

分形插值在风速时间序列中的应用

郭秀婷^{1,2}, 朱昶胜^{1*}, 张生财¹, 赵奎鹏¹

(1. 兰州理工大学 计算机与通信学院, 兰州 730050; 2. 兰州理工大学 理学院, 兰州 730050)

(* 通信作者电子邮箱 zhucs_2008@163.com)

摘要:针对风电场风速数据中大量连续缺失数据的插值问题,提出了一种基于自适应变异粒子群优化(PSO)的分形插值算法。首先,在粒子群优化算法中引入变异因子,增强粒子的多样性,提高算法搜索精度;其次,通过自适应变异粒子群优化算法来得到分形插值算法中垂直比例因子参数的最佳取值;最后,对两组不同趋势和变化特征的数据集进行分形插值计算分析,并把所提算法与Lagrange插值和三次样条插值方法进行对比。结果表明:分形插值不仅可以保持风速曲线的整体波动特性和局部特征,而且比传统插值方法的精度更高;在基于Dataset A的实验中,分形插值的均方根误差(RMSE)分别比Lagrange插值和三次样条插值减小了66.52%和58.57%;在基于Dataset B的实验中,分形插值的RMSE分别比Lagrange插值和三次样条插值减小了76.72%和67.33%。证明分形插值更适合连续缺失且波动强烈的风速时间序列的插值。

关键词:风速;时间序列;分形插值;垂直比例因子;自适应变异;粒子群优化

中图分类号:TP301.6 **文献标志码:**A

Application of fractal interpolation in wind speed time series

GUO Xiuting^{1,2}, ZHU Changsheng^{1*}, ZHANG Shengcai¹, ZHAO Kuipeng¹

(1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu 730050, China;

2. School of science, Lanzhou University of Technology, Lanzhou Gansu 730050, China)

Abstract: A fractal interpolation algorithm based on adaptive mutation Particle Swarm Optimization (PSO) was proposed aiming at the interpolation problem of a large number of continuous missing data in wind speed data of wind farms. First, the mutation factor was introduced into the particle swarm optimization algorithm to enhance the diversity of particles and the search accuracy of the algorithm. Second, the optimal value of the vertical scaling factor in the fractal interpolation algorithm was obtained by the adaptive mutation particle swarm optimization algorithm. Finally, two datasets with different trends and change characteristics were analyzed by fractal interpolation, and the proposed algorithm was compared with Lagrange interpolation and cubic spline interpolation. The results show that fractal interpolation is not only able to maintain the overall fluctuation characteristics and local characteristics of wind speed curve, but also is more accurate than the traditional interpolation methods. In the experiment based on Dataset A, the Root Mean Square Error (RMSE) of fractal interpolation was reduced by 66.52% and 58.57% respectively compared with those of Lagrange interpolation and cubic spline interpolation. In the experiment based on Dataset B, the RMSE of fractal interpolation was decreased by 76.72% and 67.33% respectively compared with those of Lagrange interpolation and cubic spline interpolation. It is verified that fractal interpolation is more suitable for the interpolation of wind speed time series with strong fluctuation and continuous missing data.

Key words: wind speed; time series; fractal interpolation; vertical scaling factor; adaptive variation; Particle Swarm Optimization (PSO)

0 引言

可再生能源发电具有不依赖化石燃料、减少温室气体排放、保护环境等一系列优点,因此得到了快速发展;其中风能作为一种低成本、分布域广的清洁能源得到了各国的高度重视^[1]。但风的间歇性和随机性给风力发电的规划及管理带来了很大困扰^[2-3],而风速预测和风电功率预测是应对风电大规

模并网运行问题的重要手段,因此风速预测和风电功率预测成为全球研究的热点^[4-6]。受天气因素、地理环境因素、传感器精度、电磁干扰以及通信故障等因素影响,采集到的风速数据会出现一定的异常值甚至一段时间的缺失值,但在研究过程中这些异常值和缺失值却至关重要,因此研究适用于风速的异常值检测及插值算法具有重要的意义^[7-8]。

