

文章编号:1671-1637(2012)03-0113-07

## 基于压缩传感的交通流量数据压缩方法

李清泉<sup>1,2</sup>, 周尧<sup>2,3</sup>, 乐阳<sup>1,2</sup>, 叶嘉安<sup>4</sup>

(1. 武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079; 2. 武汉大学 时空数据智能获取技术与应用教育部工程研究中心, 湖北 武汉 430079; 3. 武汉大学 测绘学院, 湖北 武汉 430079; 4. 香港大学 城市研究及城市规划中心, 香港)

**摘要:**为准确获得用于数据压缩的变换矩阵, 引入了基于压缩传感的交通流量数据压缩方法, 在数据压缩端无需考虑变换矩阵的选择问题, 直接通过高斯投影实现高效数据压缩。首先验证了交通流量数据在经过 K-SVD 方法训练过的字典上能够实现稀疏表达; 然后在数据压缩端, 通过具有限制性等距条件的随机矩阵将原始高维数据投影到低维空间上, 实现数据的高效快速压缩; 最后在数据传输后, 通过凸优化算法在交通信息处理端完成数据解压缩。以美国某高速公路线圈传感器采集到的交通流量数据, 对本文方法进行了验证。试验证明: 该方法能够实现快速高效的压缩编码, 当压缩比为 4 : 1 时, 解压缩相对误差仅为 0.060 8。

**关键词:**智能交通系统; 压缩传感; 数据压缩; 冗余字典; 高斯投影;  $L_1$ -合成算法

**中图分类号:**U491.112

**文献标志码:**A

## Compression method of traffic flow data based on compressed sensing

LI Qing-quan<sup>1,2</sup>, ZHOU Yao<sup>2,3</sup>, YUE Yang<sup>1,2</sup>, YEH Anthony Gar-On<sup>4</sup>

(1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei, China; 2. Engineering Research Center for Spatio-Temporal Data Smart Acquisition and Application of Ministry of Education, Wuhan University, Wuhan 430079, Hubei, China; 3. School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, Hubei, China; 4. Centre of Urban Studies and Urban Planning, The University of Hong Kong, Hong Kong, China)

**Abstract:** In order to obtain transformation matrix accurately, a new compression method of traffic flow data based on compressed sensing was introduced. The original data were projected into the low-dimension space directly by Gauss projection regardless of transformation matrix selection at the data compression side. Firstly, traffic flow data were proved to have sparse representation under the K-SVD trained dictionary. Secondly, original high-dimension data were projected into low-dimension space at the data compression side by using the random matrix with restricted isometry property, which made efficient and rapid data compression possible. Finally, after data transmission, data decompression were accomplished by convex algorithm at the data processing side. The traffic flow data obtained from the coil sensors located on a certain highway of America were used to validated the new method. The experimental result shows that the data compression method is fast and efficient. When the compression ratio is 4 : 1, the relative error of data decompression is only 0.060 8. 5 tabs, 8 figs, 18 refs.

**Key words:** intelligent transportation system; compressed sensing; data compression; redundant

收稿日期: 2012-01-17

基金项目: 国家自然科学基金项目(41171348, 60872132, 40830530); 香港研究资助局项目(HKU 754109)

作者简介: 李清泉(1965-), 男, 安徽天长人, 武汉大学教授, 工学博士, 从事地理信息系统与智能交通研究。

dictionary; Gauss projection;  $L_1$ -synthesis algorithm

**Author resume:** LI Qing-quan(1965-), male, professor, PhD, +86-27-68778222, qqli@whu.edu.cn.

## 0 引 言

随着智能交通系统(Intelligent Transportation System, ITS)的飞速发展,交通传感器数据采集能力不断增强,传感器不间断地采集到海量交通数据。庞大而复杂的数据既不利于前期的数据存储与传输,也不利于后期的数据分析,因此,需要设计高效的交通数据压缩方法,用于海量数据压缩,以节省存储空间,提高传输效率和质量,服务于实时及后期的数据处理与分析。

