基于压缩感知的 WSN 数据压缩与重构

李继楼,柯家龙

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘 要:在无线传感器网络(WSN)中,传统的处理方式是采用奈奎斯特技术对信号进行采样并重构,而随着信号频率的增加,应用奈奎斯特技术会使成本急剧增加,这是人们所不乐见的。针对这一问题,近年来出现一种新的技术即压缩感知技术,它能利用更少的数据和合适的重构方法得到更精确的原始信号。将稀疏贝叶斯学习(SBL)和压缩感知联合起来,形成了一种在噪声的情况下更好重建可压缩信号的方法,并进一步将这种方法应用在 WSN 中,可以在误差允许的范围内有效控制测量数据的维数,所以在保证了一定的误差的同时还减少了成本,提高了算法的效率。

关键词:无线传感网络;压缩感知;贝叶斯模型;信号重构

中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2015)09-0111-04

doi:10.3969/j.issn.1673-629X.2015.09.024

Data Compression and Recovery of WSN Based on Compressive Sensing

LI Ji-lou,KE Jia-long

(College of Communication and Information Engineering,Nanjing University of Posts and Telecommunications,
Nanjing 210003,China)

Abstract:In wireless sensor networks,signal is sampled and reconstructed using the technology of Nyquist in the past. But it requires a substantial increase in the cost with the growth of the signal frequency,which is that people do not like to see. Recently a new technology is emerged,which is called compressive sensing technology,is a good way to solve this problem. Compressive sensing can use less data and appropriate reconstruction method to get a more accurate original signal. Put Sparse Bayesian Learning (SBL) and compressive sensing together to form a better reconstruction compressible signal under the noise. This method can effectively control the dimension of measurement data within the range of allowed error in WSN,so you can ensure a certain degree of error while reducing the cost,improving the efficiency of the algorithm.

Key words:wireless sensor networks;compressive sensing;Bayesian model;signal reconstruction

0 引 言

随着信息技术的发展,人们对信息需求量的不断增加,网络通信、存储技术、多媒体技术的发展越来越快,网络的规模也不断扩大,寻找高效的信息处理技术来降低数据量成为无线传输系统中亟待处理的问题之一。因为数字化的各类信息的数据量非常庞大,若不对其进行有效的压缩就难以得到实际的应用,因此,数据压缩技术成为人们研究的一项重要技术。无线传感器网络是近来研究的热点方向之一。它是由分布在监测区域内的大量微型传感器节点通过无线电通信而形成的一个自组织网络系统。这个系统的目的是协作地感知、采集和处理网络覆盖区域里被监测对象的信息,并将结果发送给用户。在一个传感器网络中,常常包含大量传感器节点,每个传感器都会采集大量的数据。这些数据将会被传输到一个控制中心,也会在各个节点之间传输,在这种分布式传感网络中,数据传输功耗和带宽需求非常大,所以,如何对这样的分布式信号进行压缩,从而减小通信开销已经成为非常紧迫的需求。应用近年来形成的压缩感知技术就是一种有效途径,压缩感知技术[1-3]在国内外许多领域都得到了广泛的应用并取得了很显著的效果。在信号的重构方面,相对于常用的贪婪追踪重构算法实现点的估计,文中借助贝叶斯模型得到整体数据后验概率的分布。在有噪声条件下,OMP[4]和 ROMP[5]等方法都不能很好地重构原信号,但贝叶斯模型可以。并且通过大量实验表明,选择合适的反馈系数控制采集数据的维数可以在

收稿日期:2014-11-07 **修回日期**:2015-02-11 **网络出版时间**:2015-08-26

基金项目:国家自然科学基金资助项目(60972041,60972045)

作者简介:李继楼(1988-),男,硕士研究生,研究方向为在无线网络中压缩感知的应用。

网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20150826.1603.072.html

保证允许误差的前提下尽量减少测量次数，降低成本，提高算法效率。

1 压缩感知基本理论

传统通信系统中的采样遵循的是奈奎斯特抽样定理。该定理指出，在获取信号时，为了不丢失信号的信息，抽样速率必须大于信号中最大频率的两倍，才能精确重构信号。但是随着科技的不断发展，高分辨率的数码装置的采样产生了庞大的数据，如何更高效地处理这些数据并最大限度地节省存储和传输的成本是一大难题。压缩感知则是保存原始信号结构的线性投影，然后再从这些投影中将信号重构出来，其速率远远低于奈奎斯特抽样率。在该理论框架下，信号的采样速率不再取决于采样带宽，而是取决于信号的结构和内容，也就是稀疏性和非相干性。只要信号是可压缩的或者是稀疏的。具体表述为：假设信号 x 在正交基 $\boldsymbol{\Psi}$ 上是稀疏的，即：

$$\boldsymbol{x}^{(k)} = \boldsymbol{\Psi s}^{(k)} \tag{1}$$

其中，$\boldsymbol{x}^{(k)} \in R^N$；$N$ 维向量 $\boldsymbol{s}^{(k)} \in R^N$ 是 $\boldsymbol{x}^{(k)}$ 稀疏表示。