陈伟等^[8]在处理风电异常数据时采用三次样条插值进行

收稿日期:2020-02-15;修回日期:2020-04-09;录用日期:2020-04-16。

基金项目:中国博士后科学基金面上—等资助项目(2014M560371);兰州理工大学红柳杰出人才计划项目(J201304)。

作者简介:郭秀婷(1982—),女,山东聊城人,讲师,博士研究生,主要研究方向:数据挖掘、时间序列预测;朱昶胜(1972—),男,甘肃兰州人,教授,博士,主要研究方向:高性能计算、大数据;张生财(1984—),男,甘肃兰州人,副教授,博士研究生,主要研究方向:大数据及云计算;赵奎鹏(1992—),男,甘肃武威人,硕士研究生,主要研究方向:大数据及云计算。

数据重构,但自然风速时间序列是一个波动强烈的非线性时间序列, Lagrange、三次样条和线性插值等传统插值算法中,任意两个信息点之间用直线或者光滑的曲线连接,掩盖了相邻插值点间的局部变化特征,因此用于波动强烈的风速时间序列效果并不理想;武艳强等^[9]提出的多点三次样条插值算法对于频率相对稳定的低频和中频曲线插值效果比较理想,但对于包含多种频率混合的非线性风速曲线却不适合;杨茂等^[10]提出用支持向量机重构风电数据,但由于参数的敏感性因素导致误差较大;陈伟等^[11]提出一种粒子群优化的最小二乘支持向量机对风速数据进行重构,但将缺失风速置为零后的时间序列参与模型训练,如果连续缺失数据较多,则参与模型训练的数据本身误差较大,无法保证插值效果,且计算量大;自然风速时间序列具有典型的分形特征^[12-13];文献[14-15]用分形插值算法及分段分形插值算法处理 TurbSim 仿真软件在特定条件下生成的模拟风速时间序列,这种理想风速时间序列不受瞬时天气条件的影响,与实测风速有一定的差距。

本文将自适应粒子群算法优化的分形插值算法应用到甘肃省某风电场实测风速时间序列的连续缺失值的重构问题,实验证明针对波动强烈、连续缺失数据量大的风速时间序列,分形插值算法可以取得比传统插值算法更高的精度。

1 分形插值原理

分形插值的概念是由美国数学家 Barnsley 于 1986 年提出来的。分形插值^[16]与多项式插值和曲线插值不同,它是由一类特殊的迭代函数系统 (Iterated Function System, IFS) 产生的,它是针对集合而非针对点的,能够反映出相邻已知相关点之间的局部特征,得到比传统插值方法更高的精度。

1.1 数据集

假设点集 $\{(x_i, y_i) \in \mathbf{R}^2 | i = 0, 1, \dots, N\}$, 其中 $\{(x_i, y_i) \in \mathbf{R}^2 | i = 0, 1, \dots, N\}$, 对应于这个数据集的插值函数是一个连续函数 $f: [x_0, x_N] \rightarrow \mathbf{R}$, 使得:

$$f(x_i) = y_i; i = 0, 1, \dots, N \tag{1}$$

点 $(x_i, y_i) \in \mathbf{R}^2$ 称为插值点, 函数 f 内插该数据且 f 的图形穿过该插值点。吸引子 $G = \{(x, f(x)); x \in [x_0, x_N]\}$ 是内插数据的连续函数的图像。

1.2 构造 IFS

考虑 IFS $\{\mathbf{R}^2; W_n, n = 1, 2, \dots, N\}$, 其中 W_n 是具有如下形式的仿射变换:

$$W_n \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_n & 0 \\ c_n & d_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_n \\ f_n \end{bmatrix} \tag{2}$$

并且满足如下条件:

$$\begin{cases} W_n \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ y_{n-1} \end{bmatrix} \\ W_n \begin{bmatrix} x_N \\ y_N \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} \end{cases} \tag{3}$$

其中 d_n 为自由变量, 取 $|d_n| < 1$, 否则 IFS 不收敛, 解方程得:

$$\begin{cases} a_n = (x_n - x_{n-1}) / (x_N - x_0) \\ e_n = (x_N x_{n-1} - x_0 x_n) / (x_N - x_0) \\ c_n = [y_n - y_{n-1} - d_n(y_N - y_0)] / (x_N - x_0) \\ f_n = [x_N y_{n-1} - x_0 y_n - d_n(x_N y_0 - x_0 y_N)] / (x_N - x_0) \end{cases} \tag{4}$$