目前的数据压缩技术主要分为 2 大类:第 1 类是以统计编码与字典编码为代表的传统无损压缩技术,压缩比较低,在海量交通数据压缩中一般不会单独使用;第 2 类是基于变换域的压缩技术,如小波变换。肖扬等利用小波变换及霍夫曼编码进行了交通数据的去噪与压缩<sup>[1]</sup>。基于小波变换的数据压缩技术通过小波函数将原始数据转换到小波域上,保留了重要的小波系数,能够达到较高的压缩比,但是这种基于变换域的压缩方法,其压缩效果的好坏主要取决于变换矩阵的选择,数据压缩端受制于其有限的计算与存储资源,并不总是能准确找到最佳变换矩阵来实现数据压缩<sup>[2-6]</sup>。

赵志强等将 K-L 变换用于交通流量数据压缩<sup>[7-8]</sup>;耿彦斌等结合离散余弦变换与小波变换,对交通流量数据进行小波变换后,再对高频分量进行离散余弦变换,并结合 LZW 编码与霍夫曼编码,实现了较为高效的数据压缩<sup>[9]</sup>。

但是以上研究均主要采用无损编码以及变换域数据压缩技术,无法避免这些技术自身的一些缺陷,例如,在交通数据压缩端,只有性能优良,且与数据本身特征相符合的变换矩阵才能实现高效数据压缩。考虑到不同时间、地点的交通流量数据特性并不相同,最合适的变换矩阵也理应不同,而在交通信息采集端计算资源受限制的情况下,目前传感器数据压缩端并不总是能准确找到最佳变换矩阵来实现数据压缩。

针对上述问题,本文引入了一种新的数据获取、压缩理论,基于压缩传感的交通数据压缩方法,尝试对交通流量数据进行压缩。压缩传感是近几年诞生于应用数学和信号处理领域的一种新的数据获取与压缩技术,近些年来受到了广泛的关注。

压缩传感理论目前尚处于起步阶段,研究热点主要集中在信号处理、图像处理、信源信道编码、传感器网络、雷达等领域。目前,在交通数据压缩方面尚没有相关的研究成果。理论上,使用压缩传感方法,可以在交通数据采集端采用通用的随机矩阵来实现数据压缩,矩阵不需要根据数据的结构来设计,从而使得数据压缩过程更加简便通用。这对数据压缩端有限的计算资源而言,无疑具有重要意义。虽然数据经过传输之后,在信息处理端进行解压缩的过程较为复杂,但是此过程往往可以在室内具有较高的计算能力的大型计算机上进行,对实际应用不会造成太大影响。那么,压缩传感用于交通数据压缩能达到怎样的效果?与已有方法相比,有何优缺点?针对这些问题,本文研究了基于压缩传感理论的交通流量数据压缩方法,涉及流量数据的稀疏表达、传感矩阵的构造、原始数据解压缩等问题,并以布设在美国明尼苏达州附近高速公路上的线圈传感器采集的交通流量数据为例,做出了相关验证。

## 1 压缩传感理论

压缩传感是建立在数据稀疏表达基础上的一种全新数据获取与压缩理论,由 Candes 等于 2006 年首次提出<sup>[10-11]</sup>。当数据可以被稀疏表达时,将原始数据通过传感矩阵进行简单的线性投影,把高维数据投影到低维空间,以实现数据压缩,然后在低维空间上的少量数据通过相应的恢复算法精确恢复出原始数据<sup>[12]</sup>。

基于压缩传感的数据压缩具有显著的非对称性。与传统变换域数据压缩相比:压缩传感在数据压缩端,只需要通过简单的随机投影就能实现高效的数据压缩;数据传输后,在信息处理端进行解压缩的过程中才需要去考虑变换矩阵选取问题。在大多数实际应用的数据获取与分析系统中,通常数据采集端为廉价、省电、计算能力较低的设备,而数据分析端往往为室内具有较高计算能力的大型计算机,因此,基于压缩传感数据压缩所具有的非对称性很适用于这种实际应用需求。

压缩传感主要包括原始数据的稀疏表达、利用传感矩阵实现数据压缩以及数据恢复 3 个部分。压缩传感方法通过高斯投影把原始数据投影到低维空间实现数据压缩,主要流程见图 1。传统变换域数

