J, Tang SJ, Yin BC, Li XY. Data gathering in wireless sensor networks through intelligent compressive sensing. In: Greenberg A, ed. *Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM 2012)*. Orlando: IEEE Press, 2012. 603–611.
- [14] Xu LW, Qi X, Wang YX, Moscibroda T. Efficient data gathering using compressed sparse functions. In: Marsan M, ed. *Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM 2013)*. Turin, IEEE Press, 2013. 310–314.
- [15] Lu W, Liu YZ, Wang DS. A distributed secure data collection scheme via chaotic compressed sensing in wireless sensor networks. *Circuits System and Signal Processing*, 2013,32(3):1363–1387.
- [16] Xu ZQ. Compressed sensing: A survey. *Scientia Sinica Mathematica*, 2012,42(9):865–877. (in Chinese with English abstract)
- [17] Blumensath T, Davies M. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 2009, 27(3):265–274.
- [18] Patterson S, Eldar Y, Keidar I. Distributed sparse signal recovery for sensor networks. In: Ward R, ed. *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP'13)*. Vancouver: IEEE Press, 2013. 4494–4498.

附中中文参考文献:

- [5] 秦智超,周正,赵小川.一种分簇无线传感器网络中的分布式信源编码算法. *电子与信息学报*,2013,35(2):328–334.
- [6] 戴琼海,付长军,季向阳.压缩感知研究. *计算机学报*,2011,34(3):425–434.
- [9] 何风行,余志军,刘海涛.基于压缩感知的无线传感器网络多目标定位算法. *电子与信息学报*,2012,34(3):716–721.
- [11] 赵小川,周正,秦智超.基于双簇头交替和压缩感知的 WSN 路由协议. *软件学报*,2012,23:17–24.
- [16] 徐志强.压缩感知. *中国科学*,2012,42(9):865–877.



李国瑞(1980—),男,山西夏县人,博士,副教授,主要研究领域为无线传感器网络,压缩感知.
E-mail: lgr@mail.neuq.edu.cn