在 WSN 中，观测数据可以表示为：

$$\boldsymbol{y}^{(k)} = \boldsymbol{\Phi x}^{(k)} \tag{2}$$

其中，在 k 时刻，观测向量 $\boldsymbol{y}^{(k)} \in R^N$；在无线传感网络中 $\boldsymbol{\Phi}^{L\times N}$ 被认为是路由选择矩阵，每一行最多有一个元素为 1，每一列也最多有一个元素为 1 且最多有 L 列非 0。最后可得：

$$\boldsymbol{y}^{(k)} = \boldsymbol{\Phi x}^{(k)} = \boldsymbol{\Phi \Psi}^{\mathrm{T}} \boldsymbol{s}^{(k)} = \boldsymbol{\Theta s}^{(k)} \tag{3}$$

要想得到重构信号的稀疏表示 $\boldsymbol{s}^{(k)}$，就要求式(4)的最优化解：

$$\min \ \| \boldsymbol{s} \|_{l_0} \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{\Phi \Psi}^{\mathrm{T}} \boldsymbol{s} \tag{4}$$

这个系统是病态的，即方程的数量 L 小于变量的数量 N。所以要对式(4)采用一些特殊的处理技术[6-9]才可以求得信号的估计值 $\boldsymbol{s}^{(k)}$，再代入式(1)才可得所感兴趣的信号重构 $\boldsymbol{x}^{(k)}$。但是这样处理数据有一个必要条件就是信号是稀疏的并且 $\boldsymbol{\Phi\Psi}$ 满足 RIP 条件，即要求稀疏基和测量基之间满足非相干性，也就是要求这两个基之间不能相互稀疏描述。另外在有噪声的情况下，记为：

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{n} \tag{5}$$

利用拉氏定理可以改写为：

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \{ \| \boldsymbol{y} - \boldsymbol{\Phi\Psi}^{\mathrm{T}} \boldsymbol{s} \|_2^2 + \lambda \ \| \boldsymbol{s} \|_1 \} \tag{6}$$

目前重构算法可以归为以下三类：

(1)凸松弛法。这个方法是将非凸优化问题转化为凸优化问题，进而找到信号的逼近。现有的凸优化方法主要包括基追踪(Basis Pursuit，BP)算法、迭代阈值法(Iterative Thresholding，IT)、迭代硬阈值法(Iterative Hard Thresholding，IHT)、梯度投影稀疏重建法(Gradient Projection for Sparse Reconstruction，GPSR)、Bregman 迭代法(Bregman Iterative，BI)等。

在这三类算法中，凸优化给出了信号稀疏重构的最强保证，在观测矩阵满足一定的条件下，它能以较少的测量次数重构所有的稀疏信号，但其最大的缺点是算法复杂度高、重构速度慢，对于大尺度的重构问题实现困难。

(2)贪婪追踪算法。这类方法是通过每次迭代时寻找一个局部最优解来不断逼近原始信号。包括匹配追踪算法(Matching Pursuit，MP)、正交匹配追踪(Orthogonal Matching Pursuit，OMP)、正则化的正交匹配追踪(Regularized Orthogonal Matching Pursuit，ROMP)、分级正交匹配追踪(Stagewise Orthogonal Matching Pursuit，StOMP)、广义正交匹配追踪(Generalized Orthogonal Matching Pursuit，GOMP)、多路径追踪(Multipath Matching Pursuit，MMP)、压缩采样追踪(Compressive Sampling Matching Pursuit，CoSaMP)等。

贪婪追踪算法是这三类算法中重建速度最快的算法，基于这个优点，这类算法得到广泛应用

(3)组合算法。这类方法要求信号的采样支持通过分组测试快速重建，如傅里叶采样、链式追踪和 HHS 追踪等。

目前，重构算法可以很大程度上精确地重构信号或图像，但这些稀疏重构算法仍然都存在一些无法改善的缺点。此外，现有算法对含有噪声信号或采样过程中收入噪声信号的重构效果较差，鲁棒性也较差，如何改善有待进一步研究。