据压缩通过保留变换域上的重要系数,舍弃不重要系数来实现,主要流程见图2。

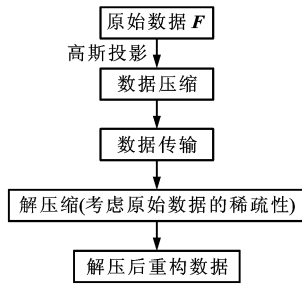


图1 压缩传感流程

Fig. 1 Compressed sensing flow

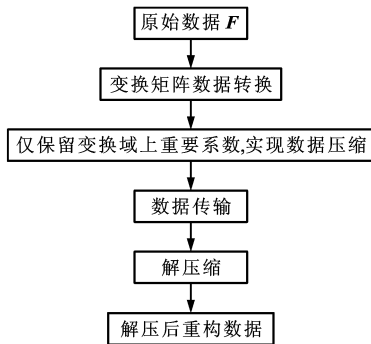


图2 传统数据压缩流程

Fig. 2 Traditional data compression flow

对比图1、2可以发现:基于压缩传感的数据压缩直接将原始数据投影到低维空间,从而实现数据压缩,与传统变换域数据压缩的区别是,在数据压缩端不需要选择变换矩阵,也不涉及保留重要系数,去除不重要系数这一过程。压缩传感的具体流程如下。

设原始数据是一个维度为  $N$  的列向量  $F$ ,在维度为  $N \times K$  的变换矩阵  $D$  下,原始数据可以表示为

$$\begin{cases} F = DX \\ D \in \mathbb{R}^{N \times K} \\ X \in \mathbb{R}^K \end{cases} \quad (1)$$

式中: $X$  为原始数据  $F$  在变换矩阵  $D$  下的系数向量。

在变换矩阵作用下,系数向量  $X$  的维度为  $K$ 。当  $X$  中绝大部分系数为0,非零系数的个数远小于向量维度  $K$ ,即  $X$  的0范数满足

$$\|X\|_0 \ll K \quad (2)$$

$F$  在变换矩阵  $D$  下可以实现稀疏表达。根据  $D$  中的列向量是否相互正交,变换矩阵  $D$  可以分为完备库与冗余字典两类,具体分析见2.1节。

在验证原始数据能够稀疏表达后,将原始数据  $F$  经过维度为  $M \times N$  的传感矩阵  $A$  投影( $M$  远小于

$N$ ),可将原始数据  $F$  转换到更低维度的空间上,记压缩后的数据为  $Y$ ,其维度为  $M$ 。由于压缩后数据维度远小于原始数据维度,经过线性投影后的数据长度远小于原始数据长度,以此实现数据压缩。压缩传感理论表明,当传感矩阵  $A$  满足限制性等距条件(Restricted Isometry Property, RIP)时,原始数据由高维空间投影到低维空间,信息的损失是极少的。常用的传感矩阵有高斯随机矩阵、二值随机矩阵、傅里叶观测矩阵以及不相干观测矩阵<sup>[13]</sup>。数据压缩的过程为

$$Y = AF = ADX \quad (3)$$

压缩后的数据经过传输后,在信息处理端进行的解压缩等价于式(3)中由  $Y$  求解  $F$ 。由于  $M$  远小于  $N$ ,式(3)为一个欠定方程,由低维数据  $Y$  求解高维数据  $F$  具有不唯一的解,看似求解  $F$  似乎并不可行。但是,综合考虑传感矩阵的 RIP 性质以及之前所提到的数据稀疏性,可以将问题转换为  $L_0$  范数最优优化问题,即以求解系数向量  $X$  的  $L_0$  范数  $\|X\|_0$  远远小于  $K$  最小为目标,约束条件为  $Y = AF$ 。

由于  $L_0$  范数优化为 NP 完全问题,即求解  $X$  中的非零值需要进行穷举,在实际应用中并不可行。通常会采用  $L_1$  范数最小法、匹配追踪算法或迭代阈值法求次优解。相对于数据压缩过程而言,数据解压过程要相对复杂得多。

## 2 基于压缩传感的交通流量数据压缩

本文将压缩传感理论引入交通领域,根据压缩传感数据压缩时非对称性的特点,对交通流量数据进行压缩,在交通传感器端,通过随机投影完成原始数据的压缩,然后把复杂的数据解压缩转移到交通信息处理端。对于计算资源相对有限的交通传感器端,这种简易的压缩方法具有重要意义。基于压缩传感的交通流量数据压缩流程见图3。



