文章编号:1674-2974(2024)10-0174-07 DOI:10.16339/j.cnki.hdxbzkb.2024210

基于改进SVT的电力负荷数据恢复算法

成达[1,2],熊素琴[1],马力[2],唐求[2],闫森[2†]
(1. 中国电力科学研究院有限公司,北京 100192;
2. 湖南大学 电气与信息工程学院,湖南 长沙 410082)

摘 要:针对传统奇异值阈值(Singular Value Thresholding,SVT)数据恢复算法在对电力负荷数据恢复中忽视数据先验信息以及大规模数据计算效率低等问题,提出一种基于相空间重构与自适应变步长的改进SVT的数据恢复算法.为解决传统SVT容易忽视数据先验信息的问题,引入相空间重构算法将原始缺失数据映射到高维空间,利用数据间的关联性和结构特征,为后续数据恢复算法提供先验知识;结合对数与Sigmoid函数构建变步长基础函数,并利用等比项提高前期步长,构建自适应变步长SVT算法,克服传统SVT在大规模数据情况下计算效率低的问题.结合多项公用电力负荷数据集及多种常用电力负荷数据恢复算法进行对比实验分析,结果表明,改进SVT算法可获得更好的数据恢复效果,收敛速度、精度以及稳定性得到提升,具有较强的工程实用性.

关键词:电力负荷数据;数据处理;奇异值阈值;相空间方法;自适应变步长
中图分类号:TP391 **文献标志码**:A

An Enhanced Algorithm for Power Load Data Recovery Based on Improved SVT

CHENG Da[1,2],XIONG Suqin[1],MA Li[2],TANG Qiu[2],YAN Sen[2†]
(1. China Electric Power Research Institute Co., Ltd., Beijing 100192, China;
2. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: In response to the issues of neglecting prior information and low computational efficiency of the traditional Singular Value Thresholding (SVT) algorithm in power load data recovery, a novel improved SVT algorithm based on phase space reconstruction and adaptive variable step length is proposed. To address the problem of neglecting prior information in traditional SVT, a phase space reconstruction algorithm is introduced to map the original missing data into a high-dimensional space, leveraging data correlation and structural features as prior knowledge for subsequent data recovery algorithms. By combining logarithmic and sigmoid functions to construct the variable step length base function, and utilizing geometric progression to enhance the initial step length, an adaptive variable step length SVT algorithm is built to overcome the low computational efficiency issue of traditional SVT in large-scale data scenarios. Comparative experimental analysis is conducted using multiple publicly available power

* **收稿日期**:2023-12-13
基金项目:国家电网有限公司总部管理科技项目资助:基于端云协同的计量防护单元数据感知和精益化运检技术研究(5700-202255203A-1-1-ZN),State Grid Corporation of China Headquarters Management Science and Technology Project Funding: Research on Metering Protection Unit Data Sensing and Lean Operation and Maintenance Technology Based on Edge-Cloud Collaboration (5700-202255203A-1-1-ZN)
作者简介:成达(1986—),男,湖南衡阳人,湖南大学博士,中国电力科学研究院有限公司高级工程师,英国IET皇家特许工程师
† 通信联系人,E-mail: 2960068599@qq.com

load datasets and various commonly used power load data recovery algorithms. The results demonstrate that the improved SVT algorithm achieves better data recovery performance, with enhanced convergence speed, accuracy, and stability, showcasing strong engineering practicality.

Key words: electricity load data; data processing; singular value thresholding; phase space methods; adaptive variable step size

随着智能电网建设的推进,大规模监测设备纳入电力系统,使得电力数据呈指数式增长[1]. 电力负荷数据在优化能源利用、设备运行分析等方面发挥重要作用. 由于设备故障、通信问题等,电力负荷数据易出现缺失问题,影响相关电力决策制定[2]. 为此,需研究数据恢复算法提升数据质量,为电力负荷预测、电网优化配置等领域提供技术支撑.

目前,针对电力数据恢复国内外主要关注三个方面的研究[3]:统计学方法,机器学习方法,矩阵补全方法. 统计学方法基于现有电力数据特征,利用统计学原理对缺失数据进行插补和填充. 常见的统计学方法有多项式插值法[4]和三次样条插值法[5]等. 这两种方法应用广泛,计算复杂度低,但对数据点的选择和分布要求较高,分别存在过拟合和插值不平滑的缺陷,且随着缺失数据量增多,易忽视电力数据的时空特性,从而导致数据恢复精度不高.

