

DOI: 10.7652/xjtuxb202002008

采用压缩感知的流程工业异常监测数据 检验与修复方法

徐光南, 高智勇, 梁艳杰, 高建民, 刘倩倩, 程亚辉

(西安交通大学机械制造系统国家重点实验室, 710049, 西安)

摘要: 针对由于流程工业生产系统具有变量多、维度高、耦合关系复杂等特性导致的异常监测数据检验与修复难题,提出了一种采用压缩感知的流程工业异常监测数据检验与修复方法。压缩感知算法能够使用极少的测量数据重构出原稀疏信号或能够在某稀疏基上得到稀疏表达的信号。结合压缩感知原理及重构成功率指标,构建出重构非线性系统方程的数学模型,该模型仅使用少量的监测数据就可以重构出反映状态感知网络节点动态关系的系统方程。将重构的系统方程代入考虑系统偏差的求解结构中得到监测数据的解析值,对比监测数据的实际值与解析值,若实际值超差则用解析值替代实际值,从而实现异常监测数据的检验与修复。通过某煤化工企业压缩机子系统应用案例验证所提方法的有效性,结果表明:在监测数据正常的情况下,实际数据与修复数据的相对误差波动极小,而当监测数据出现异常时,相对误差发生剧烈变化,证明提出的算法能够检测到异常监测数据,并能够较好地恢复监测数据的原貌。

关键词: 压缩感知;流程工业;重构;异常数据检验;数据修复

中图分类号: TH17 **文献标志码:** A **文章编号:** 0253-987X(2020)02-0059-12

Verification and Restoration Method of Abnormal Monitoring Data by Compressive Sensing for Process Industry

XU Guangnan, GAO Zhiyong, LIANG Yanjie, GAO Jianmin, LIU Qianqian, CHENG Yahui

(State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Aiming at verifying and restoring abnormal monitoring data from the production system of process industry with multi-variables, high-dimensionality and complex coupling relationship, a verification and restoration method for abnormal monitoring data is proposed by compressive sensing. The compressive sensing algorithm can reconstruct original sparse signal with very few measurement data or obtain the signal of sparse expression on a sparse basis. Combining the compressive sensing algorithm with the success rate of reconstruction, a mathematical model is established, which can reconstruct the system equation reflecting the dynamic relationship of the nodes in state-aware network with only a small number of monitoring data. The analytical value of monitoring data is obtained by substituting the reconstructed system equation into the solving structure considering system deviation. Compared the actual value with the analytical value of monitoring data, if the actual value is overproof, the analytical value is used instead of the actual value to verify and restore abnormal monitoring data. The validity of

收稿日期: 2019-06-25。 作者简介: 徐光南(1993—),男,硕士生;高智勇(通信作者),男,副教授。 基金项目: 国家自然科学基金资助项目(51375375)。

网络出版时间: 2019-10-16

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20191015.1553.002.html>

the proposed method is verified by an application case of compressor subsystem in a coal chemical enterprise. When monitoring data are normal, the relative error between the actual data and the restored data fluctuates slightly; when monitoring data are abnormal, the relative error changes dramatically.

Keywords: compressive sensing; process industry; reconstruct; abnormal data verification; data restoration

状态感知网络是由传感器、自动化仪表等组成的感知系统状态的信息采集网络,是化工生产、核能发电、航空航天等复杂机电系统进行自动化、智能化精准控制的基础。化工生产系统长期处于高温高压、有毒有害等恶劣工况,其状态感知网络中的监测仪表不可避免的会因为震动、温度、环境污染、部件老化等原因而发生故障或精度下降,从而导致状态感知网络中的监测节点信息失准,由于生产系统各单元之间高度耦合,因此会造成系统中多个耦合节点之间的信息冲突,进而引发系统复杂故障或异常激励,导致灾难性后果^[1]。根据国家应急管理部的统计数据,2008~2017年全国共发生重特大化工生产安全事故129起,其中约7%由仪表故障引起,造成了数亿元的经济损失。由于化工生产系统具有变量多、维度高、耦合关系复杂等特性^[2],导致异常监测数据的检测与修复成为难题,需提出新的研究方法。

目前,国内外针对流程工业异常监测数据的检验与修复方法已有一些研究成果,通常是通过判断监测数据的特征值是否偏离正常范围来实现的。去趋势波动分析(DCCA)^[3]是一种处理非线性非平稳性信息序列间长程幂律相关性的方法,用来对两个同步采集的非平稳信息序列间的耦合性进行分析。由于DCCA提取的分形特性过于单一,因此在其基础上提出了多重分形去趋势交叉分析(MF-DXA)^[4],用以分析两个交叉相关非平稳信息序列间耦合关系的多重分形特性。将MF-DXA中的去趋势协方差扩展到多变量间得到耦合去趋势波动分析(CD-FA)^[5],可以获取多变量耦合多重分形特征谱,最终实现异常监测数据的检验。

此外,广义似然比法(GLR)^[6]、节点检验法(NT)^[7]和整体测量检验法(GT)^[8]等方法,都能根据测量数据的统计特性检验监测数据时间序列中的异常点,但这些方法的实现都要以变量间明确的平衡约束条件为前提,因此,难以直接应用于运行机制复杂的流程工业。文献[9]提出了基于改进局部离群因子(LOF)的流程工业异常监测数据检验方法,

提高了数据实时检验的速度和效率,降低了传统算法的运算时间,但该方法局限于单变量异常监测数据的检验,无法适用于多维变量。以上方法虽然从不同角度实现了异常监测数据的检验,但无法对异常监测数据进行修复。文献[10]通过时间序列预测技术实现了异常监测数据的检验,但检验效果很大程度上取决于神经网络的收敛参数设置以及训练样本选取,并且输入数据与预测结果之间,存在着我们无法洞悉的隐层也称为“黑箱”,导致在实际应用中受到限制。文献[11]中的多元非线性回归方法可以用于多变量参与下的数据预测,到目前为止还没有一种非线性回归模型建模方法,通常需要把非线性回归模型转化为线性回归模型,然而这种非线性线性化已经被证明在变量替换涉及到参数的情形是不成立的,不适用于生产关系复杂的流程工业。

本文针对传统异常数据检验方法的不足,提出了一种采用压缩感知的流程工业异常监测数据检验与修复方法,利用压缩感知理论构建了重构非线性系统方程的数学模型,仅使用少量的监测数据就可以重构出反映状态感知网络节点动态关系的系统方程,并通过重构的系统方程得到监测数据的解析值,从而实现异常监测数据的检验与修复。最后通过某煤化工企业压缩机组子系统应用案例验证了本文所提方法的有效性和可行性。

1 压缩感知理论

常规把模拟信号转变为数字信号的过程,离不开奈奎斯特(Nyquist)采样定律,但该采样定律经常导致过多的冗余采样,为了解决在处理多维海量数据时所需要面临的储存和传输问题,通常采用数据压缩技术。传统的数据压缩过程首先通过模数转换器(A/D)完成采样,其中含有大量的冗余数据,而后通过变换域挖掘信号的稀疏性,最后通过压缩算法实现压缩,这个过程其实造成了巨大的浪费,首先采集了大量的冗余数据,而后又在压缩过程中把这些冗余数据去除掉^[12]。为了能够通过远低于奈奎斯特采样频率的采样完全恢复原始信号,文献[13]建

立了基于压缩感知理论的采样体系。压缩感知理论指出稀疏的或具有稀疏表达的有限维数的信号可以利用远少于奈奎斯特采样数量的线性、非适应的测量值无失真地重构出来。该理论一经提出,便在信息论、信号/图像处理、医疗成像、射电天文、模式识别、光学/雷达成像和信道编码等诸多领域引起高度关注,并取得了极佳的应用效果,然而在数据的检验检测方面该理论尚无广泛应用。