图3 压缩流程

Fig. 3 Compression flow

### 2.1 字典下交通流量数据的稀疏表达

压缩传感理论建立在数据稀疏表达的基础上,将压缩传感引入交通领域首先需要验证交通流量数据是否可以通过变换矩阵  $D$  的作用,在特定的变换域上实现稀疏表达。式(1)中的变换矩阵  $D$  依据是否正交可以分为以下2类。

(1)当  $K = N$ ,且  $D$  中的列向量相互正交时,则

变换矩阵  $\mathbf{D}$  为正交变换矩阵,称为完备库。常见的完备库有傅里叶变换矩阵、离散余弦变换矩阵等。完备库的构造比较简单,但是考虑到交通流量数据的复杂性,其在单一的完备库上稀疏表达效果并不理想。

(2)当  $K > N$  时,  $\mathbf{D}$  中的列向量不再满足正交性,则称变换矩阵  $\mathbf{D}$  为过完备库,即冗余字典。在单一正交变换下不能得到较好稀疏表达时,可以考虑使用冗余字典,冗余字典大大增加了稀疏表达的灵活性,拓宽了压缩传感的应用范围<sup>[14-15]</sup>。

冗余字典中的元素被称为原子,其构造是灵活多样的。字典可以是固定的,也可以通过对样本训练得来<sup>[16-17]</sup>。变换矩阵为冗余字典时,记为

$$\mathbf{D} = (\mathbf{d}_t \mid t = 1, 2, \dots, K)$$

式中:  $\mathbf{d}_t$  为字典  $\mathbf{D}$  中的第  $t$  个列向量,为 Hilbert 空间的单位矢量。

由于  $K > N$ ,字典具有冗余性,原子之间也不需要满足正交性。字典的选择应尽可能符合被逼近数据的结构,其构成可以没有任何限制。

对于维度为  $N$  的原始数据  $\mathbf{F}$ ,选取字典中少量的原子来表达数据的主要部分

$$\mathbf{F} \approx \sum_{j=0}^{h-1} c_j \mathbf{d}_{t_j} \quad h \ll N \quad (4)$$

式中:  $c_j$  为原始数据  $\mathbf{F}$  在原子  $\mathbf{d}_{t_j}$  上的分量;  $j$  为原始数据分解的次数;  $t_j$  为字典原子对应的序列;  $h$  为稀疏表达中所选取的原子总个数,远小于原始数据长度。

当  $h \ll N$  时,原始数据在冗余字典上实现了稀疏表达。交通数据类型多样,本文以常见的线圈传感器所采集的交通流量数据为研究对象,采用基于 K-SVD<sup>[16]</sup> 算法的训练字典。K-SVD 算法是建立在  $K$  均值聚类基础上的一种训练字典的方法:首先通过已有的数据对字典进行初始化;然后根据距离测度,将训练数据聚类到距离最近的少数几个字典原子中去;最后根据现有的分类结果,更新字典。其主要思路是通过  $L_1$  范数约束以及奇异值分解,同时更新字典与系数向量。

在 K-SVD 算法中:设  $\mathbf{w}_l \in \mathbf{R}^N$  为训练数据;  $\mathbf{b}_l \in \mathbf{R}^K$  为第  $l$  个训练数据的稀疏表达系数向量;训练数据  $\mathbf{w}_l$  的矩阵为  $\mathbf{W}$ ;系数向量  $\mathbf{b}_l$  的矩阵为  $\mathbf{B}$ 。

K-SVD 算法的目标为

$$\begin{aligned} \min(\|\mathbf{W} - \mathbf{DB}\|_2) \\ \text{s. t. } \|\mathbf{B}_l\|_0 \leq T_0 \end{aligned}$$

式中:  $\|\mathbf{W} - \mathbf{DB}\|_2$  为  $\mathbf{W} - \mathbf{DB}$  的  $L_2$  范数,以求解

$\|\mathbf{W} - \mathbf{DB}\|_2$  最小时对应的字典  $\mathbf{D}$  与系数向量矩阵  $\mathbf{B}$  为目标;  $\|\mathbf{B}_l\|_0$  为系数向量  $\mathbf{B}_l$  的  $L_0$  范数,即系数向量中不为 0 的分量个数;  $T_0$  为系数向量不为 0 分量的数目的上界。