机器学习方法基于历史数据的趋势,构建预测模型实现缺失样本预测并补全. 随机森林填补法[6]是一种兼具灵活性和鲁棒性的方法,但在处理长期趋势和周期性变化的数据时有一定的局限性. 此外,去噪自编码器[7]、改进生成式对抗网络[8]和改进轻量级梯度提升机[9]等机器学习方法也被用于缺失数据重建中. 上述机器学习方法考虑电力数据的时空特性,提高了数据恢复精度,但当历史数据不足时,可能会影响补全效果,且计算复杂度高.

近年来,矩阵补全方法在图像处理、视频修复等领域广泛应用[10],矩阵补全方法采用矩阵分解和重构,并确定优化目标函数得到最佳的结果. 许多学者将矩阵补全方法引入电力数据恢复中,以解决算法速度和精度的平衡问题. Zhuang 等[11]提出自适应奇异值分解线性插值数据补全方法. Donti 等[12]针对配电系统中低可观测性问题,提出低秩矩阵补全的电压估计方法,能准确预估电压相量,并具有良好的鲁棒性. Ghasemkhani 等[13]提出正则化低秩张量补全方法,与传统方法相比,所提方法有更好的插补精度. 王毅等[14]建立基于低秩矩阵分解的母线异常数据辨识与修复模型,提高数据预测精度. 上述矩阵补全方法所需计算资源远少于机器学习方法,通用性较好.

目前,研究表明,电力负荷数据、低压台区用电数据等电力数据都具有周期性、相关性、随机性等多种特征,均表现出低秩性[3]. 奇异值分解(Singular Value Decomposition, SVD)作为一种优秀的矩阵补全算法,被许多学者用于数据处理方面. SVT 是基于 SVD 提出的数据恢复算法,可用于实现电力数据恢复,并有更好性能:收敛速度快、精度高[10]. 尽管 SVT 算法在数据补全方面取得了众多进展,但仍有不足之处:忽视数据固有的先验信息,大规模数据计算效率低[15].

为此,本研究针对海量电力数据中含有缺失、异常值等问题,开展基于改进 SVT 的电力负荷数据恢复算法的研究,以实现精度与速度的最优平衡.

1　相空间重构

1.1　相空间重构过程

相空间重构是一种基于时间序列数据的分析方法,主要用于从一维时间序列数据中提取系统的多维动态行为. 基于 Takens 定理,通过将一维时间序列嵌入到多维相空间中,恢复原始系统信息.

Packard 等[16]在 1980 年提出两种时间序列重构相空间的方法:导数重构法和坐标延迟重构法. 然而在实际应用过程中,导数重构法存在坐标间尺度差异过大、对噪声的敏感性导致计算误差太大,因此,通常采用坐标延迟重构法进行时序数据的相空间重构[17]. 式(1)为任意某个序列总长度为 N 的单变量时间序列 $\{x(i), i = 1, 2, \cdots, N\}$ 经过坐标延迟重构法重构相空间后的转换结果.

$$\boldsymbol{X} = \begin{bmatrix} x(1) & x(2) & \cdots & x(d) \\ x(1+t) & x(2+t) & \cdots & x(d+t) \\ \vdots & \vdots & & \vdots \\ x[1+(m-1)t] & x[2+(m-1)t] & \cdots & x[d+(m-1)t] \end{bmatrix} \tag{1}$$

式中:t为延迟时间;m为嵌入维数,这里为矩阵的行数;d为相空间重构矩阵的列数.如果延迟时间过短,相空间向量中的两个坐标分量之间的数值非常接近,无法相互区分,将导致无法提供两个独立的坐标分量.相反,如果延迟时间过长,两个坐标分量完全独立,将导致重构矩阵中相邻两行的数据毫无相关性[18].嵌入维数m的确定直接影响后续的分析和建模,m过大,将放大数据的噪声,且导致模型的计算量增大;m过小,将降低时序数据处理与分析的准确度[19].因此,相空间重构的关键在于嵌入维数m和延迟时间t的选择.

1.2 延迟时间的确定

在相空间重构和时序数据分析过程中,延迟时间的选择很重要,需采用适当方法确定合适的延迟时间,在独立性和相关性间取得平衡.自相关函数法(Autocorrelation Function, ACF)和互信息法(Mutual Information, MI)是目前被广泛使用的方法,其中,ACF通过观察延迟时间对自相关函数的影响,进而找到一个合适的延迟时间,使得在相空间重构中能够捕捉到系统的重要动力学特征,算法简便有效,但只能提取时序数据的线性相关性[20].

相较于ACF,MI具有更强的非线性表达力,可以衡量两个变量间的相关性和依赖关系,从而更好地平衡相空间内部的关系[21].考虑到电力时序数据之间强非线性,本文使用MI来确定相空间重构的最佳延迟时间.首先,计算互信息,得到延迟因子,以便后续的数据分析和处理.MI计算式为:

$$H(x,y)=\sum_{y}\sum_{x}p(x,y)\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{2}$$

式中:x为单变量时间序列;y为x延迟t后得到的新序列;$p(x)$和$p(y)$分别为变量x和y的边际概率分布;$p(x,y)$为变量x和变量y的联合概率分布.