1.1 压缩感知的数学模型

针对长度为 N 的一维信号 $x \in \mathbf{R}^N$, 压缩感知问题可以描述为在已知某个测量矩阵 $\Phi \in \mathbf{R}^{M \times N}$ ($M \ll N$) 以及信号 x 在该矩阵下的线性测量值 $y \in \mathbf{R}^M$ 的基础上, 求解方程组

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

得到原始信号 x 。由于 y 的维数远远低于 x 的维数, 信号 x 的重构问题是一个欠定问题, 即式(1)有无穷多个解, 无法重构出原始信号。如果原始信号 x 的稀疏度为 K (即 x 中只包含 K 个非零值), 并且 y 与 Φ 满足一定条件, 能够证明信号 x 可以由测量值 y 通过求解最小化 L_0 范数问题精确重构

$$\hat{x} = \operatorname{argmin} \|x\|_0; \quad \text{s.t. } \Phi x = y \quad (2)$$

式中: $\|x\|_0$ 为向量的 L_0 范数, 表示向量 x 中非零元素的个数。Candès 等人提出, 如果要精确重构 K 稀疏信号 x , 测量次数 M (即 y 的维数) 必须满足 $M = O(K \lg(N))$, 并且矩阵 Φ 必须满足约束等距性 (RIP) 条件^[14]。

常见的自然信号在时域内几乎都是不稀疏的, 因而上述信号重构过程不能直接应用于自然信号的重构, 信号稀疏表示理论指出, 自然信号 f 可以通过某种变换 Ψ 进行稀疏表示, 即 $f = \Psi x$, x 为该信号在 Ψ 变换域的稀疏表示^[15]。考虑测量公式 $y = \Phi f$, 并且 f 是可以稀疏表示的, 即 $f = \Psi x$, 则有

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

式中: $A = \Phi \Psi$ 为 $M \times N$ 的矩阵, 称为感知矩阵。自然信号的压缩感知线性测量过程如图 1 所示。图中 y 可以看作是稀疏信号 x 关于感知矩阵 A 的测量

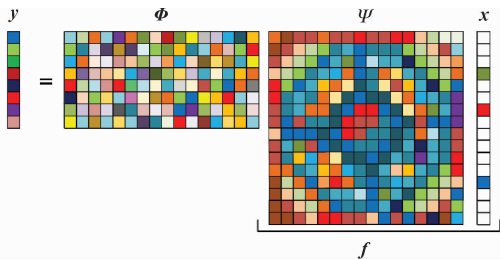


图1 自然信号的压缩感知线性测量过程

值, 如果 A 满足约束等距性条件, 可以通过求解一个类似式(2)的最小化 L_0 范数问题来重构稀疏信号 x , 即

$$\hat{x} = \operatorname{argmin} \|x\|_0; \quad \text{s.t. } Ax = y \quad (4)$$

得到 f 的稀疏表示 $\hat{x}^*$ 后, 可以进一步由变换基 Ψ 通过下式求解出原始信号 f'

$$f' = \Psi x^* \quad (5)$$

具体的压缩感知流程如图 2 所示。

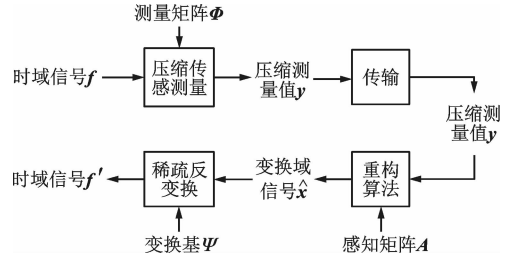


图2 压缩感知理论框架下的信号采集与重构

1.2 信号重构算法

信号重构算法是压缩感知理论的核心, 是指由 M 维测量向量 y 重构长度为 N ($M \ll N$) 的稀疏信号 x 的过程。早期研究采用最小化 L_2 范数作为约束, 然而针对该优化问题的求解通常并不具备稀疏性。Candès 等人证明了信号重构问题可以通过求解最小化 L_0 范数问题加以解决, 但 Donoho 指出, 最小化 L_0 范数问题是一个不确定多项式难题 (NP-hard) 问题^[16], 需要穷举向量 x 中非零元素的所有 C_N^K 排列方式, 因而无法求解。由此, 研究人员提出了一系列的信号重构算法, 主要可以分为以下 3 类^[17]。

第一类是凸优化算法, 由于最小化 L_0 范数问题是一个 NP-hard 问题, 故用 L_1 范数代替 L_0 范数, 得到另外一种最优化问题—— L_1 范数凸优化

$$\hat{x} = \operatorname{argmin} \|x\|_1; \quad \text{s.t. } \Phi x = y \quad (6)$$

已有人证明式(6)和式(2)是等价的。式(6)是一个凸优化问题, 可以转化成线性规划 (LP) 问题加以求解^[18], 计算复杂度为 $O(N^3)$, 这种方法也称为基追踪算法 (BP)^[19], 是最具代表性的凸优化算法。从某种意义上说, BP 是一种规则而非算法, 需要从完备的基字典 (集合) 中寻找信号的最稀疏表示, 即用尽可能少的基尽可能精确地表示原始信号。可以使用内点法或单纯形法的实现基追踪算法, 这类算法的信号恢复误差小, 重构效果好, 但计算复杂度大, 不适于应用在维度较大的信号。另外两种比较常见的凸优化算法包括: 欠定系统局域解 (FOCUSS) 算法, 该算法用 L_p 范数 ($p \leq 1$) 替代 L_0 范数

求解最优化问题;SL0 算法,通过引入光滑函数序列去逼近 L_0 范数,从而将 L_0 范数最小化问题转化为光滑函数的最优化问题。

第二类是贪婪算法,基于约束等距性,在最小化 L_1 范数条件下可以保证使用 $O(K \ln(N/K))$ 个观测数据就能够成功恢复稀疏信号。在实践中发现,当信号的维度较大时,比如百万数量级,利用最小化 L_1 范数问题恢复信号的代价极大,计算时间变得难以接受,贪婪算法可以通过迭代的方法寻找稀疏向量的支撑集,并且使用受限支撑最小二乘估计实现稀疏信号的快速恢复。典型的贪婪算法包括匹配追踪(MP)算法、正交匹配追踪(OMP)算法、正则正交匹配追踪(ROMP)算法、和引入回溯思想的压缩采样匹配追踪(CoSAMP)算法^[20]、子空间追踪(SP)算法^[21]等。以上算法都建立在已知稀疏度 K 的基础上,然而实际应用中, K 往往是未知的,因此出现了对 K 自适应的稀疏自适应匹配追踪(SAMP)算法^[22],通过固定步长、逐步逼近进行重构,可以在 K 未知的情况下获得较好的重构效果,速度也远快于 OMP 算法。

第三类是基于贝叶斯理论提出的重构算法,假定信号中的元素来自于具有突出稀疏特性的先验概率分布,然后从随机测量中重构那些符合此概率分布的非零元素。目前,这类算法包括新期望极大值(EM)算法、贝叶斯压缩感知(BCS)算法^[23]、稀疏贝叶斯学习(SBL)算法等。不过以上算法没有一种能够基于一定数量的测量值无失真地重构出原始信号,其原因是在贝叶斯信号建模的框架下,没有明确的关于无失真或重构误差的概念。

2 非线性系统方程的重构

2.1 重构非线性系统方程的数学模型

为了实现非线性系统方程的重构,需要将系统的向量场扩展为长度合适的级数形式,然后利用压缩感知理论重构出级数展开式中各项的系数^[24]。假设非线性系统方程可以写成如下形式

$$\dot{x}^* = F(x) \quad (7)$$

式中: $x \in \mathbf{R}^m$ 代表一组可从系统外部访问的变量; F 是作用在 $\mathbf{R}^m$ 上的平滑函数; x^* 与变量 x 相关,在不同性质的非线性系统中有不同的物理意义。变量 x 的第 k 个分量 $x_k (k=1, \dots, m)$ 的方程可以按级数形式表示为

$$F(x_k) = \sum_{l_1=0}^n \sum_{l_2=0}^n \cdots \sum_{l_m=0}^n a_{l_1, \dots, l_m} x_1^{l_1} x_2^{l_2} \cdots x_m^{l_m} \quad (8)$$