K-SVD 算法具体流程如下。

(1)初始化字典  $\mathbf{D}$ ,一般选择训练数据集合中的一部分数据来赋予字典  $\mathbf{D}$  初值。

(2)第一次迭代,使用稀疏分解方法计算系数向量,目标为

$$\begin{aligned} \min(\|\mathbf{w}_l - \mathbf{D}\mathbf{b}_l\|_2) \\ \text{s. t. } \|\mathbf{b}_l\|_0 \leq T_0 \end{aligned}$$

(3)更新字典,字典  $\mathbf{D}$  第  $t$  列的误差  $\mathbf{E}_t$  为

$$\mathbf{E}_t = \mathbf{W} - \sum_{l \neq t} \mathbf{d}_l \mathbf{b}_l \quad (5)$$

对  $\mathbf{E}_t$  进行奇异值分解,得到的向量作为字典第  $t$  列的更新。

(4)重复步骤(2)、(3),逐列更新字典,直到达到迭代次数为止。

通过对大样本的交通流量数据进行不断的学习与训练,得到的基于 K-SVD 方法的训练字典可以综合考虑线圈传感器数据的特点。线圈传感器交通流量数据在该字典上较好地实现了稀疏表达。

## 2.2 线性投影实现数据压缩

本文采用高斯随机矩阵,在交通传感器端根据式(3)对原始交通流量数据进行投影,实现数据压缩。由于  $M$  远小于  $N$ ,经过传感矩阵投影后,数据长度远小于原始数据长度,在此基础上实现数据压缩。高斯随机矩阵能够以较高的概率满足 RIP 条件,保证压缩后的数据  $\mathbf{Y}$  中依然包含足够的信息量,在交通信息处理端通过恢复算法解压缩得到原始数据。

## 2.3 数据解压缩

在交通信息处理端,压缩后的数据  $\mathbf{Y}$  通过恢复算法可以得到完整的原始数据  $\mathbf{F}$ 。解压缩过程中需要综合考虑原始数据的稀疏性。

本文中采用的训练字典为一种高度相关字典,高斯随机矩阵  $\mathbf{A}$  虽然能够以较高的概率满足压缩传感所需的 RIP 条件,但是与字典  $\mathbf{D}$  的高度相关性,使  $\mathbf{AD}$  不再满足 RIP 条件,因此,不能够精确恢复稀疏系数  $\mathbf{X}$ 。对此,本文选用  $L_1$ -合成算法<sup>[15]</sup>,虽然不能得到精确的系数  $\mathbf{X}$ ,但是可以通过求解  $\mathbf{F} = \mathbf{DX}$  得到较为精确的原始数据。 $L_1$ -合成算法的目标为

$$\begin{aligned} \min (\|x\|_1) \\ \text{s. t. } ADx = Y \\ F' = Dx \end{aligned}$$

式中: $x$ 为系数向量 $X$ 的优化解; $\|x\|_1$ 为 $x$ 的 $L_1$ 范数。

$L_1$ -合成算法首先在以系数向量 $X$ 的 $L_1$ 范数最小为目标,在式(3)约束下求解系数向量的优化解 $x$ 。求解 $L_1$ 范数是一个凸优化问题,通过线性规划问题来解决,众多凸优化工具包为算法的实现提供了极大的便利。然后,通过系数向量 $X$ 与原始数据 $F$ 之间的关系求解原始数据。

由于本文所采用的 K-SVD 字典相关度较高,高斯矩阵 $A$ 与 $D$ 相乘所得的矩阵 $Q$ 不再满足 RIP 条件,一般来说系数向量 $X$ 不能精确恢复。但幸运的是,在数据恢复与解压缩中,通过压缩后数据 $Y$ 求解系数向量 $X$ 并非的最终目标,最为关键的是原始数据 $F$ 的精确恢复,各项数据见图4。在 $L_1$ -合成算法下,依然可以在 $F=DX$ 的约束下求解较为精确的原始数据。