取不同的t,依次计算系统变量间的互信息为:

$$I(t)=H(x)+H(x_t)-H(x,x_t) \tag{3}$$

式中:$H(\cdot)$表示系统的信息熵.

绘制互信息函数$I(t)$随t的变化图,并取第一次至极小点的t值确定相空间重构的延迟时间.

1.3 嵌入维数的确定

在嵌入维数m的选取过程中,要充分考虑模型的准确性、噪声影响和计算效率等因素.常见方法主要有基于混沌评判指标的方法和基于相空间重构效果的方法.前者通过计算不同m下的混沌评判指标,评估不同m下的结果是否收敛,选取稳定状态下的维数作为嵌入维数;后者针对不同的m开展实验,通过观察重构结果,确定最终嵌入维数[20].

本文基于相空间重构效果确定嵌入维数,通过对比不同嵌入维数下的模型效果选择最佳嵌入维数,流程如图1所示.首先输入待恢复数据,然后设置t与待确定嵌入维数,之后进行不同m下的模型训练与评估,比较模型效果,确定最佳效果的嵌入维数.

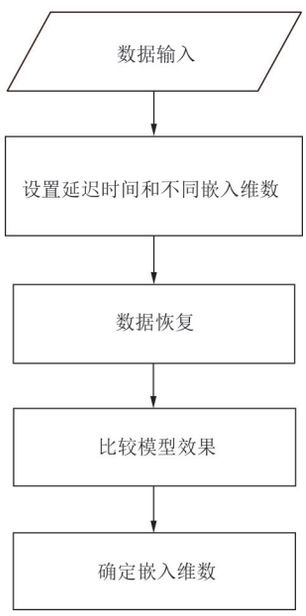


图1 相空间重构嵌入维数m选择流程

Fig.1 Phase space reconstruction m selection process

相空间重构将含缺失值的海量电力时序数据映射到高维空间,充分利用数据间的关联性和结构特征,为后续数据恢复算法的运行提供先验知识.

2 基于改进SVT的数据恢复算法

与传统的数据恢复算法相比,SVD在数据恢复过程中表现出明显的线性无偏性、最佳逼近性、自适应性与鲁棒性等优势[19].在此基础上,SVT目前正被广泛应用于各种领域的数据恢复任务.

2.1 基于SVT的数据恢复算法

与SVD不同,SVT在对矩阵进行奇异值分解后通过奇异值进行软阈值处理,将矩阵的秩降低到数据所需的级别,并重新构造矩阵.其主要步骤如下.

1)输入数据准备.对于由含缺失值数据的单变量序列进行相空间重构后的矩阵,将其表示为一个已知部分和一个未知部分之和.即

$$\boldsymbol{M}_{\text{input}}=\boldsymbol{M}_{\text{known}}+\boldsymbol{M}_{\text{unknown}} \tag{4}$$

式中:$\boldsymbol{M}_{\text{input}}$为进行奇异值分解前的输入矩阵;$\boldsymbol{M}_{\text{known}}$和$\boldsymbol{M}_{\text{unknown}}$分别为输入矩阵中已知部分的矩阵和未知部分的矩阵.

2)矩阵初始化.对于未知部分的矩阵$\boldsymbol{M}_{\text{unknown}}$,本文进行零值填充处理,并且使用零矩阵初始化输入矩阵$\boldsymbol{M}_{\text{input}}$的类似矩阵$\boldsymbol{M}_{\text{init}}$.

3)迭代更新过程.通过迭代过程逐步改善未知部分的矩阵的估计,主要过程如下.

步骤1:奇异值分解.对矩阵$\boldsymbol{M}_{\text{init}}$进行奇异值分解,得到$\boldsymbol{M}_{\text{init}}$的奇异值和奇异向量,分解式为:

$$\boldsymbol{M}_{\text{init}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \left[u_1, u_2, ..., u_m\right]\begin{bmatrix}\sigma_1 & 0 & \cdots & 0 \cdots & 0\\ 0 & \sigma_2 & \cdots & 0 \cdots & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \sigma_m \cdots & 0\end{bmatrix}\times \left[v_1, v_2, ..., v_m, ..., v_n\right]^{\mathrm{T}} \tag{5}$$

式中:$\boldsymbol{U}$、$\boldsymbol{V}$为正交矩阵;$\boldsymbol{\Sigma}$为对角矩阵;对角线上的元素为奇异值;m、n分别为矩阵的行数和列数.

