

参与式感知系统中基于压缩感知的数据采集算法

于瑞云, 周岩

(东北大学 软件学院, 辽宁 沈阳 110819)

摘 要: 基于压缩感知理论提出了一种在参与式感知系统中进行数据采集的算法. 该算法通过对节点社会关系的分析, 估计出部分未被传输的节点感知数据, 在此基础上对观测矩阵进行更新, 使压缩感知算法可以利用已传输的数据和估计得出的数据进行重构. 该算法能显著减少参与式感知系统中传输的数据量, 同时能够保证较好的数据重构精度. 采用随机漫步移动模型进行了仿真实验, 验证了算法的可行性. 实验表明, 与传统的压缩感知算法相比, 上述算法在重构成功率相同的情况下, 可以显著减少网络传输的数据量, 从而降低网络消耗.

关键词: 参与式感知; 压缩感知; 社会关系; 数据采集; 数据估计

中图分类号: TP 393.0 文献标志码: A 文章编号: 1005-3026(2015)02-0194-05

Compressed Sensing Based Data Acquisition Algorithm in Participatory Sensing System

YU Rui-yun, ZHOU Yan

(School of Software, Northeastern University, Shenyang 110819, China. Corresponding author: YU Rui-yun, associate professor, E-mail: yury@mail.neu.edu.cn)

Abstract: A data acquisition algorithm in participatory sensing systems based on the compressed sensing theory was proposed. In this algorithm, the un-transmitted data was estimated by analyzing social relationship between mobile nodes. And then the observation matrices were refreshed using estimated data. Finally the compressed sensing algorithm was exploited to reconstruct original data according to both transmitted and estimated data. The proposed algorithm could greatly reduce the amount of data transmitted in the participatory sensing systems while still achieve good data reconstruction accuracy. The random walk mobility model was exploited in the simulations to validate the feasibility of this algorithm. The simulation results showed that, compared with the traditional compressed sensing algorithm, the amount of data transmitted over the network could be remarkably reduced without losing data fidelity, and hence the network overhead could be decreased.

Key words: participatory sensing; compressed sensing; social relationship; data acquisition; data estimation

参与式感知(participatory sensing)也称城市感知(urban sensing),是人们利用集成了特定传感器的智能移动终端对人类社会状态信息(包括环境、交通、社会活动等)进行采集、分类、传输和分析,进而作出智能决策,为人类生活和社会活动提供服务^[1-2]的一项新技术.

参与式感知应用通常需要传输大量的感知数

据,因此对数据进行压缩进而缩减数据冗余是参与式感知技术中的关键内容.传统的数据压缩方法有信源编码、分布式压缩感知等,是根据当前信号内容相关性进行数据压缩的方法.分布式信源编码(distributed source coding, DSC)从信息论中的信源编码角度实现低复杂度、分布式的信息压缩^[3].压缩感知(compressed sensing, CS)是由

Donoho^[4]提出的数据压缩算法. 该算法指出, 只要信号可以在某种基下稀疏表示, 就可以在不丢失逼近原信号所需信息的情况下, 用较少的观测数据对原始信号进行投影, 然后通过求解最优化问题重构原始信号. Baron 等^[3,5]提出了分布式压缩感知 (distributed compressed sensing, DCS) 理论, 该理论结合多信号的联合稀疏模型 (joint sparsely model, JSM) 和压缩感知理论提出了一种利用信号间以及信号内相关性来进行信号压缩编码的算法.

目前, 国内外学者已经对参与式感知^[6-7]及压缩感知在参与式感知系统中的应用^[2]进行了相关研究. 移动设备的续航能力和网络资源受限问题一直是限制参与式感知系统发展的一个瓶颈. 参与式感知系统中, 节点在感知区域内按照一定规律分布, 在节点较多的情况下, 节点所传输的数据具有一定的冗余性; 如何减少冗余信息的传输, 降低系统资源消耗, 成为参与式感知系统研究的一个主要问题.

本文主要研究参与式感知系统 (participatory sensing system, PSS) 中的数据采集算法, 并提出一种基于压缩感知的数据采集算法 (compressed sensing based data acquisition algorithm, CSDA). CSDA 的主要创新是, 在参与式感知节点较多的情况下, 通过挖掘并应用节点间的社会关系对节点感知数据的数值进行估计, 从而进一步减少压缩感知需要传输数据的数量, 节约网络能量消耗.