图4 不同域的压缩传感数据

Fig. 4 Compressed sensing data in different domains

### 3 实例结果分析

本文以美国明尼苏达州附近高速公路线圈传感器采集的交通流量数据为例,进行相关验证。

#### 3.1 试验数据

线圈传感器网络布设在美国明尼苏达州明尼阿波利斯市与圣保罗市之间的高速公路上,线圈传感器分布见图5,圆圈为线圈传感器,I94为明尼苏达州明尼阿波利斯市与圣保罗市之间的一条高速公路。

线圈传感器构成的传感器网络包含约900个检测站,根据车道数量不同,每个检测站包含1~4个线圈传感器。检测站中线圈传感器的组织形式见图6。检测站的地理位置及相关信息见表1。

表1 检测站地理位置

Tab. 1 Geographical locations of inspection stations

检测站	经度/(°)	纬度/(°)	所属高速路	方向
1	93.40	45.08	I-94	北
2	93.31	45.07		



图5 线圈传感器分布

Fig. 5 Distribution of coil sensors



图6 车道与线圈传感器的关系

Fig. 6 Relationship between coil sensors and lanes

线圈传感器的检测值包括车流量与车辆占有率,其中检测站2中线圈传感1在时间段1~20的检测值见表2,时间段之间的间隔为30s。

布设在高速公路上约4000个线圈传感器不间断工作,产生的数据量相当可观,因而必须对其数据进行压缩,以便于数据传输与分析。线圈传感器设备主要由布设在每条车道下方的感应线圈以及安装在路侧的控制主机与数据传输设备构成,控制主机的计算能力与存储能力均较为有限,因而有必要设计较为简便的数据压缩算法,以适应数据压缩端有限的计算能力,同时实现高效数据压缩。为此实例分析中选取该数据源,试验包括训练字典的构建、原始数据的稀疏表达以及原始数据的压缩与解压缩。在试验中,原始数据的采样间隔为30s,选用1h的交通流量数据为原始数据 $F$ ,原始数据量 $N$ 为120。选用10个位置的线圈传感器1个月采集的交通流量数据为样本进行训练,用于获取训练字典。

#### 3.2 字典下线圈数据的稀疏表达

试验中对样本数据采用 K-SVD 算法构建训练字典。将10个线圈传感器1个月内采集到的交通流量数据,以小时为单位划分为12000个数据单元,

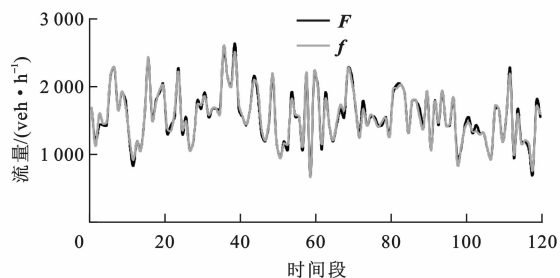
表 2 线圈传感器的检测值

Tab. 2 Detection values of coil sensors

时间段	车流量/(veh·h <sup>-1</sup> )	占有率/%
1	2 760	10
2	1 680	7
3	1 920	6
4	2 400	10
5	2 520	6
6	1 320	7
7	1 440	7
8	2 880	11
9	2 160	6
10	1 440	8
11	2 040	7
12	2 040	6
13	2 280	10
14	2 280	4
15	960	8
16	2 160	8
17	1 440	10
18	3 480	10
19	2 640	13
20	2 400	7

得到维度为  $120 \times 120\ 000$  的训练样本。从样本中抽取 2 000 个数据初始化字典  $D$ , 字典维度为  $120 \times 2\ 000$ 。

用训练样本对字典  $D$  采用 K-SVD 方法训练, 迭代 200 次后得到维度为  $120 \times 2\ 000$  的训练字典后, 用正交匹配追踪算法<sup>[18]</sup>从字典中找到与原始数据  $F$  最为匹配 10 个原子, 对原始数据进行逼近, 实现对交通流量数据的稀疏表达。稀疏分解重构数据为  $f$ 。原始数据  $F$  与稀疏分解重构数据  $f$  的比较见图 7。稀疏表达的相对误差为 0.033 2。原始数据在 K-SVD 字典下的稀疏表达误差分析见表 3。从字典中选择