将已知点代入方程(4), 不断迭代, 使系统收敛到其构造的吸引子 G 附近, 即完成了插值运算。

1.3 分形维数

分形维数是刻画分形特征的定量指标, 它可以衡量两个分形的相似程度, 是衡量分形物体不规则程度的参数。对于风速时间序列, 分形维数表示研究时间段内风速波动的强烈程度, 分形维数越大表明风速波动越剧烈^[17]。分形维数的计算遵循如下定理: 假定数据集 $\{(x_i, y_i)\}$ 不共线且 $\sum_{n=1}^N |d_n| > 1$, 则满足 G 的分形维数 D 是如下方程的唯一常数解:

$$\sum_{n=1}^N |d_n| a_n^{D-1} = 1 \tag{5}$$

否则, 满足 G 的分形维数等于 1。

式(5)表明分形维数 D 不依赖于因变量(数据集), 而自由变量垂直比例因子 d_n 较为关键, 它能影响分形维数 D 的确定。目前常用盒维数法求解分形维数, 即采用边长为 $\delta_i (i = 1, 2, \dots)$ 的正方形的网格覆盖分形体, 数出覆盖分形体所需盒子的数目 $N(\delta_i)$, 随着 $\lim \delta_i \rightarrow 0$, 得到一组数据 $(\ln \delta_i^{-1}, \ln N(\delta_i)) (i = 1, 2, \dots, M)$, 用最小二乘法拟合, 得到的拟合直线的斜率即为分形体的分形维数。

2 自适应变异粒子群优化算法

2.1 粒子群优化算法

粒子群优化 (Particle Swarm Optimization, PSO) 算法是一种群体智能的优化算法, 源于对鸟类捕食行为的研究。这种算法最早由 Kennedy 和 Eberhart 在 1995 年提出, 是一种非常成熟的求解优化问题的方法。该算法及其改进算法已在各种行业得到广泛应用^[18-20], 并取得了很好的效果。粒子群算法收敛速度快, 具有很强的通用性。算法中每个粒子都代表问题的一个潜在最优解, 用位置、速度和适应度值三项指标表示粒子的特征。位置用来存储用来优化的变量, 在本文中则表示分形插值的垂直比例因子参数值, 适应度值表示粒子的优劣, 粒子的速度决定粒子移动的方向和距离, 速度随着自身及其他粒子的移动经验进行动态调整。最初随机初始化粒子, 计算每个粒子的适应度值, 根据适应度值更新个体极值和群体极值, 然后粒子在解空间中运动, 通过跟踪个体极值和群体极值更新个体的位置和速度, 粒子每更新一次位置, 就重新计算一次适应度值, 通过比较新粒子的适应度值和个体极值、群体极值的适应度值更新个体极值和群体极值, 最终的群体极值即最优解。

每次迭代过程中, 粒子通过个体极值和群体极值更新自身的位置和速度, 公式如下:

$$\begin{cases} V_i^{k+1} = \omega V_i^k + c_1 r_1 (P_i^k - X_i^k) + c_2 r_2 (P_g^k - X_i^k); \\ X_i^{k+1} = X_i^k + \gamma V_i^{k+1} \end{cases}; \tag{6}$$

$$i = 1, 2, \dots, m$$

其中: V_i 为粒子的速度, X_i 为粒子的位置, P_i 为个体极值, P_g 为群体极值, ω 为惯性权重, c_1 和 c_2 为加速度因子, 是非负的常数, k 为当前迭代次数, r_1 和 r_2 为分布在 $[0, 1]$ 区间的随机数, γ 为约束因子。

其中 ω 的计算公式如下:

$$\omega^k = \omega_{start} - (\omega_{start} - \omega_{end}) * k / T_{max} \tag{7}$$

其中: ω_{start} 和 ω_{end} 分别为初始惯性权重和迭代至最大次数时

的惯性权重,分别取值 $\omega_{\text{start}} = 0.9$ 、 $\omega_{\text{end}} = 0.4$, T_{max} 为最大迭代次数。