步骤2:奇异值的软阈值处理.将对角矩阵中的每个元素都减去一个阈值,并将小于零的元素置零,即去除小于阈值的奇异值,并保留大于等于阈值的奇异值,该过程可以表示为:

$$\sigma_i' = \begin{cases}\sigma_i - \tau, & \sigma_i > \tau\\ 0, & \sigma_i \leqslant \tau\end{cases} \tag{6}$$

式中:σ_i'为处理后的第i个奇异值;σ_i为原始的第i个奇异值;τ为设定的奇异值阈值.

步骤3:矩阵重构.使用经阈值处理后的奇异值和奇异向量,重构估计矩阵M^*,重构过程为:

$$\boldsymbol{M}^* = \boldsymbol{U}\boldsymbol{\Sigma}'\boldsymbol{V}^{\mathrm{T}} = \left[u_1, u_2, ..., u_m\right]\begin{bmatrix}\sigma_1' & 0 & \cdots & 0 \cdots & 0\\ 0 & \sigma_2' & \cdots & 0 \cdots & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \sigma_m' \cdots & 0\end{bmatrix}\times \left[v_1, v_2, ..., v_m, ..., v_n\right]^{\mathrm{T}} \tag{7}$$

步骤4:更新.通过给定的更新步长,结合输入矩阵$\boldsymbol{M}_{\text{input}}$和估计矩阵$\boldsymbol{M}^*$的差值,对奇异值分解矩阵进行更新,更新过程为:

$$\boldsymbol{M}_{\text{init}} = \boldsymbol{M}_{\text{init}} + \Delta\boldsymbol{W}(\boldsymbol{M}_{\text{input}} - \boldsymbol{M}^*) \tag{8}$$

式中:$\boldsymbol{M}_{\text{init}}$为初始化全零矩阵;$\Delta$为给定步长值;$\boldsymbol{W}$为原始输入矩阵的掩码矩阵.

步骤5:终止条件检查.通过判断算法是否达到设定的最大迭代次数或矩阵的重构误差是否达到设定阈值来检查迭代是否达到终止条件,如果终止条件满足,算法结束;否则,返回步骤3.以矩阵的二范数式表示重构误差.

$$\varepsilon = \frac{\left\| \boldsymbol{M}_{\text{input}} - \boldsymbol{M}^* \right\|_2}{\left\| \boldsymbol{W}\boldsymbol{M}_{\text{input}} \right\|_2} \tag{9}$$

4)输出结果.在算法收敛后,得到最终的估计矩阵$\boldsymbol{M}^*$作为原始矩阵的恢复结果.然而,考虑到原始样本为含缺失值的单变量时间序列数据,因此,在最终数据缺失值的恢复上,需要对恢复矩阵进行重构使其变为一维时序数据.具体为:

$$l_{\min} = \begin{cases}0, & q \leqslant K\\ \left\lfloor \dfrac{q - K - 1}{t} \right\rfloor, & q > K\end{cases} \tag{10}$$

$$l_{\max} = \begin{cases}\left\lfloor \dfrac{q - 1}{t} \right\rfloor, & q \leqslant N - K\\ L - 1, & q > N - K\end{cases} \tag{11}$$

$$X_q = \frac{1}{l_{\max} - l_{\min} + 1}\sum_{i = l_{\min}}^{l_{\max}} \boldsymbol{M}^*_{i, q - i*t - 1} \tag{12}$$

式中:X_q为重构后原始单变量样本对应的第q个数;N为原始单变量样本的数据总量;L为$\boldsymbol{M}_{\text{input}}$的行数;$K$为$\boldsymbol{M}_{\text{input}}$的列数;$l_{\min}$和$l_{\max}$分别为$X_q$对应$\boldsymbol{M}^*$矩阵中相关数据所存在的最小行与最大行,通过计算得出X_q大小.

2.2　改进的SVT数据恢复算法

SVT数据恢复算法的收敛性和速度受步长选择影响,步长过大会导致算法发散,无法收敛到最优解,步长过小则使收敛速度变慢,增加计算成本.针对上述问题,本文提出自适应变步长SVT算法,动态调整步长大小,确保算法收敛性和速度的最佳平衡.