式中:共包含 $(1+n)^m$ 个乘积项,每个乘积项对应的待定系数为 $a_{l_1, \dots, l_m} \in \mathbf{R}^M$ 。

下面,以系统中的一个变量为例说明重构非线性系统方程的数学模型。由式(8)可知,待定系数的总数是 $(1+n)^m$,即稀疏向量 $\mathbf{a}$ 包含 $(1+n)^m$ 个元素。假设系统只包含 3 个变量 x, y 和 z (即 $m=3$),且级数展开式的最高幂次为 3 (即 $n=3$),则变量 x 的方程可以表示为

$$F(x) = a_{0,0,0} x^0 y^0 z^0 + a_{1,0,0} x^1 y^0 z^0 + \cdots + a_{3,3,3} x^3 y^3 z^3 \quad (9)$$

使用向量 $\mathbf{a} = [a_{0,0,0}, a_{1,0,0}, \dots, a_{3,3,3}]^T$ 表示方程 $F(x)$ 的系数。若可以在任意时刻 t 对变量 x, y 和 z 进行测量,则能得到向量

$$\mathbf{g}(t) = [x(t)^0 y(t)^0 z(t)^0, \dots, x(t)^3 y(t)^3 z(t)^3] \quad (10)$$

因此式(9)也可以表示为

$$F(x(t)) = \mathbf{g}(t) \cdot \mathbf{a} \quad (11)$$

由式(7)、式(11),可得

$$x^*(t) = F(x(t)) = \mathbf{g}(t) \cdot \mathbf{a} \quad (12)$$

选择变量 x 在 $t_1, t_2, \dots, t_w$ 时刻的 $x^*(t)$ 值作为测量向量

$$\mathbf{X} = [x^*(t_1), x^*(t_2), \dots, x^*(t_w)]^T \quad (13)$$

由式(12)、式(13)得到满足压缩感知理论一般形式的方程 $\mathbf{X} = \mathbf{G} \cdot \mathbf{a}$,展开式如下

$$\begin{bmatrix} x^*(t_1) \\ x^*(t_2) \\ \vdots \\ x^*(t_w) \end{bmatrix} = \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_w) \end{bmatrix} \cdot \mathbf{a} \quad (14)$$

为了使矩阵 $\mathbf{G}$ 满足约束等距性,需要对 $\mathbf{G}$ 进行归一化处理,即每一列的元素除以该列元素的 L_2 范数

$$\mathbf{G}'_{ij} = \mathbf{G}_{ij} / L_2(j) \quad (15)$$

式中: $L_2(j) = \sqrt{\sum_{i=1}^M [G_{ij}]^2}$ 。此时方程 $\mathbf{X} = \mathbf{G} \cdot \mathbf{a}$ 改写为 $\mathbf{X} = \mathbf{G}' \mathbf{a}'$,通过压缩感知的信号重构算法得到向量 $\mathbf{a}'$ 后,稀疏向量 $\mathbf{a}$ 可以通过下式得到

$$\mathbf{a}_j = \mathbf{a}'_j / L_2(j) \quad (16)$$

式中: $\mathbf{a}_j$ 和 $\mathbf{a}'_j$ 分别代表向量 $\mathbf{a}$ 和 $\mathbf{a}'$ 的第 j 个分量, $L_2(j)$ 同式(15)。当需要确定其他变量(如 y)的方程时,不需要改变矩阵 $\mathbf{G}$,只需将测量向量替换为 $\mathbf{X} = [y^*(t_1), y^*(t_2), \dots, y^*(t_w)]$,重复上述步骤最终就能得到完整的非线性系统方程。

2.2 重构 Lorenz 方程

以典型的非线性系统方程——Lorenz 方程的重构为例说明上述数学模型^[19]。Lorenz 方程的数

学表达式为

$$\left. \begin{aligned} \dot{x} &= -\sigma(x-y) \\ \dot{y} &= rx-y-xz \\ \dot{z} &= -bz+xy \end{aligned} \right\} \quad (17)$$

式中: σ 、 r 和 b 是方程的参数,当 $\sigma=10$ 、 $r=28$ 、 $b=8/3$ 时,系统会进入混沌状态。假设系统按设定参数运行但方程表达式未知,只能获取有限长度的时间序列 $x(t)$ 、 $y(t)$ 和 $z(t)$ 。为了便于说明,仅将系统方程按级数形式展开至三次项

$$\left. \begin{aligned} \dot{x} &= a_0 + a_1x + a_2y + \cdots + a_{17}x^3 + a_{18}y^3 + a_{19}z^3 \\ \dot{y} &= b_0 + b_1x + b_2y + \cdots + b_{17}x^3 + b_{18}y^3 + b_{19}z^3 \\ \dot{z} &= c_0 + c_1x + c_2y + \cdots + c_{17}x^3 + c_{18}y^3 + c_{19}z^3 \end{aligned} \right\} \quad (18)$$

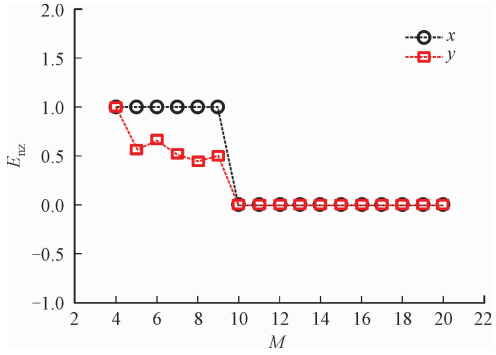
将 Lorenz 方程 3 个分量 $\dot{x}$ 、 $\dot{y}$ 、 $\dot{z}$ 的展开式分别代入重构非线性系统方程的数学模型,每个分量都有 20 个待定系数且其中非零系数的数量(即稀疏度 K)未知。分别选择属于凸优化算法的 BP 算法和属于贪婪算法的 SAMP 算法对 Lorenz 方程进行重构,因为这两种算法都不需要已知稀疏度 K ,并且可以通过重构结果得到算法各自的特点,为下文的算法选择提供依据。

以 1 Hz 的频率从时间序列 $x(t)$ 、 $y(t)$ 、 $z(t)$ 中分别采集 M 个数据用于 Lorenz 方程的重构。方程的重构误差可以分为两类:一类是非零系数(即原方程中存在的项)的重构误差 E_{nz} ;另一类是零系数(即原方程中不存在的项)的重构误差 E_z 。 E_{nz} 和 E_z 越接近于 0 说明重构误差越小,当 M 达到一定的数值 $M_{\min}$ 时,随着 M 的增加, E_{nz} 和 E_z 保持 0 值不变,此时的 $M_{\min}$ 值就是能够实现方程重构的最小数据量。

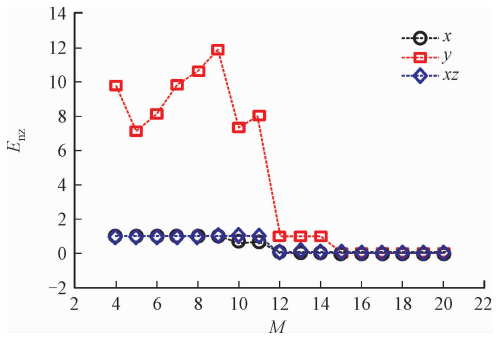
重构误差 E_{nz} 与采样点数 M 之间的关系如图 3 所示,从图中可以看出对于非零系数, $\dot{y}$ 分量中两类算法实现方程重构的最小数据量都是 15,其余分量 BP 算法实现方程重构的最小数据量均要小于 SAMP 算法,且重构误差 E_{nz} 随 M 的增加波动程度较小,说明 BP 算法的重构效果要好于 SAMP 算法。重构误差 E_z 与采样点数 M 之间的关系如图 4 所示,两类算法实现方程重构的零系数最小数据量都是 15,但以 E_z 随 M 的波动情况来看, BP 算法的重构效果仍要好于 SAMP 算法。结合从图 3 和图 4 中得到的结论可知,当数据量 $M \geq 15$ 时,两类算法都能以极小的重构误差重构出 Lorenz 方程。由于 SAMP 算法的求解速度要远快于 BP 算法,因此算法的选择要依数据特点而定:当处理低维信号时,由于数据量较少,为了得到更精确的重构结果,应选择