本文中参数取值为 $c_1 = c_2 = 1.49$, $\gamma = 0.5$, 最大迭代次数 $T_{\text{max}} = 100$, 种群大小 $p = 15$ 。

2.2 自适应变异粒子群算法

粒子群算法虽然具有收敛快、通用性强的优点,但同时存在易早熟收敛、搜索精度低等缺点,因此在本文中提出将遗传算法中的变异思想引入 PSO 算法,即在粒子每次更新之后,引入变异算子,以一定的概率重新初始化粒子。变异操作拓展了在迭代中不断缩小的搜索空间,使粒子能够跳出当前搜索到的最佳位置,在更大的空间中开展搜索,保持了粒子的多样性,同时提高算法搜索到更优值的可能性。本文中采用的变异算子公式如下:

$$X_j = \begin{cases} 2 * \text{rand} - 1, & r > 0.9 \\ X_j, & r \leq 0.9 \end{cases} \quad (8)$$

其中 r 为在 $[0, 1]$ 区间的随机数。

3 衡量分形插值结果的指标

3.1 插值误差

插值曲线与实际曲线之间的误差称为插值误差。本次实验采用均方根误差(Root Mean Square Error, RMSE)对插值结果进行评价:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{x}(i) - x(i))^2}{N}} \quad (9)$$

其中: $\hat{x}(i)$ 为插值; $x(i)$ 为真实值; N 为插值点的总数量。

3.2 纳什效率系数

纳什效率系数(Nash-Sutcliffe Efficiency coefficient, NSE),用于表示模拟结果与实际结果的吻合程度,本文用于验证插值曲线与实际曲线的吻合程度,吻合程度越高表示插值效果越好,公式如下:

$$NSE = 1 - \frac{\sum_{i=1}^N (\hat{x}(i) - x(i))^2}{\sum_{i=1}^N (\hat{x}(i) - \bar{x})^2} \quad (10)$$

其中: $\hat{x}(i)$ 为插值; $x(i)$ 为真实值; N 为插值点的总数量, \bar{x} 为真实值的平均值。NSE 值越接近 1 表明两个时间序列吻合程度越高,即插值效果越好。

4 实验结果与讨论

所有实验运行环境均为: 1.60 GHz CPU (8 CPUs), 8.00 GB RAM, Windows 10, Matlab R2019a。

4.1 实验数据统计分析

采用甘肃省酒泉风电场 2017 年 1 月 9 日 09:55 至 22:20 的间隔 5 min 的 150 条数据集和 2017 年 1 月 10 日 00:00 至 14:30 间隔 5 min 的 175 条数据集进行实验,简称 Dataset A 和 Dataset B,这两组数据统计指标即:均值(Mean),最小值(Min),最大值(Max)和标准方差(Standard deviation, Std.),如表 1 和图 1 所示,由此可知该组数据表现出非线性和较强的波动性。在两组数据集中分别抽取 30 个数据点和 40 个数据点作为插值初始点。

表 1 实验数据集的统计指标

Tab. 1 Statistical indexes of experimental datasets

Dataset	N	Mean/ ($\text{m} \cdot \text{s}^{-1}$)	Min/ ($\text{m} \cdot \text{s}^{-1}$)	Max/ ($\text{m} \cdot \text{s}^{-1}$)	Std. / ($\text{m} \cdot \text{s}^{-1}$)
A	150	7.386 1	3.03	9.70	1.283 5
B	175	9.126 2	7.43	12.45	1.224 8