变步长自适应算法通常根据当前误差调整步长,常见调节函数包括双曲正切和Sigmoid函数.本文提出改进的Sigmoid函数变步长方法,结合等比项,先基于对数和Sigmoid函数构建变步长基础函数,再引入等比项以加速前期收敛.具体为:

$$\mu(n) = \beta\{\ln[1 + \mathrm{e}^{-\alpha(|e(n)|)^{-2}}]\} \tag{13}$$

$$\gamma(n) = q\gamma(n - 1), 0 < q < 1 \tag{14}$$

$$\Delta(n + 1) = (\mu(n) + \gamma(n)k(|e(n)| - |e(n - 1)|))d \tag{15}$$

式中:$\mu(n)$为步长因子调节函数;β为垂直尺度因子;α为水平尺度因子,通过调节β和α可调节函数幅值与形状;$e(n)$为第n次迭代的误差大小;$\gamma(n)$为等比项,由于$0 < q < 1$,该项可提高前期收敛速度而不影响后期效果.式(15)为步长更新方法,d为矩阵元素总数(含缺失)与不含缺失数据量比;k被用于调控误差差值大小.

改进后,原本基于SVT的数据恢复算法的步骤4,变为利用式(15)计算出的更新步长,结合输入矩阵$\boldsymbol{M}_{\text{input}}$和估计矩阵$\boldsymbol{M}^*$的差值,对奇异值分解矩阵进

行更新,更新过程为:

$$\boldsymbol{M}_{\text{init}} = \boldsymbol{M}_{\text{init}} + \Delta(n)W(\boldsymbol{M}_{\text{input}} - \boldsymbol{M}^{*}) \tag{16}$$

改进后的数据恢复流程如图2所示.第1步对数据进行预处理,第2步对数据进行相空间重构并提取隐藏信息,第3步进行SVT数据恢复,第4步判断是否达到指定精度或迭代次数最大指定值,若没有则更新步长与分解矩阵,返回第3步,若达到则进行第5步时间序列数据恢复.

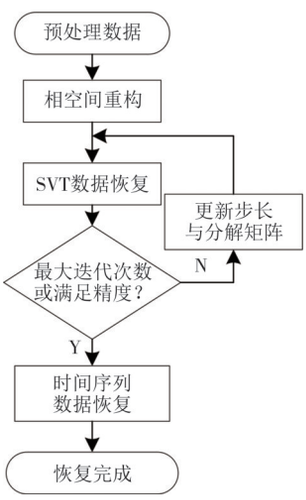


图2 数据恢复流程图

Fig.2 Flowchart for data recovery

3 实验分析

实验采用Python语言,在PyCharm平台进行.通过比利时Elia电网总负载数据集(以下简称Elia数据集)和美国每日负荷公开数据集(以下简称美国数据集),针对不同缺失度验证本文算法的恢复精度和有效性.对比方法包括多项式插值法、三次样条插值法、随机森林填补法及传统SVT算法.

3.1 相空间重构参数选取实验

经MI选取合适的延迟时间后,对嵌入维数进行实际数据分析,选取均方根误差(Root Mean Square Error,RMSE)和决定系数R^2作为模型评价指标.

RMSE是恢复结果与原始数据之间误差的平方和均值的平方根,能够反映整体准确性.RMSE值越小表示模型的恢复能力越好,其计算式为:

$$R_{\text{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{17}$$

式中:R_{RMSE}为均方根误差;n为样本数据的个数;y_i表示真实值;$\hat{y}_i$表示填补值.

决定系数R^2是一种常用的统计指标,取值范围为0~1.它可以直观地表示模型对数据的拟合程度,R^2越接近1表示模型能够很好地解释观测值的方差.其计算式为:

$$R^2 = 1 - \sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \Big/ \sum_{i=1}^{n}\left(\bar{y} - y_i\right)^2 \tag{18}$$

式中:$\bar{y}$表示样本均值.

本文通过相空间重构效果确定嵌入维数,比较不同维度下模型效果,选择最佳维数.图3和图4展示了数据随机缺失30%时,不同嵌入维数m下R_{RMSE}和R^2变化.结果显示,嵌入维数增加提升了模型性能,但过高的嵌入维数增加了计算量,并可能放大噪声,如Elia数据集中评价指标产生了浮动.为平衡计算量与数据恢复效果,本研究选取m = 130作为Elia数据集的嵌入维数,m = 100作为美国数据集的嵌入维数.

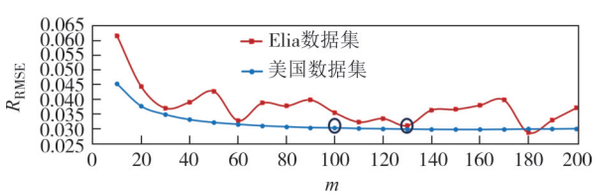


图3 不同m下R_{RMSE}变化趋势

Fig.3 Variations of R_{RMSE} with different m

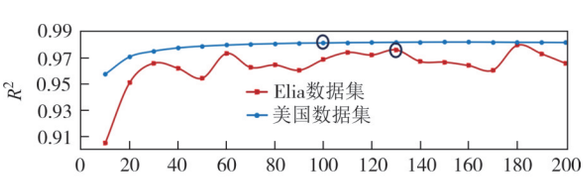


图4 不同m下R^2变化趋势

Fig.4 Variations of R^2 with different m

3.2 改进SVT数据恢复算法的效果验证

将Elia数据集和美国数据集随机缺失30%用于验证改进SVT数据恢复算法的数据恢复效果,结果分别如图5和图6所示,由图可知,均显示出良好效果.图7和图8对比了改进前后SVT算法的收敛效果,结果表明,改进SVT算法在两个数据集上通过自适应变步长机制,显著提升迭代收敛速度、精度和稳定性.