BP 算法;当处理高维信号时,受运算效率的影响,可以优先选择 SAMP 算法。

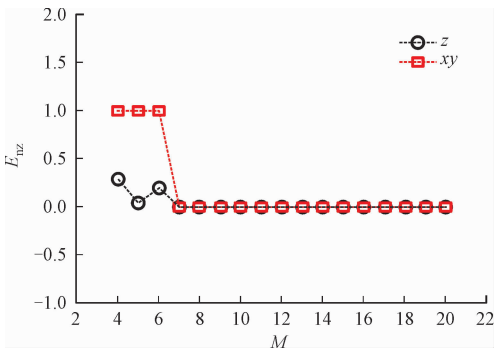
结合从图 3 和图 4 中得到的结论可知,当数据量 $M \geq 15$ 时,两类算法都能以极小的重构误差重构



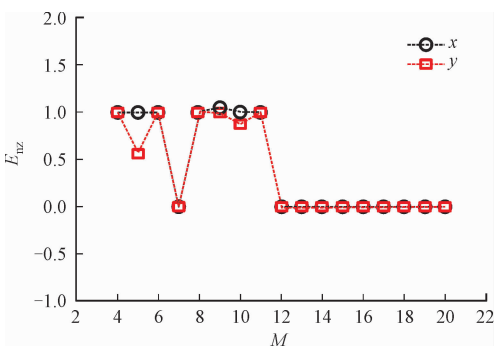
(a)BP 算法下 $\dot{x}$ 分量



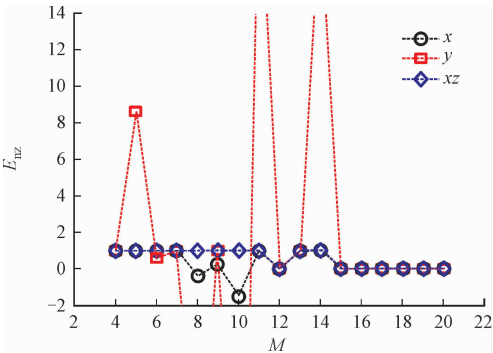
(b)BP 算法下 $\dot{y}$ 分量



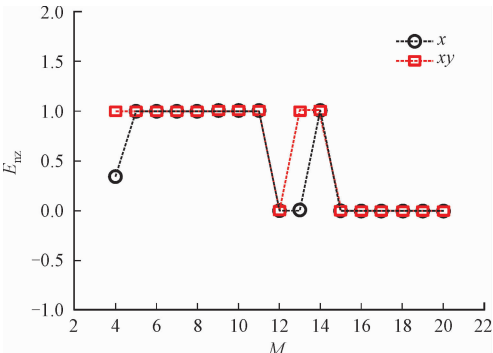
(c)BP 算法下 $\dot{z}$ 分量



(d)SAMP 算法下 $\dot{x}$ 分量

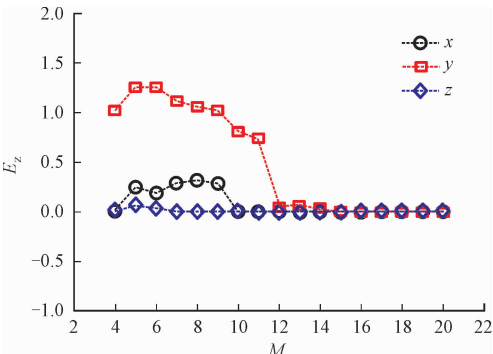


(e)SAMP 算法下 y 分量

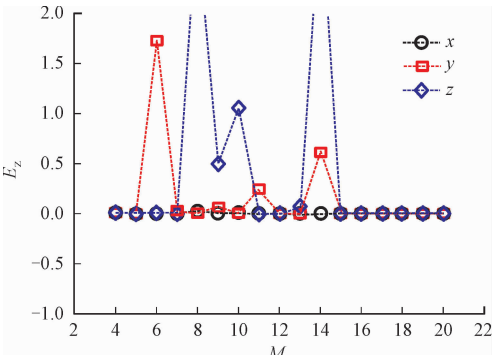


(f)SAMP 算法下 z 分量

图 3 重构误差 E_{nz} 与采样点数 M 之间的关系



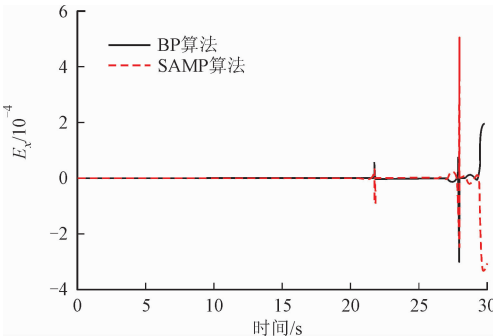
(a)BP 算法



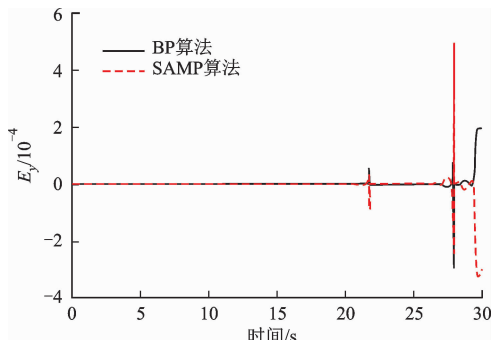
(b)SAMP 算法

图 4 重构误差 E_z 与采样点数 M 之间的关系

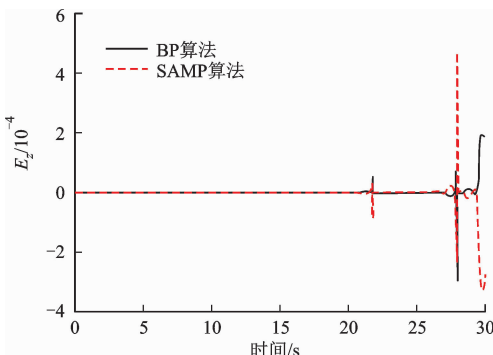
出 Lorenz 方程。当 $M=15$ 时,BP 算法和 SAMP 算法重构的 Lorenz 方程以相同的初始条件产生的时间序列与真实时间序列的相对误差随时间的变化曲线见图 5,其中 E_x 、 E_y 、 E_z 分别表示时间序列 $x(t)$ 、 $y(t)$ 、 $z(t)$ 的相对误差。重构 Lorenz 方程所需的 15 个数据全部采集自时间序列的 0~14 s,从图中可以看出,14~20 s 左右的一段时间内 E_x 、 E_y 、 E_z 保持了同前 14 s 一样的稳定状态,证明通过重构非线性系统方程的数学模型,仅使用有限的时间序列就可以对系统方程在数据采集点及之后的一段时间内进行精确重构。随着时间的推移,20 s 后的某一时刻起,3 者都出现了剧烈的波动,说明从此时刻开始重