图 7  $F$  与  $f$  的比较Fig. 7 Comparison of  $F$  and  $f$ 

10 个原子就能够很好地表达原始数据。由于选用原子个数远小于原始数据长度, 车流量数据在字典下可以实现稀疏表达。

表 3 稀疏表达误差分析

Tab. 3 Error analysis of sparse representation

原始数长度	字典原子个数	稀疏表达选用原子个数	相对误差 $\ F-f\ _2 / \ F\ _2$
120	2 000	10	0.033 2

### 3.3 传感矩阵实现数据压缩

考虑到高斯随机矩阵能够以很高概率满足 RIP 条件, 试验中选取高斯随机矩阵为传感矩阵, 维度为  $30 \times 120$ 。通过高斯传感矩阵对原始数据的线性投影, 将原始数据长度从 120 个降低到 30 个, 压缩比为 4:1。

### 3.4 数据解压缩

采用  $L_1$ -合成算法进行数据解压缩, 解压缩后的数据为  $F'$ 。解压缩后数据与原始数据的相对误差为 0.060 8。解压缩数据与原始数据的比较见图 8。压缩比为 4:1 时, 压缩传感数据压缩误差见表 4。从表 4 可以看出, 对压缩后的数据进行解压后, 解压缩相对误差为 0.060 8, 能够比较精确地恢复出原始数据。

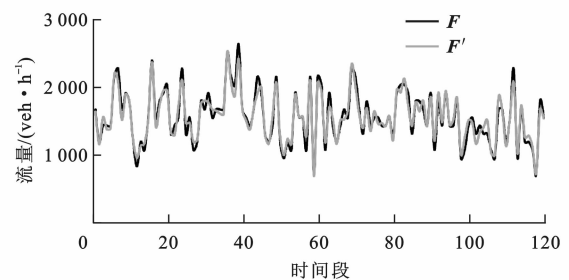
图 8  $F$  与  $F'$  的比较Fig. 8 Comparison of  $F$  and  $F'$ 

表 4 压缩传感误差分析

Tab. 4 Error analysis of compressed sensing

原始数据长度	压缩后数据长度	压缩比	相对误差 $\ F-F'\ _2 / \ F\ _2$
120	30	4:1	0.060 8

由上述试验可以看出: 基于压缩传感的交通流量数据压缩方法, 可以在传感器数据压缩端无需考虑变换矩阵的情况下, 以较为简单的编码策略实现高效的数据压缩, 解压缩之后, 可以得到较为精确的原始数据。

### 3.5 压缩传感方法与传统方法的对比

与传统变换域数据压缩相比, 压缩传感方法除了可以避免变换矩阵的选取, 采取较为通用的随机

投影策略即可实现数据压缩之外,在数据压缩端的计算量也较小。

本文对压缩传感方法与传统方法在数据压缩端的计算量进行了对比分析。原始交通流量数据为  $\mathbf{F}$ , 维度为  $N$ , 高斯随机矩阵  $\mathbf{A}$  维度为  $M \times N$ 。线性投影实现数据压缩通过矩阵  $\mathbf{A}$  与向量  $\mathbf{F}$  相乘完成, 在这个过程中, 需要进行  $MN$  次乘法和  $M(N-1)$  次加法, 计算量为  $O(MN)$ ,  $O$  为函数数量级。

传统变换域数据压缩建立在正交变换基础上, 变换矩阵维度为  $N \times N$ 。在不考虑快速算法时, 正交变换矩阵与向量  $\mathbf{F}$  的相乘需要进行  $N^2$  次乘法和  $N(N-1)$  次加法, 计算量为  $O(N^2)$ 。二者之间的对比见表 5。

表 5 数据压缩端对比

Tab. 5 Comparison at data compression side

	乘法次数	加法次数	计算量
正交变换	$N^2$	$N(N-1)$	$O(N^2)$
压缩传感	$MN$	$M(N-1)$	$O(MN)$

在压缩传感中, 考虑到  $M$  远小于  $N$ , 其计算量明显小于传统变换域数据压缩中的正交变换, 因而更加适合计算资源有限的传感器数据压缩端。

## 4 结 语

本文将压缩传感首次引入交通数据压缩领域, 提出了基于压缩传感的交通流量数据压缩方法, 该方法直接对原始数据进行线性投影就能实现较高压缩比的数据压缩。

试验首先证明了线圈传感器交通流量数据可以在经过 K-SVD 方法训练过的字典上实现稀疏表达, 在此基础上, 基于压缩传感的数据压缩能够实现较高的压缩比; 通过  $L_1$ -合成算法可以比较完美地实现解压缩。