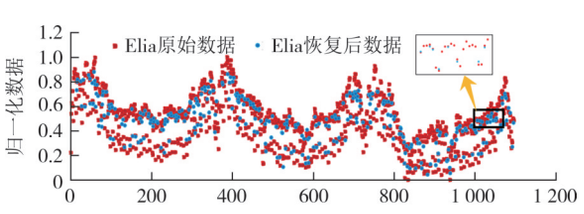


图5 Elia数据集数据恢复效果

Fig.5 Data recovery effectiveness of Elia data set

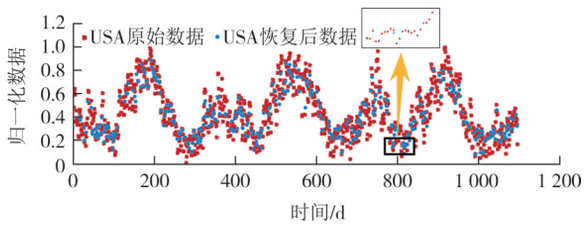


图6　美国数据集数据恢复效果

Fig.6　Data recovery effectiveness of USA data set

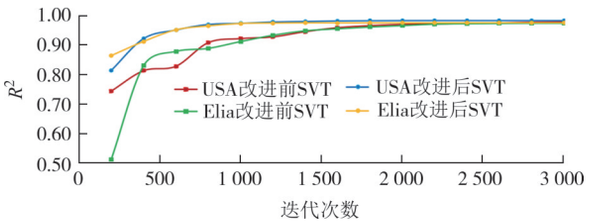


图8　不同迭代次数下R^2变化趋势

Fig.8　Variations of R^2 with different iteration counts

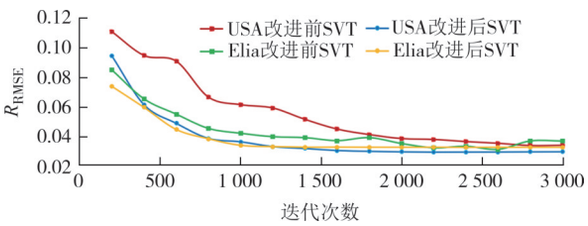


图7　不同迭代次数下R_{RMSE}变化趋势

Fig.7　Variations of R_{RMSE} with different iteration counts

3.3　对比实验

在Elia数据集和美国数据集进行实验时,完全随机地使数据缺失度从10%变化到80%,增幅为10%. 同时,采取在电力负荷数据恢复领域应用广泛的多项式插值法、三次样条插值法、随机森林填补法、传统SVT算法为对比对象,比较改进SVT算法与其他算法的恢复效果差异,结果分别如表1和表2所示. 由表1和表2可知,多项式插值法与三次样条插值法的数据恢复整体精度较差;随机森林填补法缺失度较低的时候效果尚可,缺失度超过30%时恢复效果急速下降;传统SVT算法恢复效果较好,但在恢复准确度与拟合效果精度均不如改进SVT算法;改进SVT算法在数据缺失度从10%增加到30%时,R_{RMSE}分别为0.014 8、0.027 1与0.032 2(Elia数据集)远超过其他几种算法,缺失度升高后,恢复准确度与拟合效果虽然略有下降,但依然优于其他算法,且在两个数据集的恢复效果都很好,具有良好泛用性.

表1　不同数据缺失度下5种方法性能对比(Elia数据集)

Tab.1　Performance comparison of five methods under different levels of data missingness (Elia data set)