1 基于压缩感知的数据采集算法

CSDA 主要分为数据发送和数据处理两部分: 在数据发送端, 簇头节点利用 CSDA 算法通过对原始数据的压缩感知随机采样, 得到压缩后的采样数据并将其传输到数据处理端; 数据处理端在采样数据的基础上利用节点间的社会关系对其他未传输数据进行估计, 然后利用压缩感知的数据重构算法对处理后的数据进行重构, 完成对原始数据的恢复.

1.1 压缩感知随机采样

PSS 中的节点具有移动性, 这些节点在同一空间中的不同时刻呈现不同的分布, 而同一节点在不同时刻呈现出按一定规律移动的社会性.

在 PSS 中数据传输过程需要分簇路由协议等数据传输算法的支持. 各分布节点将全部数据 X 传给簇头节点 A, 簇头节点 A 再根据压缩感知算法, 对数据观测采样后将采样数据 Y 传输给

Sink 节点 B, Sink 节点 B 对数据进行重构. 数据的采集过程需要根据地点进行划分, 具体划分方法如定义 1 所述.

定义 1 节点的活动范围可以划分为若干个事件区域 (event area, EA), 并将 EA 划分为 N 个感知网格 (sensing grid, SG).

如图 1 所示, 每个感知网格相当于一个信号源, SG 的集合可表示为

$$\text{Grid} = \{\text{grid}_n\}, n \in \{1, 2, \dots, N\}. \quad (1)$$

则系统随机采集过程表述如下:

设系统采集数据周期数为 D , 每个周期内采集 T 次, $t \in \{1, 2, \dots, T\}$ 表示采集时刻的序数. 令 X_d 为第 d 个周期 $N \times T$ 的数据矩阵, 矩阵中的元素 x_m^d 表示 grid_n 在周期 d 中第 t 时刻的原始数据值.

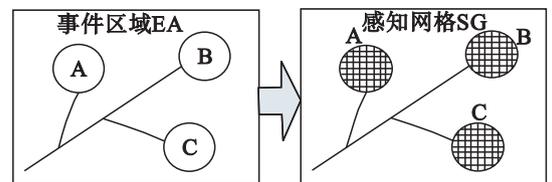


图 1 事件区域与感知网格
Fig. 1 Event area and sensing grid

根据压缩感知算法理论可知, 观测矩阵 Φ 一般选取高斯随机矩阵, 其中高斯随机矩阵决定了在获得观测向量前, 获取全部网络节点的数据. PSS 中的节点多为私人设备, 为了保证用户设备的电池使用寿命, 可能无法将节点完全实时开放, 因此本文采用置换矩阵 (permutation matrix) 作为观测矩阵, 该矩阵可以在节点不需要传输数据时不激活该节点, 从而节省节点能量. 令列向量 $x_n^d = [x_{1n}^d, x_{2n}^d, \dots, x_{Tn}^d]^T$ 为网格 grid_n 在所有时刻的原始数据向量, Φ_n^d 为 grid_n 的观测矩阵, Ψ 为 x_n^d 的稀疏基并且满足 $x_n^d = \Psi \Theta_n^d$, Θ_n^d 为 x_n^d 的稀疏向量, 则对于 grid_n 的观测过程如式 (2) 所示:

$$y_n^d = \Phi_n^d x_n^d = \Phi_n^d \Psi \Theta_n^d. \quad (2)$$

其中 Φ_n^d 为 $T^* \times T$ ($T^* < T$) 的置换矩阵, 所有 SG 的观测矩阵组成观测矩阵集合 $\Phi = \{\Phi_n^d\}$, $n \in \{1, 2, \dots, N\}$, $d \in \{1, 2, \dots, D\}$, 每个周期 Sink 节点为每个 SG 分配不同的观测矩阵. 若令第 d 个周期的采样数据矩阵为 Y_d , 则式 (2) 中列向量 $y_n^d = [y_{1n}^d, y_{2n}^d, \dots, y_{T^*n}^d]^T$ 表示网格 grid_n 在 T^* 个时刻的采样数据向量.

图 2 为一个采样过程的示例, 由于置换矩阵的特性, 数据在传输到 Sink 节点后的时刻顺序是随机的, 但其数据值不变.

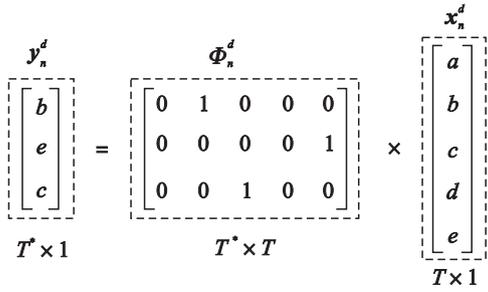


图 2 压缩感知随机观测过程示例

Fig. 2 Random observation process based on CS

在数据传输过程中,簇头节点通过社会关系减少了需要传输的数据量,并将采样后的数据发送给接收端,从而显著降低了网络负载.同时节点在传输数据时要记录所采集数据的节点信息和数据所属时刻,以便进行数据集的转换以及数据的重构.

