

基于海量物联网数据的压缩感知及其并行处理

周 健, 张永平, 张功萱

(南京理工大学 计算机学院, 江苏 南京 210094)

摘 要: 物联网是当前人们的研究热点, 本文提出使用压缩感知理论处理大规模的物联网中产生的海量数据. 压缩感知是一种能够在采样的同时实现数据压缩的采样方法, 它可以通过降低采样率显著减少采集的数据量, 但压缩感知算法的计算复杂度高、对信号的适应性差. 针对压缩感知方法的缺点, 本文尝试对压缩感知算法并行处理方法以提高压缩感知的计算速度, 同时引入冗余字典构造稀疏变换基以提高压缩感知对信号的适应性.

关键词: 海量数据; 压缩感知; 优化重构; 并行处理

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)11-0116-04

Research of Parallel Compressed Sensing for Mass Data in Internet of Things

ZHOU Jian, ZHANG Yong-ping, ZHANG Gong-xuan

(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Internet of Things is one of the most popular scientific and technical terminologies. In this paper, we use compressed sensing theory to process mass data in Internet of Things. Compressed sensing is a sampling method that data sampling and compressing can be done simultaneously. Compressed sensing can significantly lower the data size by reducing sampling rates of sensors, but its algorithm has high computational complexity and its transformation basis is nonadaptive. This paper puts forward parallel processing of compressed sensing algorithm for high computational complexity. At the same time, we introduce redundant dictionary into compressed sensing for increasing the flexibility.

Key words: mass data; compressed sensing; optimal recovery; parallel processing

1 引言

物联网的应用领域非常广泛, 每年产生的数据将达到 PB 量级. 海量数据的采集和处理, 将严重影响智能感知系统的寿命、物联网的数据处理和存储能力及其覆盖范围. 为了降低物联网的数据量, 使用压缩感知方法对数据进行处理是一个不错的选择.

2 压缩感知

压缩感知 (Compressed sensing or compressive sampling, CS) 作为一种新的采样理论, 它利用信号的稀疏特性, 以远小于传统的 Nyquist 频率对信号

进行采样, 得到压缩后的数据, 再利用数值优化算法重构原始信号^[1].

2.1 压缩感知理论的描述

假设一维线性原始信号 $X: N \times 1$ 在某组稀疏变换基 $\Psi: N \times N$ 上的转换系数是稀疏的. 那么 X 基于 Ψ 的转换系数 Θ 可以表示为:

$$\Theta = \Psi^T X \quad (1)$$

其中转换系数 Θ 是稀疏的, 这里所说的 Θ 是稀疏意思是矩阵 Θ 中的元素大部分是 0.

接着, 设计一个与稀疏变换基 Ψ 不相关的 $M \times N (M \ll N)$ 维观测矩阵 Φ , 对 Θ 进行观测得到线性观测集合 Y :

收稿日期: 2012-05-10; 修回日期: 2012-06-18

基金项目: 江苏省“九七三”项目 (BK2011022); 国家自然科学基金 (61170035); 南京理工大学重点基金 (2011YBXM18)

$$Y = \Phi\Theta = \Phi\Psi^T X \quad (2)$$

其中 $Y: M \times 1$ 就是压缩后的数据,表达式(2)矩阵表示如图1描述。

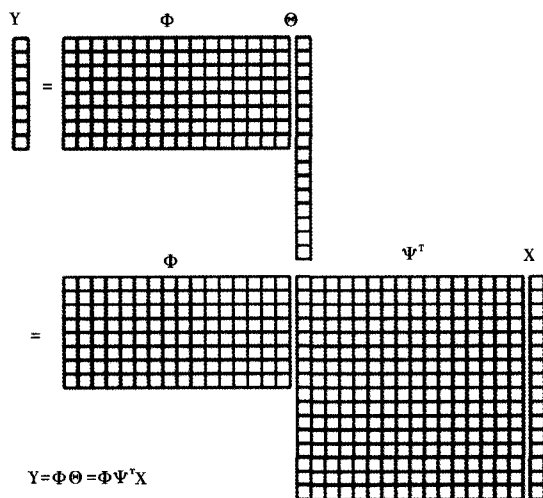


图1 压缩感知观测过程的矩阵表示

通过上面的描述可知,压缩感知理论的压缩就是使用变换基 Ψ 和度量矩阵 Φ 把 N 维信号 X 变换成 M 维的观测集合 Y ,从而实现了数据的压缩。根据压缩感知理论,当重构信号时,只需要求解:

$$\min \|\Theta\|_0 \text{ s.t. } \Phi\Theta = \Phi\Psi^T X = Y \quad (3)$$

其中 $\|\Theta\|_0$ 表示 X 中非零项的个数。压缩感知理论的计算过程如图2所示。

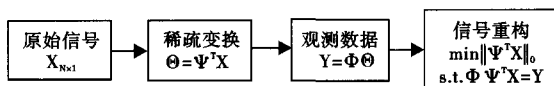


图2 压缩感知理论框架

2.2 压缩感知的关键问题

压缩感知理论的实现包含三个关键问题^[2]:

(1)稀疏表示:信号的稀疏表示问题可以归结于稀疏变换基的设计,对于某个信号 X ,如何找到简洁、直接、最佳的稀疏正交基 Ψ ,使 X 的表示是稀疏的,这就是信号的稀疏表示问题。压缩感知理论提出之初均假设稀疏变换基为标准正交基,但标准正交基不具有适应性,不能根据信号类型的不同而改变。Peyr^[3]把稀疏变换基的条件由标准正交基扩展到了由多个标准正交基构成的正交基字典。在正交基字典中,可以自适应地选择逼近某一种信号特征的最优正交基,使信号的稀疏表示具有了一定的适应性。对于某个信号 X ,如何找到简洁、直接、最佳的稀疏正交基 Ψ ,使 X 的表示是稀疏的,这就是信号的稀疏表示问题,也是压缩感知研究的一个热点。

(2)观测矩阵:方程 $Y = \Phi\Theta = \Phi\Psi^T X$ 是一个欠定的,一般很难求解,但是如果假设 $\Phi\Psi^T$ 具有有限等距性质(RIP)^[4],这个方程就可以求得确定的解。RIP性质的等价条件就是 Ψ 和 Φ 不相关。随机高斯矩阵^[5]、Rademacher 随机矩阵^[6]等随机矩阵与大多数固定正交基矩阵不相关,这一特性决定了选它作观测矩阵、其它正交基矩阵作稀疏变换基时, $\Phi\Psi^T$ 一般满足 RIP 性质,所以在应用中一般采用随机矩阵作观测矩阵。但是,目前普遍采用的观测矩阵都不能保证 100% 的精确重构信号,观测矩阵的设计仍然是一个很热门的研究课题。

(3)信号重构:重构信号是一个非线性优化问题,但在 0-范数意义下自表达式(3)中求得 X 的精确或近似逼近解 $\tilde{X}$,这是一个 NP 难问题,很难得到结果。然而,Donoho、Chen 和 Saunders 指出^[7]:当度量矩阵 Φ 与变换基 Ψ 不相关时,表达式(3)与表达式(4)有相同的解:

$$\min \|\Theta\|_1 \text{ s.t. } \Phi\Theta = \Phi\Psi^T X = Y \quad (4)$$

而表达式(4)是一个凸优化问题,可以很容易的化简为线性规划问题进行求解。

2.3 重构算法

当前,压缩感知的重构算法可以归为以下三类:

(1)贪婪算法:贪婪算法都是迭代的,每次迭代都选择一个局部最优解,从而不断逼近原始信号。匹配追踪(Matching Pursuit, MP)^[8]、分级正交匹配追踪(Stagewise OMP, StOMP)^[9]等都属于这类算法。

(2)凸松弛法:凸松弛法是最小 1-范式线性规划方法,它可以把非凸问题转化为凸问题求解。这类方法包括基追踪(Basis Pursuit, BP)^[7]等算法。

(3)组合算法:组合算法是通过分组测试实现原始信号的快速重构,它包括 Fourier 表示法(Fourier representations)^[10]等算法。

3 并行压缩感知算法架构

压缩感知是一种数据采集和数据压缩同时进行的数据采样方法,它的引入可以大大减少物联网中的数据采集量、减轻网路的通信负担、提高数据的实时性、增加智能感知系统的寿命,对物联网的发展有很大的促进作用。但压缩感知的稀疏变换基适应性差、算法的计算复杂度高。例如:采用 BP 算法对于长度为 8192 的信号进行分解,如利用小波字典,则等价于求解一个长度为 8192×212992 的线性规划问题^[7]。