缺失度/%	多项式插值法		三次样条插值法		随机森林填补法		传统SVT算法		改进SVT算法	
	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2
10	0.044 6	0.950 1	0.031 3	0.975 4	0.031 8	0.974 6	0.034 5	0.970 0	**0.014 8**	**0.994 5**
20	0.076 5	0.853 4	0.053 9	0.927 3	0.050 3	0.936 6	0.037 8	0.964 1	**0.027 1**	**0.981 5**
30	0.099 6	0.751 4	0.074 4	0.861 2	0.066 7	0.888 5	0.039 3	0.961 3	**0.032 2**	**0.974 0**
40	0.122 7	0.622 7	0.104 3	0.727 6	0.083 0	0.827 4	0.049 0	0.939 7	**0.045 7**	**0.942 5**
50	0.135 4	0.541 2	0.123 4	0.618 6	0.097 3	0.763 0	0.054 4	0.925 7	**0.051 2**	**0.930 2**
60	0.143 0	0.488 0	0.148 8	0.445 8	0.114 0	0.674 6	0.059 3	0.911 9	**0.056 3**	**0.915 4**
70	0.145 7	0.468 7	0.178 8	0.199 7	0.126 8	0.597 0	0.076 9	0.851 7	**0.073 2**	**0.861 2**
80	0.146 6	0.466 5	0.189 3	0.103 1	0.136 4	0.533 7	0.091 5	0.790 2	**0.084 1**	**0.819 8**

表2　不同数据缺失度下5种方法性能对比(美国数据集)

Tab.2　Performance comparison of five methods under different levels of data missingness (USA data set)

缺失度/%	多项式插值法		三次样条插值法		随机森林填补法		传统SVT算法		改进SVT算法	
	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2	R_{RMSE}	R^2
10	0.064 5	0.912 8	0.016 9	0.993 9	0.021 1	0.990 6	0.027 7	0.983 8	**0.018 1**	**0.993 0**
20	0.110 1	0.746 2	0.030 7	0.980 2	0.030 2	0.980 8	0.031 2	0.979 5	**0.023 3**	**0.988 6**
30	0.147 4	0.545 5	0.049 5	0.948 7	0.048 9	0.949 9	0.035 2	0.974 0	**0.030 1**	**0.981 0**
40	0.174 3	0.364 6	0.065 2	0.910 8	0.056 1	0.934 1	0.043 1	0.961 0	**0.041 5**	**0.969 2**
50	0.192 3	0.226 3	0.082 3	0.866 1	0.063 8	0.914 6	0.045 4	0.956 8	**0.044 0**	**0.959 3**
60	0.202 6	0.141 3	0.095 2	0.810 2	0.083 3	0.854 7	0.057 6	0.930 5	**0.055 3**	**0.934 8**
70	0.206 2	0.110 1	0.108 1	0.755 4	0.096 6	0.804 7	0.076 1	0.878 8	**0.074 6**	**0.883 4**
80	0.207 0	0.103 8	0.127 1	0.661 8	0.101 4	0.784 6	0.092 3	0.821 5	**0.085 2**	**0.847 9**

4 结 论

本文提出了一种基于改进SVT的数据恢复算法用于电力负荷数据恢复.利用相空间重构获取电力负荷时间序列数据的隐藏信息和先验知识,深入挖掘时间序列更多的信息;基于SVT算法收敛速度慢、收敛精度不足的缺陷,利用对数与Sigmoid函数进行变步长基础函数构建,引入等比项以获取更快前期收敛速度,基于此构建自适应变步长SVT算法,提升算法的收敛速度、收敛精度与稳定性.通过在不同数据集上与其他常用的数据恢复算法进行对比分析,结果表明,本文提出算法收敛速度更快,精度更高且具有良好的普适性.

参考文献

[1] 张东霞,苗新,刘丽平,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-12.
ZHANG D X,MIAO X,LIU L P,et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE,2015,35(1):2-12.(in Chinese)

[2] 李富柏,焦瑞莉,薄宇,等.基于DTWKNN的电力缺失数据补全方法[J].北京信息科技大学学报(自然科学版),2023,38(5):32-38.
LI F B,JIAO R L,BO Y,et al. Power missing data completion method based on DTWKNN[J]. Journal of Beijing Information Science & Technology University,2023,38(5):32-38.(in Chinese)

[3] 洪德华,张翠翠,宫政,等. 基于改进双向GABP神经网络的电力负荷缺失数据补全方法[J]. 电气自动化,2022,44(5):41-45.
HONG D H,ZHANG C C,GONG Z,et al. Method of completing power load missing data basedon improved bidirectional GABP neural network[J]. Electrical Automation,2022,44(5):41-45.(in Chinese)

[4] 陆嘉铭,奚增辉,瞿海妮,等. 电力量测数据缺失补齐方法研究与实践[J]. 电力大数据,2023,26(7):40-49.
LU J M,XI Z H,QU H N,et al. Research and practice on power measurement data missing value imputation methods[J]. Power Systems and Big Data,2023,26(7):40-49.(in Chinese)

[5] 刘灏,杨智伟,毕天姝,等. 基于优先级分配策略的PMU数据恢复方法[J]. 电网技术,2018,42(9):2814-2820.
LIU H,YANG Z W,BI T S,et al. PMU data recovery method based on priority allocation strategy[J]. Power System Technology,2018,42(9):2814-2820.(in Chinese)

[6] 马颢. 针对多源数据的变压器状态评估方法研究[D]. 济南:山东大学,2023.
MA H. Research on transformer condition evaluation method for multi-source data[D]. Jinan:Shandong University,2023.(in Chinese)

[7] RYU S,KIM M,KIM H. Denoising autoencoder-based missing value imputation for smart meters[J]. IEEE Access,2020,8:40656-40666.