2 贝叶斯模型

应用贝叶斯模型[10-12]对信号重构的过程被视为是一种线性回归的问题。主要分为两个阶段：第一阶段是所有的未知数被认为是服从指定分布的随机量[13]。在有噪的观测模型中，通常假设噪声是服从独立的均值为 0、方差为 σ^2 的高斯分布，记为：

$$p(n) = \prod_{i=1}^{N} N(n_i \mid 0, \sigma^2) \tag{7}$$

超参数 σ^2 可以提前设置为常量，通常这个参数的估计值定为 $\sigma^2 = 0.01 \ \| y \|_2^2$，也可以通过数据进行估计。设 $\beta = \sigma^{-2}$ 可以进一步假设：

$$p(\beta / a, b) = \Gamma(\beta / a, b) \tag{8}$$

观测向量 $\boldsymbol{y}$ 可以认为是多变量的高斯似然函数：

$$p(\boldsymbol{y} \mid \boldsymbol{\Phi s}, \sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp\{ - \frac{\| \boldsymbol{y} - \boldsymbol{\Phi s} \|^2}{2\sigma^2} \} \tag{9}$$

通常对信号模型[9-10]采用如下形式:

$$p(s \mid \lambda)=\frac{\lambda}{2}\exp(-\lambda \mid s) \tag{10}$$

但是这个拉氏先验分布公式在贝叶斯分析中并不方便,因为不能和观测模型式(9)进行联合处理,所以这里采用多层先验概率进行分析。首先:

$$p(s \mid \alpha)=\prod_{i=1}^{N} N(s_i/0,\alpha_i) \tag{11}$$

其中,超参数 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_M)$,超参数是指信息先验分布通常有其自身的参数,这些参数被称作超参数。进一步把参数 α 看作变量并由超参数 a,b 决定:

$$p(\alpha \mid a,b)=\Gamma(\alpha_i \mid a,b)= \Gamma(a)^{-1}b^{\alpha}\alpha_i^{\alpha-1}\exp(-b\alpha_i) \tag{12}$$

其中, $\Gamma(a)=\int_0^{\infty}t^{\alpha-1}\exp(-t)\mathrm{d}t$ 。

为了简化模型,初始化时可以假设这些参数是某些固定的值,例如文献[14]中假设 $a=b=c=d=0$;文中采用类似的处理手段,固定 $a=1$。所以有:

$$p(s/\alpha=1,b)=\int p(s/\alpha=1,b)\mathrm{d}\alpha = \prod_i \int p(s_i/\alpha_i)p(\alpha_i/a=1,b)\mathrm{d}\alpha = \frac{\lambda^{N/2}}{2^N}\exp(-\sqrt{\lambda}\sum_i s_i) \tag{13}$$

注意上式进行了 $\lambda=2b$ 的改写,把概率论中著名的恒等式,应用到文中可得 s 的后验概率:

$$p(s/y,\lambda,\sigma^2)=p(y/s,\sigma^2)/p(y/\lambda,\sigma^2) \tag{14}$$

把式(9)、式(13)代入,计算可知式(14)服从 $N(u,\Sigma)$ 。其中:

$$\begin{aligned} &u=\sigma^2\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y}\\ &\boldsymbol{\Sigma}=(\boldsymbol{\Lambda}+\sigma^{-2}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\\ &\boldsymbol{\Lambda}=\mathrm{diag}(\alpha_i^{-1})p(y/\lambda,\sigma^2)p(s/\lambda)\mathrm{d}s \end{aligned} \tag{15}$$

通过统计分析可知,上面的均值也就是要重构的信号的稀疏表示形式,另外协方差 $\boldsymbol{\Sigma}$ 的对角线上的元素也就是重构过程中产生的误差。在无线传感网络中,允许的误差最大值为 $\boldsymbol{\Sigma}$,重构过程中产生的误差均值为 Δ,通过大量的实验可知反馈系数 p_{re}:

$$p_{\mathrm{re}}=\begin{cases}\min\{\dfrac{\Delta-\eta}{\eta}c_1,1\},\text{if } 1.2\Delta>\eta\\[2ex] \max\{-\dfrac{\eta-\frac{3}{5}\Delta}{\eta}c_2,-(1-\dfrac{K}{L}\log(\dfrac{N}{K}))\},\text{if } 1.2\Delta<\eta\end{cases} \tag{16}$$