1.2 基于社会关系的数据重构

CSDA 在数据重构过程中利用了节点间的社会关系:在重构之前通过构建社会关系矩阵对感知数据集进行更新,使其数据量达到压缩感知重构的最小数据量要求.

在事件区域(EA)中,PSS 中的节点在参与社会活动时所产生的移动轨迹可以反映节点的社会性质,其中移动轨迹相似的节点具有一定的社会关系.同时环境数据(如温度、湿度、PM2.5 浓度等)具有空间相关性,在不考虑欺骗节点的情况下,节点在其移动轨迹上所采集的数据反映了节点移动轨迹的特性.节点所采集数据的相关性可以作为社会关系的量化标准.

这些数据间的相关性是节点活动轨迹相似度的直接反映,通过数据挖掘方法^[5]对历史数据分析计算可得到相关性的各项系数,从而利用这种相关性进行有效的数据重构.

不同时刻节点可能存在不同的社会关系,因此,本文将分别维护不同时刻的数据相关性系数(即节点的社会关系系数).某个时刻的数据相关性系数可以根据多个连续周期该时刻的节点采集数据的期望和方差来衡量.

方差越趋于 0,节点间的社会关系系数越高,节点的关系越亲密;节点与自身的社会关系系数为 0.两个节点的数据趋势如图 3a 所示;两节点数据的差的分布如图 3b 所示,其中虚线为数据差的期望,方差为 0.4,可以认为两节点的社会关系协同性较高.方差过大的节点间社会关系系数较小,其数据相关性可忽略不计.设阈值 λ ,若社会关系系数大于 λ ,则令社会关系系数为 ∞ ,因此社

会关系系数的正常取值范围为 0 到 λ 之间.

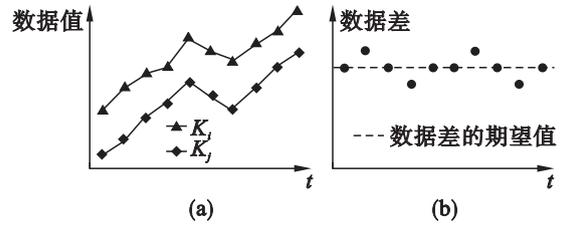


图 3 社会关系协同性

Fig. 3 Social relationship cooperativity

- (a) —节点 K_i, K_j 的数据趋势;
- (b) —节点 K_i, K_j 数据差的分布.

在 CSDA 中,可以认为带有数据的 SG 的数量小于系统中节点的数量.在节点和 SG 足够多的情况下,节点的移动在各个时刻可以基本覆盖所有的 SG,则通过数据估计可以满足压缩感知算法重构信息的最小要求.若在一个 SG 中出现多个节点,则传输信用度较高的节点的数据.

在周期 d 中,对于 $grid_n$ 有 T 个时刻需要采集数据,传统的采集方法需要全部 T 个时刻的感知数据来获取感知区域的信息.而压缩感知(CS)算法可能仅利用最少 M ($M < T$) 个时刻的数据即可恢复 T 个时刻的数据.本文提出的基于压缩感知的数据采集算法(CSDA)在收集 T^* ($T^* < M$) 个数据的情况下,利用节点的社会关系估算 $M - T^*$ 个未传输数据时刻的节点感知数据,从而达到 CS 重构所需的最小数据量.未传输的 $M - T^*$ 个数据利用节点社会关系加权估计得出.

假设在时刻 t ,与节点 k 具有社会关系的节点共有 S 个,则节点 k 与节点 s 的权重计算过程如式(3)所示.节点 k 的数据通过对 S 个节点数据加权平均获得.

$$W_s = \frac{\lambda - q_{ks}^t}{\sum_{s=1}^S (\lambda - q_{ks}^t)} \quad (3)$$

式中 W_s 为权重值, q_{ks}^t 为两节点之间社会关系系数.

压缩感知数据重构需要求解范数最小化问题^[4],同时要用估算后的数据集对采样数据集和观测矩阵进行更新,以满足重构的需要.更新过程如下.