3.1 并行算法架构

压缩感知算法的计算复杂度很高,但其算法涉及了大量的矩阵运算,比如信号的稀疏表示、数据的压缩和信号的重构等,这为压缩感知算法的并行处理带来了便利,算法的并行处理可以显著的提高计算速度,降低算法处理时间.而对于压缩感知算法稀疏变换基的非适应性问题,采用具有基选择能力的冗余字典是一个很好的尝试.

根据研究,为了适应处理物联网海量数据的需要,我们在本文中提出了一个并行压缩感知算法架构,如图 3 所示.

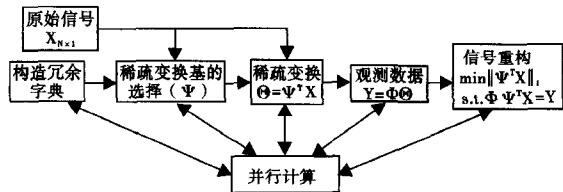


图 3 并行压缩感知算法架构

与最初的压缩感知算法相比,并行压缩感知算法架构增加了两个步骤,即冗余字典的构造和稀疏变换基的选择,而且,所有的计算均引入了并行计算方法.接下来,我们将详细介绍此并行压缩感知算法架构.

3.2 并行计算和压缩感知

并行计算(parallel computing)是一种多个指令得以同时运行的计算模式.它的主要做法是将一个大的复杂问题分解成多个子问题,然后分配给不同的处理器,各个处理器之间通过相互协同,并行地处理子任务,从而达到提高求解速度,或者加大求解应用问题规模的目的.并行计算以多核处理为主要的形式,多核处理又可以分为 CPU 多核处理和 GPU 多核处理.但 CPU 与 GPU 协同并行计算以其强劲的计算能力、高性价比和低功耗等特点也渐渐成为新的高性能计算平台.

压缩感知算法具有较高的计算复杂度,然而物联网环境对采集数据的实时性又有较高的要求,这就使得压缩感知算法的并行化处理成为必要需求.本文设计了并行压缩感知算法架构,它与原始的压缩感知算法相似,也涉及大量的矩阵运算,如冗余字典的构造、稀疏变换基的选择、观测数据的获取、信号的重构等,这为提高算法的并行性能提供了可能.目前我们已经把部分原始的压缩感知算法(如 BP 算法、OMP 算法)做了并行化处理,运行结果如图 4 所示(其中 CPU 每个核的主频是 1.8GHz).从图 4

中可知,压缩感知算法的并行处理显著地降低了算法运行时间,从而提高了数据处采集处理的实时性.由于没有把并行压缩感知算法架构的冗余字典加入算法中,所以还未能实现算法的适应性.这也是接下来要进行的研究.

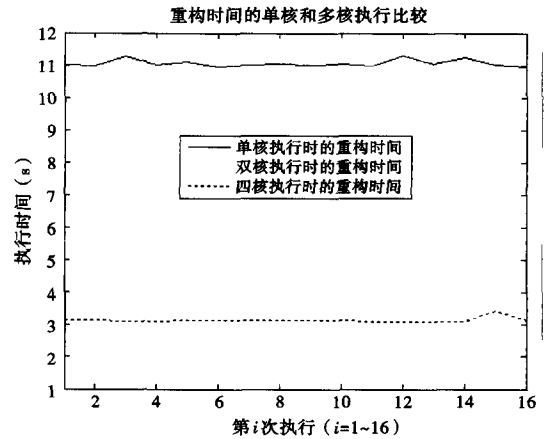


图 4 OMP 算法执行时间比较

3.3 冗余字典

压缩感知算法的稀疏变换基是不能变化的,因此原始的算法并不具有自适应性.虽然在 Peyr? 的研究中,已经把稀疏变换基的条件做了扩展,使其扩充到了正交基字典,但仍无法解决这个问题.目前,采用冗余字典来当作稀疏变换基,这是一个新的研究热点.