基于压缩传感的交通流量数据压缩方法, 在数据压缩端无需考虑选择哪种类型的基函数, 直接运用随机投影这种非自适应性的压缩策略, 就能实现原始数据的高效压缩。与传统的基于变换域的数据压缩相比, 基于压缩传感的交通流量数据压缩方法更加适合计算资源有限的传感器交通数据压缩。

此外, 在数据压缩过程中, 考虑到传感矩阵维度低于正交变换矩阵, 在数据端压缩传感方法的计算量也远小于传统变换域数据压缩。

但是, 文中采用的 K-SVD 训练字典不够稳定, 对于某些时段的交通流量数据稀疏表达效果还不好。在接下来的研究中, 将在考虑流量数据的拥堵

特性的基础上, 制作性能更加优良的训练字典。

## 参考文献:

## References :

- [1] 肖 扬, 鲁凌云, 高 爽, 等. 基于二维离散小波变换的智能交通系统数据去噪声压缩[J]. 北京交通大学学报, 2004, 28(5): 1-5.  
XIAO Yang, LU Ling-yun, GAO Shuang, et al. Traffic data denoising compression for intelligent traffic systems based on 2-D discrete wavelet transformation[J]. Journal of Beijing Jiaotong University, 2004, 28(5): 1-5. (in Chinese)
- [2] 练秋生, 王成儒, 孔令富. 心电图压缩算法中的小波基选择[J]. 计算机工程与应用, 2002, 38(18): 6-8.  
LIAN Qiu-sheng, WANG Cheng-ru, KONG Ling-fu. The choice of wavelet in electrocardiogram compression algorithm[J]. Computer Engineering and Applications, 2002, 38(18): 6-8. (in Chinese)
- [3] COIFMAN R R, WICKERHAUSER M V. Entropy-based algorithms for best basis selection[J]. IEEE Transactions on Information Theory, 1992, 38(2): 713-718.
- [4] 蔡敦虎. 多种小波基的图像去噪性能研究[D]. 武汉: 武汉大学, 2003.  
CAI Dun-hu. The research on the performance of manifold wavelet basis in image denoising[D]. Wuhan: Wuhan University, 2003. (in Chinese)
- [5] 费铭薇, 乐全明, 张沛超, 等. 电力系统故障录波数据压缩与重构小波基选择[J]. 电力系统自动化, 2005, 29(17): 64-67, 97.  
FEI Ming-wei, YUE Quan-ming, ZHANG Pei-chao, et al. Wavelets selection of compression and reconstruction algorithm based on digital recorded data from a faulted power system[J]. Automation of Electric Power Systems, 2005, 29(17): 64-67, 97. (in Chinese)
- [6] 魏玉芬, 梦艳君. 图像编码压缩小波基的选择[J]. 装备制造技术, 2009(4): 49-50, 61.  
WEI Yu-fen, MENG Yan-jun. The choice of orthogonal wavelets base in image compression[J]. Equipment Manufacturing Technology, 2009(4): 49-50, 61. (in Chinese)
- [7] 赵志强, 张 毅, 胡坚明, 等. 基于 PCA 和 ICA 的交通流量数据压缩方法比较研究[J]. 公路交通科技, 2008, 25(11): 109-118.  
ZHAO Zhi-qiang, ZHANG Yi, HU Jian-ming, et al. Comparative study of PCA and ICA based traffic flow compression[J]. Journal of Highway and Transportation Research and Development, 2008, 25(11): 109-118. (in Chinese)
- [8] QU Li, HU Jian-ming, ZHANG Yi. A flow volumes data compression approach for traffic network based on principal component analysis[C] // IEEE. Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference. Seattle: IEEE, 2007: 125-130.
- [9] 耿彦斌, 于 雷, 武 旭, 等. 基于信号处理技术的 ITS 数据压缩方法与应用[J]. 土木工程学报, 2006, 39(11): 107-113.

(下转第 126 页)