对于周期 d 中 $grid_n$ 的感知数据 $y_n^d = [y_{1n}^d, y_{2n}^d, \dots, y_{T^*n}^d]^T$,若有 L 个数据可以经过数据估计得出,则更新后的感知数据可表示为 $\hat{y}_n^d = [y_{1n}^d, y_{2n}^d, \dots, y_{T^*n}^d, y_{(T^*+1)n}^d, \dots, y_{(T^*+L)n}^d]^T$,其中 $[y_{(T^*+1)n}^d, \dots, y_{(T^*+L)n}^d]^T$ 为 L 个估计得出的数据,同时要在观测矩阵 Φ_n^d 的最后一行下面添加 L

行;若第 l 行的数据是 t 时刻经过计算得出的数据,则在 Φ_n^d 添加的第 l 行中,第 t 列为 1,其余元素为 0,更新后的观测矩阵表示为 $\hat{\Phi}_n^d$. 更新过程如图 4 所示,图中 \hat{d} 表示估计得出的数据.

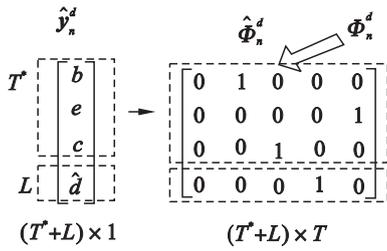


图 4 采样数据及观测矩阵更新过程

Fig. 4 Refresh process of sensing data and observation matrix

PSS 需要 Sink 节点对每个 SG 的数据分别进行重构,对于网格 $grid_n$,Sink 节点通过 1-范数最小化算法对 \hat{y}_n^d 进行精确重构:利用式 $\hat{y}_n^d = \hat{\Phi}_n^d \Psi \Theta_n^d$,得到 \hat{x}_n^d 的稀疏表示 $\hat{\Theta}_n^d$,然后根据式 $\Psi \hat{\Theta}_n^d = \hat{x}_n^d$,得到重构数据 \hat{x}_n^d .

2 性能分析

2.1 实验环境配置

为验证 CSDA 算法的性能,本文利用仿真实验生成移动路径及节点感知数据集,其中感知数据采用自由空间扩散模型(FSD)生成,利用随机漫步模型(RWM)^[6]生成移动路径.压缩感知算法中,稀疏矩阵可以选择 DCT 矩阵、FFT 矩阵或者小波变换矩阵.本文采用 DCT 矩阵作为稀疏矩阵,置换矩阵与 DCT 矩阵的约束等距性^[4-5]为 $\mu(\Phi, \Psi) \approx 2.453$,根据本文对压缩感知理论的介绍可知,可以对数据较精确地恢复.

本实验设置 SG 数量 $N = 173$,运行的周期数 $D = 31$,每个周期采集数据次数 $T = 50$,理论最大采集数据量为 $173 \times 50 = 8650$ 个.

同时,本实验利用前 30 次数据为已知数据集(也称训练集),利用 CSDA 算法计算第 31 次的数据,并与模拟得出的第 31 次数据进行比较,观察算法效果.重构误差可以用两种数据的差的绝对值与模拟数据的比来表示,则误差小于给定重构误差的重构数据可视为其重构成功,重构成功率为重构成功的数据量与总数据量之比.

2.2 不同重构误差下 CSDA 与 CS 算法的分析

给定不同的重构误差,分析成功率达到 95% 时两个算法所需要的最少数据量,如图 5 所示.

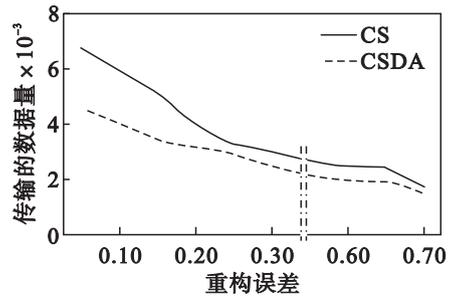


图 5 不同重构误差下 CSDA 与 CS 算法的比较
Fig. 5 Comparison of CSDA and CS with different recovery errors

图 5 表明,在满足重构成功率大于 95% 的情况下,CS 算法所需要的数据量大于 CSDA.而随着重构误差的扩大,两种算法的数据量差异缩小;重构误差越小,两种算法所需的数据量差异越大.

2.3 CSDA 在社会关系阈值不同的数据集下的效果分析

给定重构误差 0.2,并取阈值 λ 分别为 0.6, 0.7, 0.8, 0.9 进行仿真实验,结果如图 6 所示.