(a)时间序列 $x(t)$ 的相对误差 E_x 随时间的变化



(b)时间序列 $y(t)$ 的相对误差 E_y 随时间的变化



(c)时间序列 $z(t)$ 的相对误差 E_z 随时间的变化

图 5 Lorenz 时间序列相对误差随时间的变化曲线

构的 Lorenz 方程已经无法反映系统的真实运行状态。这是因为混沌状态下的非线性系统具有初值敏感性,由于数据本身的原因重构误差不可避免,因此方程的微小偏差会随着时间的推移被不断放大。

3 异常监测数据的检验与修复

3.1 状态感知网络节点动态关系建模

实现流程工业异常监测数据的检验与修复,关键是要重构出能够反映状态感知网络节点动态关系的系统方程。在重构系统方程之前,要先根据系统的特点选择合适的系统方程形式,流程工业生产系统是典型的非线性连续系统,一般以微分方程组的形式表示如下

$$\left. \begin{aligned} \dot{x}_1 &= f_1(x_1, \dots, x_n; c_1, \dots, c_n) \\ \dot{x}_2 &= f_2(x_1, \dots, x_n; c_1, \dots, c_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \dots, x_n; c_1, \dots, c_n) \end{aligned} \right\} \quad (19)$$

式中: $x_1, \dots, x_n$ 为状态变量, $c_1, \dots, c_n$ 为控制变量,两者分别形成了系统的状态空间和参量空间。

实际用于重构系统方程的所有数据都是通过集散控制系统(DCS)采集得到的,DCS 监测变量多为工艺变量,如温度、压力、流量等。考虑到变量间影响关系的传递速度等因素,采样间隔一般选择为 1 min,因此采集到的数据都是具有较大时间间隔的时间序列,无法像重构 Lorenz 方程那样,通过 Runge-Kutta 法可以得到任意时间点各变量的数据。由于化工生产环境恶劣,监测数据不可避免地会受到噪声的污染。因为监测数据具有上述缺陷,所以无法得到各状态变量准确的一阶导数,也就无法重构出式(19)的系统方程。原因在于,它是对真实物理系统的抽象表达,却由于系统的复杂性和数据的固有缺陷而无法全面地反映系统的真实运行状态。

此外,重构非线性系统方程的数学模型中,假设系统方程的表达式为级数展开的形式,实际上,系统方程的表达式多种多样,尤其对于生产关系极为复杂的流程工业,用级数展开的形式很难重构出准确的系统方程。

为了摆脱因数据和方程形式而带来的限制,可以在继承监测数据的情况下对系统进行转化,得到只反映简单数值关系的系统方程如下

$$\left. \begin{aligned} x^* &= a_0 + a_1x + a_2y + \dots + a_{17}x^3 + a_{18}y^3 + a_{19}z^3 \\ y^* &= b_0 + b_1x + b_2y + \dots + b_{17}x^3 + b_{18}y^3 + b_{19}z^3 \\ z^* &= c_0 + c_1x + c_2y + \dots + c_{17}x^3 + c_{18}y^3 + c_{19}z^3 \end{aligned} \right\} \quad (20)$$

为了便于说明,假设系统只包含 3 个变量 x 、 y 和 z ,将系统方程按级数形式展开至三次项。式(20)将状态变量的一阶导数替换为 x^* 、 y^* 、 z^* ,表示由系统 A 向系统 B 的转化,系统间的对应关系如图 6 所示。系统 A 和 B 都是状态感知网络,由若干监测节点构成,系统 B 的原型可以是系统 A 的全部或一部分。式(20)中, x 、 y 、 z 表示由各节点产生的信号强度, x^* 、 y^* 、 z^* 表示在各节点测得的信号强度。系统 A 是数据驱动的系统,状态感知网络中的传感器是能够根据确切的物理关系将其他信号转化为电信号的实体,根据上文可知,仅依靠监测数据无法重构出能够全面反映系统真实运行状态的系统方程。系统 B 是模型驱动的系统,状态感知网络中的传感器是以级数展开形式表示的如式(20)中各分量的方程,也可以称之为虚拟传感器,方程中的各项即虚拟传感器的元器件,不同元器件的组合会对虚拟传感器的性能产生影响。虚拟传感器能够显示监测变量的真实状态信息即可,其有效性将在下文得到证明。系统 B 中的监测变量 x 、 y 、 z 都是无量纲的,并且在数值上继承自系统 A 中对应监测变量的监测数据,若能够重构出虚拟传感器的具体形式,也就在小范围内重构出了能够反映状态感知网络节点动态关系的系统方程。

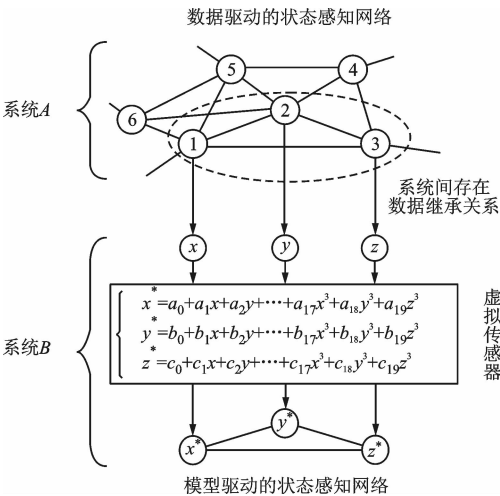


图 6 系统 A 和系统 B 的对应关系

系统方程的重构需要 x 、 y 、 z 和 x^* 、 y^* 、 z^* 的数据,其中 x 、 y 、 z 的数据继承自系统 A,若传感器有效则 x^* 、 y^* 、 z^* 与 x 、 y 、 z 的数值应该接近,接近程度受虚拟传感器的性能影响。因此,在重构系统方程时可以用 x 、 y 、 z 近似替代 x^* 、 y^* 、 z^* 。不过近似替代会使重构非线性系统方程的数学模型中,测量数据与感知矩阵的某列数据在数值上完全一致,导致

感知矩阵不再满足约束等距性条件,所以需要感知矩阵的结构进行优化。

3.2 感知矩阵的优化

为了使感知矩阵满足约束等距性条件并使系统方程达到较好的重构效果,需要对感知矩阵的结构进行优化,优化目标包括感知矩阵的列向量和行向量。为了使感知矩阵满足约束等距性条件并避免当前变量对系统方程的重构影响过大,列向量的优化方法为剔除只包含当前变量的列向量。例如:若当前要重构变量 x 对应的系统方程,则需要剔除列向量 $[x_1 \cdots x_n]^T, [x_1^2 \cdots x_n^2]^T, [x_1^3 \cdots x_n^3]^T$,列向量优化后的感知矩阵如下

$$\begin{bmatrix} 1 & y_1 & z_1 & \cdots & z_1^2 y_1 & y_1^3 & z_1^3 \\ 1 & y_2 & z_2 & \cdots & z_2^2 y_2 & y_2^3 & z_2^3 \\ 1 & y_3 & z_3 & \cdots & z_3^2 y_3 & y_3^3 & z_3^3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 1 & y_n & z_n & \cdots & z_n^2 y_n & y_n^3 & z_n^3 \end{bmatrix} \quad (21)$$

接下来,需要对感知矩阵的行向量进行优化,也就是选择合适的行向量数量即用于重构系统方程的测量数据的个数。由于 DCS 监测数据极易受到噪声的影响,为了有效地降低部分受噪声影响程度较大的数据对重构系统方程造成的影响,测量数据的选取应遵循随机抽取、逐组检验的原则。针对同一段时间序列,由于存在噪声,选取不同的测量数据会导致重构的系数向量发生变化,所以根据系数的数值及分布无法确定系统方程的重构质量。根据上文可知, x^*, y^*, z^* 与 x, y, z 在数值上是极其接近的,认为系统方程的重构质量与两者的接近程度有关。因此,选择依据 x^*, y^*, z^* 的重构精度确定系统方程的重构质量, x^*, y^*, z^* 通过将时间序列 x, y, z 代入重构的系统方程中计算得到。