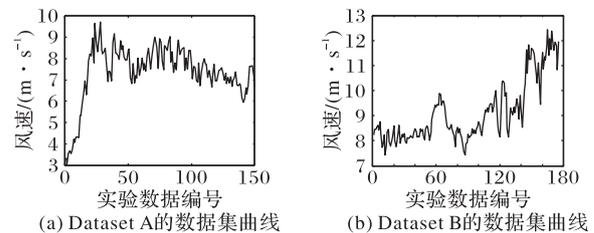


图 1 实验数据集的曲线

Fig. 1 Curves of experimental datasets

4.2 参数的设定

4.2.1 垂直比例因子

由式(4)可知,有一个自由变量 d_n ($|d_n| < 1$),即垂直比例因子,它是影响分形插值精度的关键因素,当 $d_n = 0$ 时,分形插值函数就变成分段线性插值函数。图 2 为选取不同的垂直比例因子时图形呈现的不同形态,由图 2 可以看出随着 $|d_n|$ 的增大,分形插值拟合曲线越来越粗糙,波动幅度越来越大,曲线的局部特征也更加明显。分形维数是刻画风速序列波动的强烈程度的指标,图 3 验证了 $|d_n|$ 越大,分形维数 D 越大,即风速曲线波动越强烈。

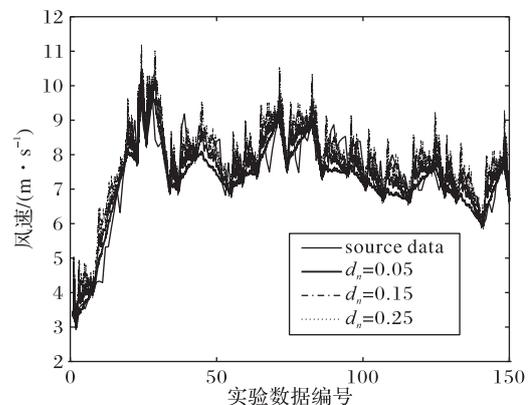


图 2 垂直比例因子对曲线的影响

Fig. 2 Influence of vertical scaling factor on curve

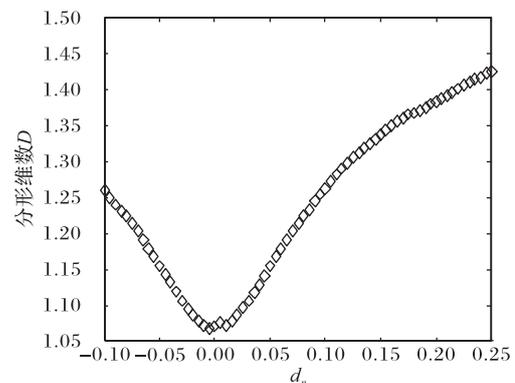


图 3 分形维数随垂直比例因子的变化

Fig. 3 Change of fractal dimension with vertical scaling factor

但 $|d_n|$ 并不是越大越好,即不代表插值曲线局部特征越复杂越好,图 4 为 RMSE 随 d_n 取值的变化,由此可知, d_n 严重影响分形插值的误差。

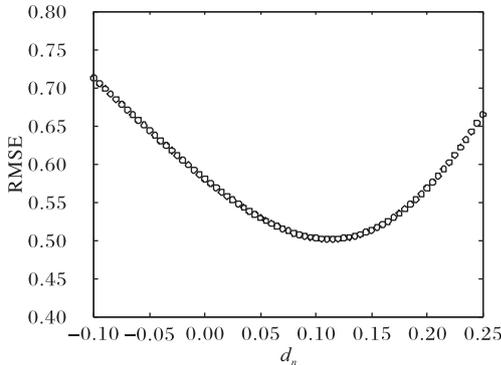
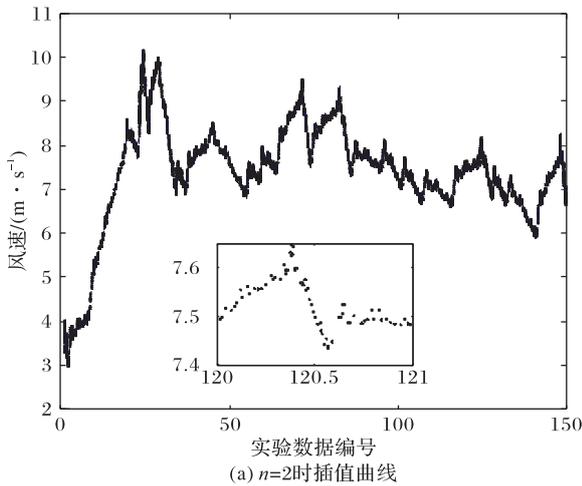