上面也就是第 2 阶段的目的,得到最匹配的超参数从而进一步获得权值参数的最优估计即待重建的信号[15-16],采用的方法是最大化超参数的边缘对数似然函数。对于边缘似然函数 $p(\lambda,\beta,\alpha \mid y)$ 的对数形式的最大化,这里采用第二类型最大似然参数估计的方法[11]。边缘似然函数的对数形式通常表示为:

由于: $p(\lambda,\beta,\alpha \mid y)=\dfrac{p(\lambda,\beta,\alpha,y)}{p(y)}\propto p(\lambda,\beta,\alpha,y)$,所以有:

$$\begin{aligned} L&=\lg(\int p(\lambda,\beta,\alpha,y,s)\mathrm{d}s)=\\ &\lg(\int p(y \mid s,\beta)p(s \mid \alpha)p(\alpha \mid a=1,\lambda)p(\lambda)p(\beta)\mathrm{d}s)=\\ &\lg((\frac{1}{2\pi})^{\frac{N}{2}}\ |\boldsymbol{C}|^{-\frac{1}{2}}\exp[-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y}]p(\alpha \mid a=1,\lambda)p(\lambda)p(\beta)) \end{aligned} \tag{17}$$

其中, $\boldsymbol{C}=\sigma^2\boldsymbol{I}+\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}$ 。对上式:

$$\begin{aligned} L(\alpha)&=-\frac{1}{2}[\lg|\boldsymbol{C}_{-i}|+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{y}+\frac{\lambda}{2}\sum_{j\neq i}\alpha_j]+\\ &\frac{1}{2}[\log\frac{1}{1+\alpha_i p_i}+\frac{q_i^2\alpha_i}{1+\alpha_i p_i}-\alpha_i\lambda]=\\ &L(\alpha_{-i})+l(\alpha_i) \end{aligned} \tag{18}$$

估计参数 λ_i,最大化 $L(\lambda)$ 因为把 λ_i 都视为常数,所以相当于最大化 $L(\lambda_i)$ 。

计算可知:当 $q_i^2-p_i>\lambda$,有

$$\alpha_i=\frac{-p_i^{\ 2}+2\lambda p_i+p_i\sqrt{p_i^2+4\lambda q_i^2}}{2\lambda p_i^{\ 2}} \tag{19}$$

其他情况下 $\alpha_i=0$。

另外设:

$$\begin{aligned} &P_i=\beta\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{\varphi}_i-\beta^2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\varphi}_i\\ &Q_i=\beta\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{y}-\beta^2\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{y} \end{aligned} \tag{20}$$

更新公式:

$$p_i=\frac{P_i}{1-\alpha_i P_i}\qquad q_i=\frac{Q_i}{1-\alpha_i P_i} \tag{21}$$

3　实验仿真

在第一个实验中考虑的是稀疏 M 为 30 且长度 N 为 512 的稀疏信号,随机选择 30 个位置并随机赋值正 1 或-1,观测矩阵为随机矩阵并进行归一化处理。仿真结果如图 1 所示(纵坐标代表信号的幅值,横坐标代表信号的长度)。