根据测量数据的选取原则可知,行向量的优化

涉及到两个参数的选择,包括测量数据的个数和测量数据的选取范围即窗口的宽度。为了完成行向量的优化,需要对以上两个参数多种组合下的矩阵结构作出评价。由于系统方程的重构质量越高,说明感知矩阵的结构设计越合理,因此将感知矩阵结构的评价方法定为:从不同长度梯度的时间序列中随机抽取若干个作为测量数据(即随机抽取过程),代入重构非线性系统方程的数学模型,得到重构的系统方程,将时间序列 x, y, z 代入重构的系统方程中得到 x^*, y^*, z^* ,然后计算时间序列 x, y, z 与 x^*, y^*, z^* 相对误差的绝对值的均值,并将其作为对应变量的重构误差。

为了减少随机误差的影响,将上述操作重复若干次完成评价的第一个阶段。接下来用滑动窗口历遍长度足够的整段正常数据,每次滑动后重复评价的第一个阶段。得到全部的重构误差后,设置关于重构误差的阈值,重构误差低于该阈值认为重构成功即系统方程的重构质量符合要求。以上就是逐组检验过程,最终得到不同矩阵结构的重构成功率即重构成功次数与重构总次数之比,重构成功率越高说明感知矩阵的结构设计越合理。另外,由于系统方程同时也是系统 B 中的虚拟传感器,重构成功说明当前虚拟传感器有效,重构成功率越高说明虚拟传感器的应用效果越好。

确定感知矩阵结构的评价方法之后,需要进行实例验证,对不同的矩阵结构作出评价,得到最优的矩阵结构。以 2018 年 3 月 1 日某煤化工企业压缩机组子系统(见图 7)中监测点位 A_API7602、A_API7612、A_API7614(详见表 1)数据为例。数据的采样间隔为 1 min,时间序列的长度为 1 440。该日 DCS 监控系统显示 3 个监测点位工作正常,未出现报警现象。

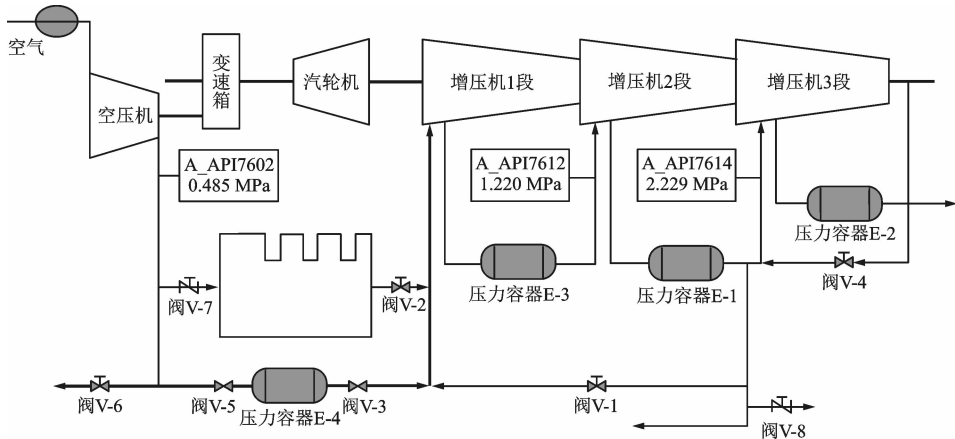
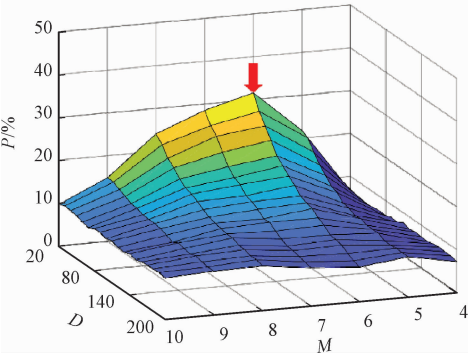


图 7 压缩机组设备连接简化图

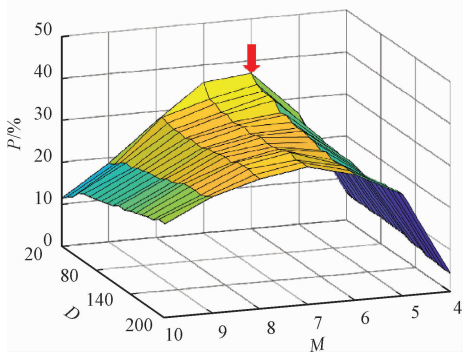
表 1 压缩机组系统监控变量表(部分)

监测点位	监测点位的物理含义
A_API7602	空压机排气压力
A_API7612	增压机 1 段排气压力
A_API7614	增压机 2 段排气压力

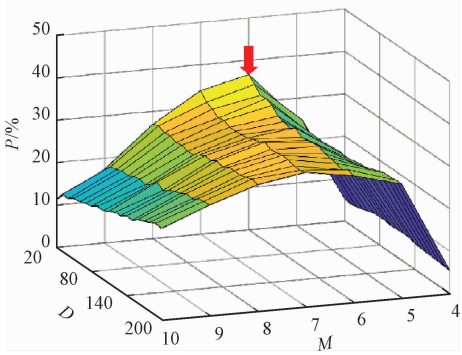
将以上 3 个点位的监测数据代入感知矩阵结构的评价方法中,由于重构成功率的求取涉及到大量的信号重构运算,为了提高运算效率,选择 SAMP 算法作为信号重构算法。另外,在系统方程的重构过程中发现,虽然 BP 算法理论上的重构精度要高于 SAMP 算法,但 SAMP 算法重构的系数向量的稀疏性要远好于 BP 算法,因此认为 SAMP 算法对 DCS 监测数据具有较强的适应性。将重构误差的阈值定为 0.01%,得到行向量优化后各监测变量的重构成功率 P 与 M 和 D 之间的关系如图 8 所示,其中 M 表示测量数据的个数, D 表示时间窗口的宽度,即一段连续时序数据的个数。图中窗口宽度的梯度设置为 10,由于当窗口宽度为 10 时,会造成可供选择的系统方程数量过少,因此设置窗口宽度的起始值为 20;同样,当测量数据的个数小于 4 时,重构成功率极低,所以设置其起始值为 4。从图中可以看出,当窗口宽度 D 为 20 且测量数据的个数 M



(a) A_API7602



(b) A_API7612



(c) A_API7614

图 8 重构成功率 P 与 M 和 D 之间的关系

为 6 时,所有的监测变量都能得到最高的重构成功率,如图中的箭头所示,此时就得到了最优的感知矩阵结构。

3.3 异常监测数据的检验与修复原理

将优化后的感知矩阵代入重构非线性系统方程的数学模型,根据采样点的不同,能够得到大量的系统方程,由于监测数据中普遍存在噪声,不是所有的系统方程都可以反映状态感知网络节点动态关系,需要对重构的系统方程进行检验,检验方法同重构成功率的检验方法:重构误差低于给定的阈值认为重构成功,即重构的系统方程符合要求。通过检验的所有系统方程都可以认为能够在一定程度上反映状态感知网络节点动态关系,这种动态关系不仅在滑动窗口内是稳定的,在接下来的一段时间内都应该是适用的,只不过随着时间推移,由于系统的混沌性,节点动态关系会发生改变,当前的系统方程不再适用。

接下来就可以通过重构的系统方程对异常监测信息进行检验与修复。以节点 x 为例,重构的系统方程具体形式应该为

$$x^* = a_0 + a_1 y + \cdots + a_7 x^2 y + \cdots + a_{16} y^3 + a_{17} z^3$$

(22)