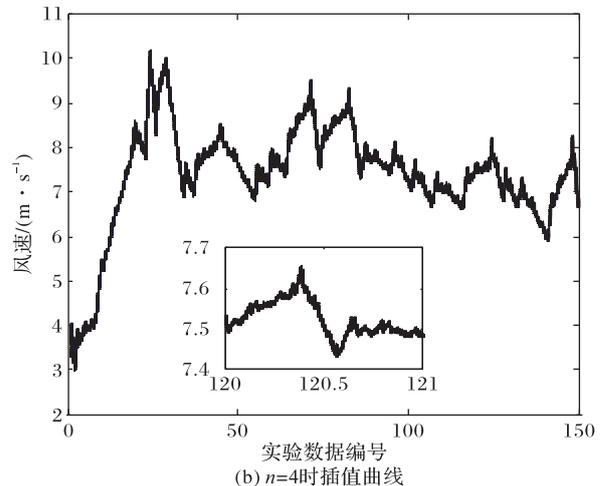
图 4 RMSE 随垂直比例因子的变化

Fig. 4 Change of RMSE with vertical scaling factor

本文采用自适应变异粒子群算法对垂直比例因子 d_n 进行寻优,粒子群中每个粒子包含垂直比例因子 $\{d_i\}_{i=1}^N$,且以式



(a) $n=2$ 时插值曲线



(b) $n=4$ 时插值曲线

图 5 迭代次数对插值点数的影响

Fig. 5 Influence of iteration number on interpolation point number

表 2 迭代次数对各指标的影响

Tab. 2 Influence of iteration number on different indexes

n	D	RMSE	运行时间/s	插值点数
2	1.2908	0.5018	0.038732	20412
3	1.3100	0.5017	1.011550	551124
4	1.3202	0.5019	36.743503	14880348

4.3 分形插值与传统插值的插值结果的对比研究

分别采用三次样条插值、Lagrange 插值等传统插值方法和分形插值算法进行比较分析, Dataset A 数据集的插值结果如图 6 所示。结果表明:分形插值较好地保持了图形的形状以及粗糙程度,能更好地凸显图形的局部细节特征;而 Lagrange 插值和三次样条插值只是保持了图形的大致形状,相邻点间用光滑曲线连接,忽略了图形局部丰富的信息。

表 3 是对 Dataset A 数据集插值结果的各种量化指标,三种插值结果的均值与原始数据相差均不大,对于表征曲线波动强烈程度的分形维数, Lagrange 插值与分形插值相当,三次样条插值最差;分形插值误差最小为 0.5018,与 Lagrange 插值算法比较,减小了 66.52%,与三次样条插值算法比较,减小了 58.57%;分形插值序列与原始序列的吻合程度同样最佳,达到 0.8120,

(9) RMSE 为适应度值,即最优垂直比例因子取值满足以下公式:

$$\min RMSE(d_1, d_2, \dots, d_n)$$

$$\text{s. t. } |d_n| < 1$$

$$(11)$$

4.2.2 迭代次数

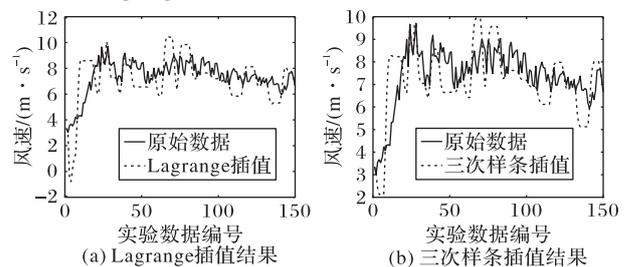
表 2 给出了迭代次数对分形维数、插值误差、运行时间及插值点数的影响。由表 2 可知,随着迭代次数的增多,分形维数基本不变,即曲线波动的强烈程度不随迭代次数的改变而改变,插值误差也基本不受迭代次数的影响,但迭代次数越多,插值得到的数据点数越大,曲线细节会更丰富,但不会变光滑。图 5 给出了迭代次数分别为 2 和 4 时的图形,由局部放大图可以明显看到迭代 2 次和迭代 4 次插值得到的数据点的密集程度的差别。迭代次数 n 与生成的插值点数 m 之间的关系^[17]为:

$$m = (q - 1)^{2^n} + 1$$

$$(12)$$

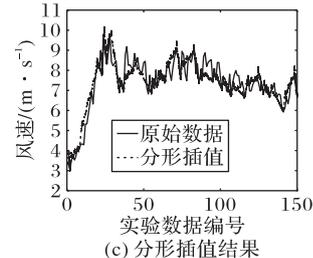
其中 q 为初始插值点的个数。虽然迭代次数越多,插值得到的曲线细节越丰富,但该参数不宜取得过大,符合实际插值要求即可,迭代次数过大会增加计算时间,降低计算效率,本次实验迭代次数选择 2 次即可满足要求。

远远超过 Lagrange 插值和三次样条插值两种传统插值方法。



(a) Lagrange 插值结果

(b) 三次样条插值结果



(c) 分形插值结果

图 6 Dataset A 插值效果比较

Fig. 6 Interpolation effect comparison of Dataset A

由以上分析可知,对于波动强烈并且连续缺失数据较多

的风速时间序列,分形插值不仅可以保持图形的整体形状和波动特性,而且插值精度更高,与原始曲线吻合更好,与传统的插值算法相比,分形插值具有明显的优势。为了进一步证实上述结论的可靠性,对 Dataset B 数据集插值计算,得到插值结果如图 7 所示。

表 3 Dataset A 插值性能对比

Tab. 3 Interpolation performance comparison of Dataset A

插值方法	风速平均值/ ($\text{m}\cdot\text{s}^{-1}$)	D	RMSE	NSE
原始数据	7.386 1	1.356 2	—	—
Lagrange 插值	7.163 2	1.293 0	1.498 7	0.363 8
三次样条插值	7.191 9	1.260 6	1.211 1	0.382 7
分形插值	7.585 1	1.290 8	0.501 8	0.812 0

注:黑体字标注值是各评价指标的最佳值。

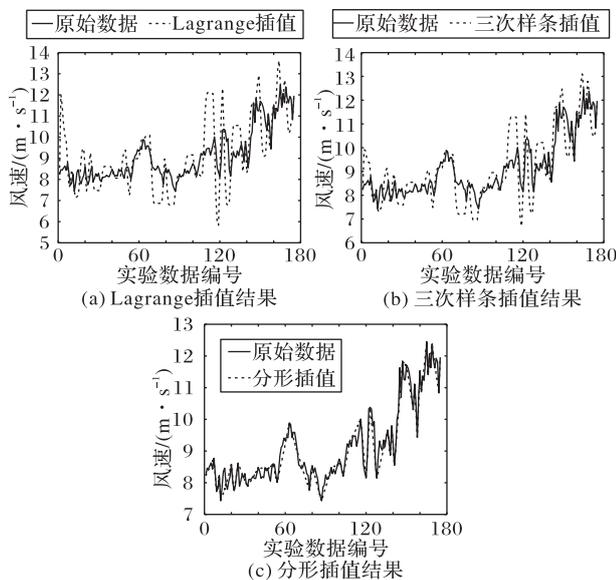


图 7 Dataset B 插值效果比较

Fig. 7 Interpolation effect comparison of Dataset B

表 4 给出了 Dataset B 数据集的各插值结果。Dataset B 数据集的各插值误差 RMSE 和纳什效率系数 NSE 的规律与 Dataset A 的规律相同,进一步说明了分形插值比传统的插值方法更适合连续缺失数据较多且波动强烈的风速序列。

表 4 Dataset B 插值性能对比

Tab. 4 Interpolation performance comparison of Dataset B

插值方法	风速平均值/ ($\text{m}\cdot\text{s}^{-1}$)	D	RMSE	NSE
原始数据	9.126 2	1.392 7	—	—
Lagrange 插值	9.176 8	1.380 2	1.170 3	0.515 6
三次样条插值	9.153 4	1.376 2	0.834 2	0.680 9
分形插值	9.289 4	1.199 0	0.272 5	0.950 2