[8] 王守相,陈海文,潘志新,等. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法[J]. 中国电机工程学报,2019,39(1):56-64.
WANG S X,CHEN H W,PAN Z X,et al. A reconstruction method for missing data in power system measurement using an improved generative adversarial network[J]. Proceedings of the CSEE,2019,39(1):56-64.(in Chinese)

[9] 李富盛,陈伟松,钱斌,等.面向低压配电网智能电表误差监测的LightGBM-EM-EC多变量缺失数据高效重建[J].中国电机工程学报,2022,42(增刊1):95-105.
LI F S,CHENG W S,QIAN B,et al. Efficient reconstruction of multivariate missing data for smart meter error monitoring in low voltage distribution network based on LightGBM-EM-EC[J]. Proceedings of the CSEE,2022,42(Sup.1):95-105.(in Chinese)

[10] 洪文慧,李钦豪,张勇军,等.基于二次矩阵补全的低压配电网相序识别算法[J].电力自动化设备,2022,42(9):133-138.
HONG W H,LI Q H,ZHANG Y J,et al. Quadratic matrix completion based phase sequence identification algorithm for low-voltage distribution network[J]. Electric Power Automation Equipment,2022,42(9):133-138.(in Chinese)

[11] ZHUANG C J,AN J W,LIU Z Q,et al. Data completion for power load analysis considering the low-rank property[J]. CSEE Journal of Power and Energy Systems,2022,8(6):1751-1759.

[12] DONTI P L,LIU Y J,SCHMITT A J,et al. Matrix completion for low-observability voltage estimation[J]. IEEE Transactions on Smart Grid,2020,11(3):2520-2530.

[13] GHASEMKHANI A,NIAZAZARI I,LIU Y C,et al. A regularized tensor completion approach for PMU data recovery[J]. IEEE Transactions on Smart Grid,2021,12(2):1519-1528.

[14] 王毅,李鼎睿,康重庆. 低秩矩阵分解在母线坏数据辨识与修复中的应用[J]. 电网技术,2017,41(6):1972-1979.
WANG Y,LI D R,KANG C Q. Application of low-rank matrix factorization in bad data identification and recovering for bus load[J]. Power System Technology,2017,41(6):1972-1979.(in Chinese)

[15] 陈雷,陈松灿. 矩阵补全模型及其算法研究综述[J]. 软件学报,2017,28(6):1547-1564.
CHEN L,CHEN S C. Survey on matrix completion models and algorithms[J]. Journal of Software,2017,28(6):1547-1564.(in Chinese)

[16] PACKARD N H,CRUTCHFIELD J P,FARMER J D,et al. Geometry from a time series[J]. Physical Review Letters,1980,45(9):712-716.

[17] 段文锋,张冀宁,黄卫星,等. 相空间导数重构法的探讨[J]. 四川大学学报(工程科学版),2001,33(5):102-106.
DUAN W F,ZHANG J N,HUANG W X,et al. Study on the derivative reconstruction method[J]. Journal of Sichuan University (Engineering Science Edition),2001,33(5):102-106.(in Chinese)

[18] XU B B,JACQUIR S,LAURENT G,et al. Analysis of an experimental model of *in vitro* cardiac tissue using phase space reconstruction[J]. Biomedical Signal Processing and Control,2014,13:313-326.

[19] LI H W,LIU J P,LI T,et al. Analysis of dynamic of two-phase flow in small channel based on phase space reconstruction combined with data reduction sub-frequency band wavelet[J]. Chinese Journal of Chemical Engineering,2015,23(6):1017-1026.

[20] 高俊杰. 混沌时间序列预测研究及应用[D]. 上海:上海交通大学,2013.
GAO J J. Research and application of chaotic time series prediction[D]. Shanghai:Shanghai Jiao Tong University,2013.(in Chinese)

[21] 陈帅,赵明,郭栋,等.基于SVD-KDR算法的工业监测数据插补技术[J]. 机械工程学报,2021,57(2):30-38.
CHEN S,ZHAO M,GUO D,et al. Missing data imputation using SVD-KDR algorithm in industrial monitoring data[J]. Journal of Mechanical Engineering,2021,57(2):30-38.(in Chinese)