由上文可知, x^* 可以用 x 近似替代。在使用式(22)对 x 进行求解时会出现一个严重的问题:由于存在 x 高次项, x 可能会有多个解析值,并且无法确认哪一个更接近真实值。为了解决这一问题,需要对系统方程进行优化,已知 x^* 与 x 不是完全一致的,它们之间应该存在较小的偏差,这个偏差是由系统 B 中虚拟传感器的固有特性引起的,将偏差 Δx 引入系统方程,得到

$$x + \Delta x = a_0 + a_1 y + \cdots + a_7 x^2 y + \cdots + a_{16} y^3 + a_{17} z^3$$

(23)

式中: Δx 是 x^* 与 x 之间的偏差, $x + \Delta x = x^*$ 。通过联立式(22)(23),将变量 y, z 下一时刻的监测数据代入,就能够得到 x 的解析值和对应的偏差 Δx ,当 x 出现多个解析值时,认为对应的偏差 Δx 小的解析值更接近真实值,这种规律也在仿真实验中得到了验证。通过对比 x 的解析值和 x 的实际值就能判断出监测数据是否出现异常,若与 x 的解析值相比 x 的实际值超差,认为监测数据出现异常,用 x 的解析值代替 x 的实际值,从而实现异常监测数据的检验与修复。异常监测数据的检验与修复算法流程如图 9 所示。

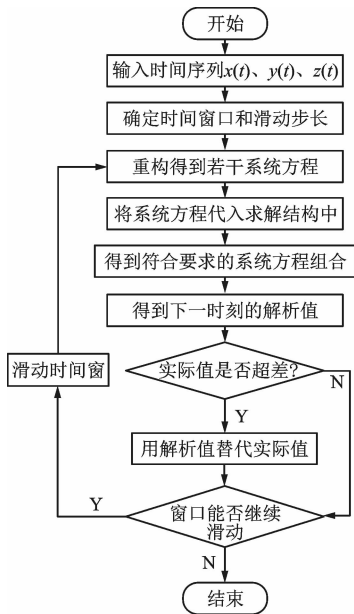


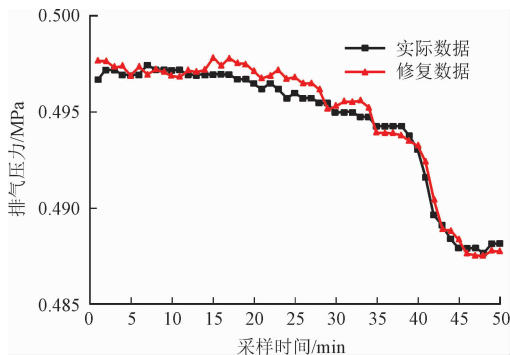
图 9 异常监测数据的检验与修复算法流程图

将 3 个监测点位 A_API7602、A_API7612、A_API7614 的监测数据代入本文算法以验证算法的有效性。图 10 是 3 个监测点位连续 50 个监测数据的修复结果,从图中可以看出,无论哪个监测变量作为待检验与修复变量,实际数据与修复数据之间的相对误差都稳定在极小值以内,在异常监测数据的检验与修复过程中解析值可以作为实际值的替代,证明本文算法有效。

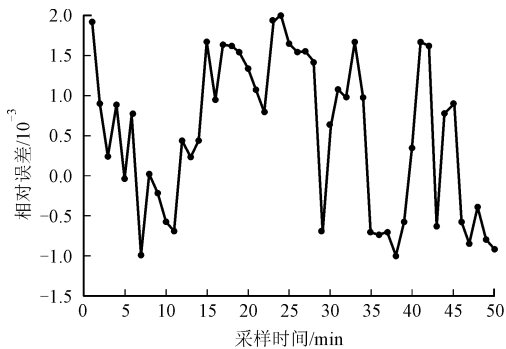
3.4 实例验证

2018 年 12 月 21 日下午 6 时,监测点位 A_API7612 出现持续性高报,时间约为 1 h。将出现异常的监测数据和其他监测点位正常的监测数据代入模型中,异常监测数据的检验与修复结果如图 11 所示。在监测数据正常的情况下,相对误差的波动极小;当监测数据出现异常时,相对误差会发生剧烈变化。从图 11 中可以看出,大约 60 个监测数据的

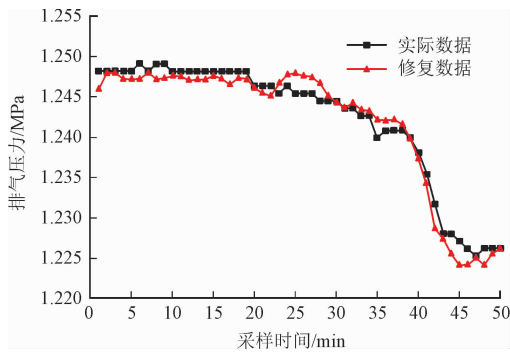
相对误差激增,说明算法检测到了异常监测数据。为了验证异常监测数据的修复效果,延长图 11a 中的实际数据得到图 11c,根据实际数据在之后时刻



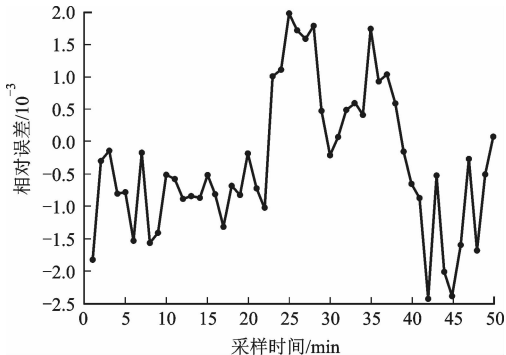
(a) A_API7602 的实际数据和修复数据



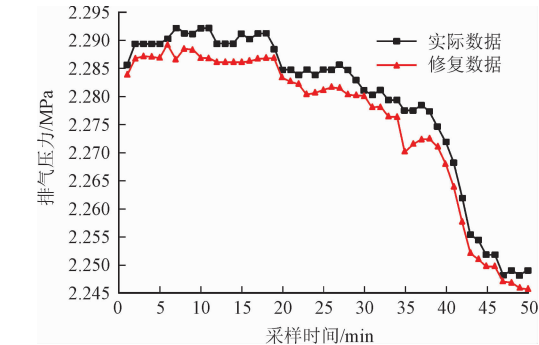
(b) A_API7602 的实际数据与修复数据之间的相对误差



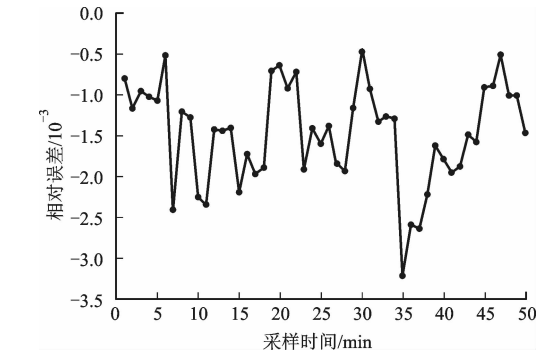
(c) A_API7612 的实际数据和修复数据



(d) A_API7612 的实际数据与修复数据之间的相对误差



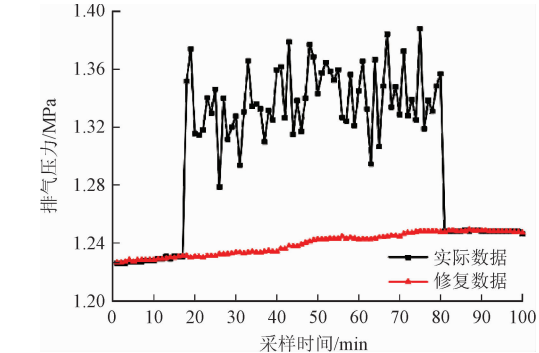
(e)A_API7614 的实际数据和修复数据



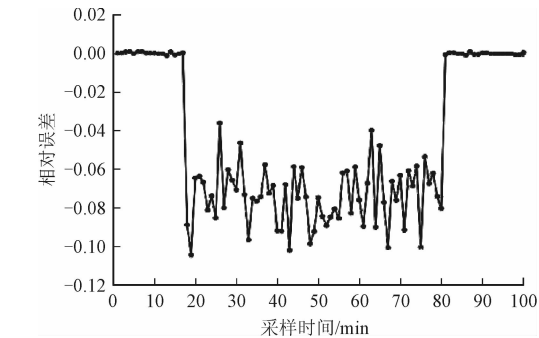
(f)A_API7614 的实际数据与修复数据之间的相对误差

图 10 3 个监测点位监测数据的修复结果

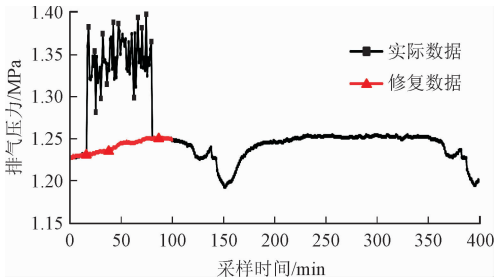
的变化趋势,认为修复结果能够较好地恢复监测数据
的原貌,证明本文算法能够实现流程工业异常监
测数据的检验与修复。