冗余字典下的信号稀疏表示是一种新的信号表示理论,它采用超完备的冗余函数系统代替传统的正交基函数,这样就为信号自适应地稀疏扩展提供了极大的灵活性,因此在并行压缩感知架构中,选择了冗余字典构造正交基.这也是我们未完成的部分.接下来,将要寻找传感环境下的冗余字典的构造方法、研究从冗余字典中快速选择最优稀疏变换基的方法及其算法并行化的问题.

3.4 观测矩阵和观测数据

构造观测矩阵必须保证观测矩阵 Φ 必须与稀疏变换基 Ψ 不相关.目前,观测矩阵大多采用随机矩阵,而它一般与固定正交基矩阵不相关.因此,在我们的实验中也采用随机矩阵作为观测矩阵,然后用它去观测信号 X ,即利用式(2)求得观测数据 Y ,这也是利用压缩感知方法压缩后的数据.采用随机矩阵作为观测矩阵易构造,但计算复杂度高.因此,选择合适的观测矩阵对信号压缩也是目前压缩感知理论的一个重要研究方向.

3.5 信号重构

如图4所示,重构算法的并行化减少执行时间,从而可以提高数据的实时性.在我们的实验中,其他条件相同的条件下,压缩感知算法使用双核CPU进行并行处理,要比只是用单核CPU运行可以节省 $1/3 \sim 3/4$ 的时间,然而若采用具有更多核的GPU进行并行处理,将能更好的减少执行时间.前面已经介绍了,目前的压缩感知重构算法根据它们的特点可以分为三类:贪婪算法、凸松弛算法和组合算法.目前,我们已经研究了凸松弛算法中的BP算法、贪婪算法中的OMP算法的并行化问题,但我们认为组合算法会更适合并行化.这也是我们接下来要进行的研究,希望能找到有更好的并行性能的压缩算法能.

4 结束语

本文提出了使用压缩感知理论处理物联网海量数据,并对压缩感知算法实现了并行化处理,这提高算法的执行速度和数据处理的实时性,对于改善物联网的数据采集、传输、处理能力都有很大的帮助,可以促进物联网的发展.对于压缩感知理论适应性差的问题,虽然已经在本文给出的并行压缩感知算法框架中提出,但仍未实现.今后,我们将对这一课题进行更深入的研究,希望取得更大的成果.

参考文献:

- [1] Cands E J, Wakin M B. An introduction to compressive sampling[J]. IEEE Mag. Signal Proc., 2008, 25(2): 21-30.
- [2] 戴琼海,付长军,季向阳. 压缩感知研究[J]. 计算机学报, 2011, 34(3): 425-434.
- [3] Peyr G. Best basis compressed sensing[J]. Lecture Notes in Computer Science, 2007(4485): 80-91.
- [4] Cands E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Trans. on Information Theory, 2006, 52(2): 489-509.
- [5] Baraniuk R. A lecture on compressive sensing[J]. IEEE Mag. Signal Processing, 2007, 24(4): 118-121.
- [6] Coifman R, Geshwind F, Meyer Y. Noiselets[J]. Appl. Comp. Harmonic Analysis, 2001, 10(1): 27-44.
- [7] Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit[J]. SIAM Review, 2011, 43(1): 129-159.
- [8] Neff R, Zakhov Z. Very low bit rate video coding based on matching pursuits[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1997, 7(1): 158-171.
- [9] Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit[J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655-4666.
- [10] Gilbert A C, Guha S, Indyk P. Near-optimal sparse Fourier representations via sampling[J]. Proceedings of the Annual ACM Symposium on Theory of Computing, Association for Computing Machinery, 2002 (STOC'02): 152-161.

作者简介:



周 健 男, (1987—), 硕士研究生.
研究方向为分布式计算和压缩感知.

张永平 男, (1979—), 博士研究生. 研究方向为 Web 服务、云计算和分布式计算、压缩感知.

张功萱 男, (1961—), 博士生导师. 研究方向为网络与分布式计算、客户/服务器计算、CORBA 技术、Web Services 和信息安全、Grid Computing 技术.