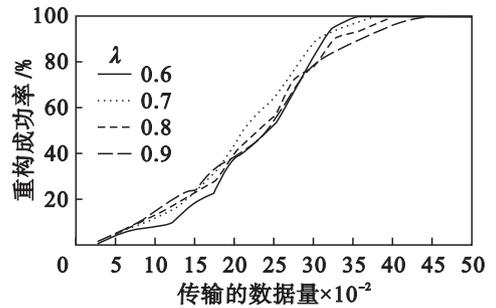


图 6 不同社会关系阈值重构成功率
Fig. 6 Success rate with different social relationship thresholds

图 6 表明, λ 越大,社会关系协同性越复杂,重构成功率达到 95% 以上需要传输数据的节点数越多,说明过于复杂的社会关系会导致节点估计的不准确;而在采集节点较少的情况下, λ 越大重构准确率越高.本实验说明 CSDA 需要合理选择社会关系阈值才能具备良好的效果.

2.4 不同移动节点数的 CSDA 效果分析

给定重构误差 0.2,设置移动模型中的节点数量 K 分别为 128, 150, 175, 200, 256, 300 进行仿真实验.对两种算法的重构成功率达到 95% 时所需要的最小数据量进行比较,结果见图 7,图 8.

图 7 表明,模型中节点越多,算法效果越好, $N = 300$ 时重构成功率达到 95% 以上需要的采集数据量为 2250.这是因为 CSDA 在移动节点较多时可以减少传输的数据量;在节点较稀疏时,应用 CSDA 的效果并不明显.

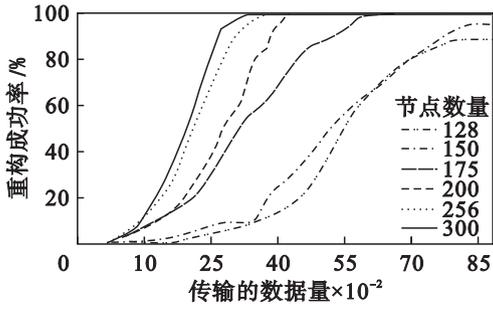


图 7 不同移动节点数的重构成功率

Fig. 7 Success rate with different number of mobile nodes

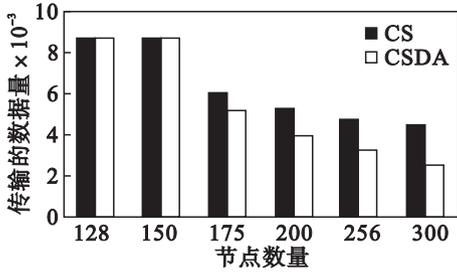


图 8 达到 95% 重构成功率所需的节点数量

Fig. 8 The number of mobile nodes in 95% success rate

图 8 表明,在节点数小于 SG 数量时,需要传输全部节点的数据量来获得全部信息;而在节点数大于 SG 数量的情况下,节点数量越多,获得 95% 以上重构成功率时 CSDA 需要传输的数据量越少. CS 算法所需传输的数据量没有明显变化,说明移动节点的增多使社会关系协同性发挥更大的作用.

3 结 语

本文提出了一种参与式系统中基于压缩感知

的数据采集算法(CSDA). 该算法将压缩感知算法与节点的社会关系结合起来,利用节点的社会关系,估计未被传输的节点感知数据,并对压缩感知重构算法进行了改进,减少了压缩感知算法在参与式感知系统中采集的数据量,进一步降低了网络的能量消耗. 实验验证,相比传统压缩感知方法,CSDA 只需传输较少的数据量就能达到理想的数据重构效果.

参考文献:

- [1] Kanhere S S. Participatory sensing: crowdsourcing data from mobile smart phones in urban spaces[C]// 2011 12th IEEE International Conference on Mobile Data Management. New York:IEEE,2011:3-6.
- [2] Akimura D, Kawahara Y, Asami T. Compressed sensing method for human activity sensing using mobile phone accelerometers[C]// 2012 Ninth International Conference on Networked Sensing Systems. New York:IEEE,2012:1-4.
- [3] Duarte M F, Sarvotham S, Baron D, et al. Distributed compressed sensing of jointly sparse signals [C]// Proceedings of the 39th Asilomar Conference on Signals, Systems and Computation. Pacific Grove,2005:1537-1541.
- [4] Donoho D L. Compressed sensing[J]. *IEEE Transactions on Information Theory*,2006,52(4):1289-1306.
- [5] Bettstetter C, Hartenstein H, Pérez-Costa X. Stochastic properties of the random waypoint mobility model [J]. *Wireless Networks*,2004,10(5):555-567.
- [6] Ramesh M V, Jacob A, Aryadevi R D. Participatory sensing for emergency communication via MANET [C]// 2012 International Conference on Data Science & Engineering. New York:IEEE,2012:181-186.
- [7] Kotovirta V, Toivanen T, Tergujeff R, et al. Participatory sensing in environmental monitoring—experiences [C]// 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. New York:IEEE,2012:155-162.