(a)A_API7612 的实际数据和修复数据



(b)A_API7612 的实际数据与修复数据之间的相对误差



(c)延长实际数据后 A_API7612 的实际数据和修复数据

图 11 异常监测数据的检验与修复结果

4 结 论

本文提出的采用压缩感知的流程工业异常监测
数据检验与修复方法,利用压缩感知原理构建重构
非线性系统方程的数学模型,仅使用少量的监测数
据就可以重构出反映状态感知网络节点动态关系的
系统方程,通过重构的系统方程得到监测数据的解
析值,从而实现异常监测数据的检验与修复。以某
煤化工企业压缩机子系统作为应用对象,结果表明
本方法能够为提高过程数据质量、保障生产系统安
全平稳运行提供技术支持。

本文所提方法的应用前提是监测数据的异常不
会引起闭环反馈系统的响应,并且忽略多个监测点
位同时出现异常的情形,接下来需要利用更多的工
业场景案例验证并优化模型,提高模型的适用性。

参考文献:

[1] 刘德荣. 复杂工业过程的先进控制 [J]. 自动化学报, 2014, 40(9): 1841-1842.
LIU Derong. The advanced control of complex industrial process [J]. Acta Automatica Sinica, 2014, 40 (9): 1841-1842.
[2] 赵皓, 高智勇, 高建民, 等. 一种采用相空间重构的多源数据融合方法 [J]. 西安交通大学学报, 2016, 50 (8): 84-89.
ZHAO Hao, GAO Zhiyong, GAO Jianmin, et al. A fusion method of multisource data using phase space reconstruction [J]. Journal of Xi'an Jiaotong University, 2016, 50(8): 84-89.
[3] PODOBNIK B, GROSSE I, HORVATÍĆ D, et al. Quantifying cross-correlations using local and global detrending approaches [J]. The European Physical Journal: B, 2009, 71(2): 243-250.
[4] ZHOU W X. Multifractal detrended cross-correlation analysis for two nonstationary signals [J]. Physical Review: E, 2008, 77(6): 066211.

- [5] HEDAYATIFAR L, VAHABI M, JAFARI G R. Coupling detrended fluctuation analysis for analyzing coupled nonstationary signals [J]. *Physical Review: E*, 2011, 84(2): 021138.
- [6] 蒋余厂, 刘爱伦. 基于 GLR-NT 的显著误差检测与数据协调 [J]. *华东理工大学学报(自然科学版)*, 2011, 37(4): 502-508.
JIANG Yuchang, LIU Ailun. Gross error detection and data reconciliation based on A GLR-NT combined method [J]. *Journal of East China University of Science and Technology (Natural Science Edition)*, 2011, 37(4): 502-508.
- [7] WANG F, JIA X P, ZHENG S Q, et al. An improved MT-NT method for gross error detection and data reconciliation [J]. *Computers & Chemical Engineering*, 2004, 28(11): 2189-2192.
- [8] SUN S C, DAO H, GONG Y X. A MT-NT-MILP combined method for gross error detection and data reconciliation [C]// *The 2nd International Conference on Information Science and Engineering*. Piscataway NJ, USA: IEEE, 2010: 3439-3442.
- [9] 何九虎, 刘飞. 工业过程数据异常检测的改进局部菌群因子法 [J]. *计算机与应用化学*, 2013, 30(1): 53-56.
HE Jiuhu, LIU Fei. The outlier detection of industry process using improved local outlier factor [J]. *Computers and Applied Chemistry*, 2013, 30(1): 53-56.
- [10] LIANG Y J, GAO Z Y, GAO J M, et al. Data fusion combined with echo state network for multivariate time series prediction in complex electromechanical system [J]. *Computational and Applied Mathematics*, 2018, 37(5): 5920-5934.
- [11] 葛新权. 回归模型应用的发展综述 [J]. *北京信息科技大学学报(自然科学版)*, 2009, 24(3): 1-6.
GE Xinquan. Applying development of regression model [J]. *Journal of Beijing Information Science & Technology University*, 2009, 24(3): 1-6.
- [12] 李坤, 马彩文, 李艳, 等. 压缩感知重构算法综述 [J]. *红外与激光工程*, 2013, 42(S1): 225-232.
LI Shen, MA Caiwen, LI Yan, et al. Survey on reconstruction algorithm based on compressive sensing [J]. *Infrared and Laser Engineering*, 2013, 42(S1): 225-232.
- [13] CANDES E J, TAO T. Near-optimal signal recovery from random projections; universal encoding strategies? [J]. *IEEE Transactions on Information Theory*, 2006, 52(12): 5406-5425.
- [14] CANDES E J, TAO T. Decoding by linear programming [J]. *IEEE Transactions on Information Theory*, 2005, 51(12): 4203-4215.
- [15] 闫敬文, 刘蕾, 屈小波. 压缩感知及应用 [M]. 北京: 国防工业出版社, 2015: 57-59.
- [16] NING W. Strongly NP-hard discrete gate-sizing problems [J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1994, 13(8): 1045-1051.
- [17] 李峰, 郭毅. 压缩感知浅析 [M]. 北京: 科学出版社, 2015: 61-79.
- [18] DONOHO D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1306.
- [19] CHEN S, SDONHO D, LSAUNDERS M A. Atomic decomposition by basis pursuit [J]. *SIAM Review*, 2001, 43(1): 129-159.
- [20] NEEDELL D, TROPP J A. CoSaMP: iterative signal recovery from incomplete and inaccurate samples [J]. *Communications of the ACM*, 2010, 53(12): 93-100.
- [21] DAI Wei, MILENKONIC O. Subspace pursuit for compressive sensing signal reconstruction [J]. *IEEE Transactions on Information Theory*, 2009, 55(5): 2230-2249.
- [22] DO T T, GAN L, NGUYEN N, et al. Sparsity adaptive matching pursuit algorithm for practical compressed sensing [C]// *2008 42nd Asilomar Conference on Signals, Systems and Computers*. Piscataway, NJ, USA: IEEE, 2008: 581-587.
- [23] JI Shihao, XUE Ya, CARIN L. Bayesian compressive sensing [J]. *IEEE Transactions on Signal Processing*, 2008, 56(6): 2346-2356.
- [24] WANG Wenxu, YANG Rui, LAI Yingcheng, et al. Predicting catastrophes in nonlinear dynamical systems by compressive sensing [J]. *Physical Review Letters*, 2011, 106(15): 154101.
- [25] YANG R, LAI Y C, GREBOGI C. Forecasting the future: is it possible for adiabatically time-varying nonlinear dynamical systems? [J]. *Chaos: an Interdisciplinary Journal of Nonlinear Science*, 2012, 22(3): 033119.

(编辑 杜秀杰)