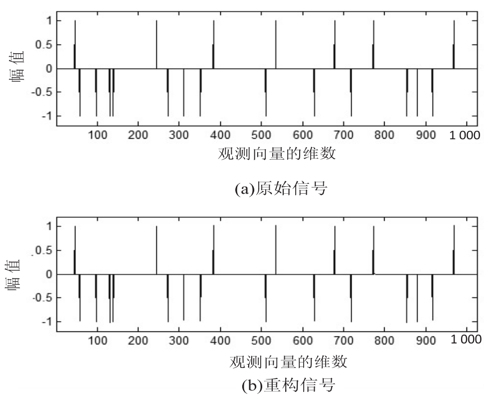

图 1　原始信号与重构信号

在反馈系数更新公式中, $c_1 = 1.7$, $c_2 = 3$。下面的实验是在 1 000 次测量进行重构的过程中,最后一次测量的实验结果。

在第二实验中,数据集的选取和实验一中的相同,主要是对观测向量 y 的维数对整体重构误差的影响进行,如图 2 所示。

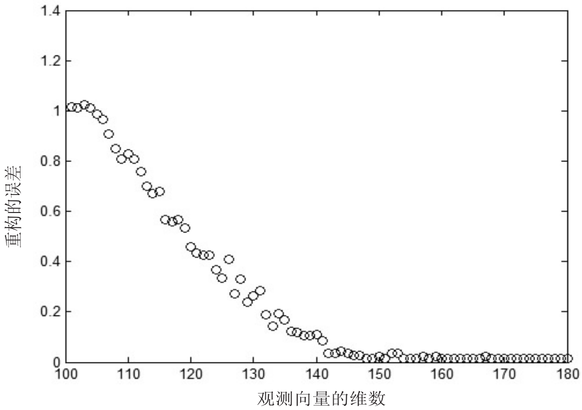

图 2 观测数据的次数与误差关系图

在自适应系统中,如果误差太大,可以通过增加测量向量的维数来减少误差,这是贝叶斯模型一个明显的优势。

在第三个实验中,对比 OMP、GOMP[17] 和文中的重构算法,如图 3 所示("＊"代表 OMP 算法的重构,"o"代表文中算法的重构;"x"代表 GOMP 的重构,其中纵坐标代表重构信号的幅值,横坐标代表信号的长度)。

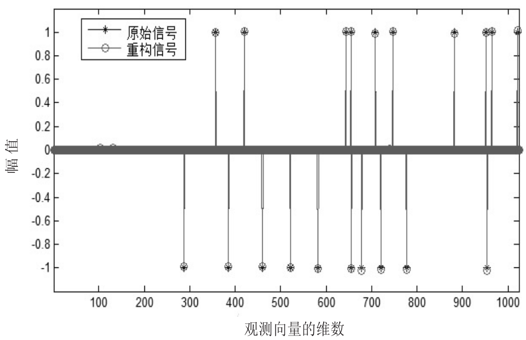

图 3 原始信号与重构信号的对比图

4 结束语

传感器的使用寿命很大程度受到无线传感网络中微型电池供电的影响,这使得无线传感网络的稳定性受到了很大影响。在现实的环境中,噪声无处不在,贝叶斯模型就是一种典型解决有噪情况下的框架模型。这就是文中把贝叶斯模型和压缩感知技术结合起来的最原始的出发点。通过大量的仿真求证反馈系数是为了自动地控制网络系统,以满足在允许误差范围内利用最少的测量数据获取精确的重构信号,降低数据的采集传输成本,延长传感器的使用寿命,从而保证无线传感网络的稳定性,为未来无线传感网络性能的优化提供了一种有效的方法。

参考文献:

[1] Donoho D. Compressed sensing[J]. IEEE Trans on Information Theory,2006,52(4):4036-4048.

[2] Candes E J, Wakin M B. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine,2008,25(2):21-30.

[3] Candes 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.

[4] Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit[J]. IEEE Trans on Information Theory,2007,53(12):4655-4666.

[5] Needell D, Vershynin R. Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit[J]. Found Comput Math,2009,9(3):317-334.

[6] 邓乃扬,田英杰. 数据挖掘中的新方法-支持向量机[M]. 北京:科学出版社,2004.

[7] Vapnik V N. 统计学习理论[M]. 许建华,张学工,译. 北京:电子工业出版社,2004.

[8] Vapnik V N. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000.

[9] 薛建儒,郑南宁,郑朝晖,等. 基于自适应高斯混合体模型的相控阵雷达 TWS 跟踪技术[J]. 电子学报,2003,31(3):433-436.

[10] Babacan S D, Molina R, Katsaggelos A K. Bayesian compressive sensing using Laplace priors[J]. IEEE Trans on Image Processing,2010,19(1):53-63.

[11] Tipping M. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research,2001,1:211-244.

[12] Figueiredo M, Nowak R, Wrigh S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems[J]. IEEE Trans on Selected Topics in Signal Processing,2007,1(4):586-597.

[13] 王平波,蔡志明,刘旺锁. 混合高斯概率密度模型参数的期望最大化估计[J]. 声学技术,2007,26(3):498-502.

[14] 岳 佳,王士同. 高斯混合模型聚类中 EM 算法及初始化的研究[J]. 微计算机信息,2006,22(11-3):244-246.

[15] Seeger M W, Nickisch H. Compressed sensing and Bayesian experimental design[C]//Proc of ICML. [s. l.]:[s. n.],2008:912-919.

[16] Faul A C, Tipping M E. Analysis of sparse Bayesian learning[C]//Proc of NIPS. [s. l.]:[s. n.],2002:383-389.

[17] Wang J, Kwon S, Shim B. Generalized orthogonal matching pursuit[J]. IEEE Trans on Signal Processing,2012,60(12):6202-6216.

基于压缩感知的WSN数据压缩与重构

作者： 李继楼， 柯家龙， LI Ji-lou， KE Jia-long
作者单位： 南京邮电大学 通信与信息工程学院,江苏 南京,210003
刊名： 计算机技术与发展ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2015(9)