5 结语

自适应变异粒子群优化的分形插值算法为风速序列的连续缺失数据插值提供了一种新途径,并对分形插值算法中的两个重要参数进行讨论分析,利用自适应变异粒子群算法对垂直比例因子寻优,最后对分形插值算法与传统插值算法的进行了对比研究。通过计算和验证得出如下结论:

1) 垂直比例因子是影响分形插值精度的重要因素,采用

自适应变异粒子群优化垂直比例因子,分形插值取得了很好的插值效果,证明了其有效性。

2) 针对连续缺失的波动强烈的风速序列,分形插值比传统插值方法更有优势,分形插值不仅可以保持风速曲线的整体变化趋势,而且可以反映数据的局部波动特征,并且具有更高的插值精度和吻合度。

参考文献 (References)

- [1] Global Wind Energy Council. Global wind report 2018 [EB/OL]. [2019-09-11]. <https://gwec.net/global-wind-report-2018/>.
- [2] 杨悦,李国庆. 大规模风电并网对互联电力系统低频振荡影响研究[J]. 太阳能学报, 2017, 38(10):2665-2674. (YANG R, LI G Q. Influence of large scale grid connected wind power on low frequency oscillation of interconnected power system [J]. Acta Energetica Solaris Sinica, 2017, 38(10):2665-2674.)
- [3] CUI B, XUE A, BI T, et al. Analysis on the static voltage stability of an actual power system with large-scale wind farm [C]// Proceedings of the 2nd IET Renewable Power Generation Conference. Stevenage: IET, 2013:1-5.
- [4] AHMED A, KHALID M. An intelligent framework for short-term multi-step wind speed forecasting based on Functional Networks [J]. Applied Energy, 2018, 225:902-911.
- [5] XIU C, WANG T, TIAN M, et al. Short-term prediction method of wind speed series based on fractal interpolation[J]. Chaos, Solitons and Fractals, 2014, 68:89-97.
- [6] NAIK J, BISOI R, DASH P K. Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression [J]. Renewable Energy, 2018, 129(Pt A):357-383.
- [7] 赵永宁,叶林,朱倩雯. 风电场弃风异常数据簇的特征及处理方法[J]. 电力系统自动化, 2014, 38(21):39-46. (ZHAO Y N, YE L, ZHU Q W. Characteristics and processing method of abnormal data clusters caused by wind curtailments in wind farms [J]. Automation of Electric Power Systems, 2014, 38(21):39-46.)
- [8] 陈伟,吴布托,裴喜平. 风电机组异常数据预处理的分类多模型算法[J]. 电力系统及其自动化学报, 2018, 30(4):137-143. (CHEN W, WU B T, PEI X P. Classification multi-model algorithm for abnormal data preprocessing in wind turbines [J]. Proceedings of the CSU-EPSA, 2018, 30(4):137-143.)
- [9] 武艳强,黄立人. 时间序列处理的新插值方法[J]. 大地测量与地球动力学, 2004, 24(4):43-47. (WU Y Q, HUANG L R. A new interpolation method in time series analyzing [J]. Journal of Geodesy and Geodynamics, 2004, 24(4):43-47.)
- [10] 杨茂,翟冠强,李大勇,等. 基于风速升降特性及支持向量机理论的异常数据重构算法[J]. 电力系统保护与控制, 2018, 46(16):31-37. (YANG M, ZHAI G Q, LI D Y, et al. An algorithm of abnormal data reconstruction based on RISE-FALL-feature of the wind speed and support vector machine [J]. Power System Protection and Control, 2018, 46(16):31-37.)
- [11] 陈伟,王敏,裴喜平. 一种风电场风速异常数据预处理的新方法[J]. 兰州理工大学学报, 2019, 45(5):91-96. (CHEN W, WANG M, PEI X P. A new method for data preprocessing of wind speed anomalies at wind farm [J]. Journal of Lanzhou University of

- Technology, 2019, 45(5):91-96.)
- [12] 钟莉,李杰. 风速时程的分形特征分析[J]. 华中科技大学学报(城市科学版), 2008, 25(4):273-275. (ZHONG L, LI J. Fractal characteristic analysis of wind speed time history [J]. Journal of Huazhong University of Science and Technology (Urban Science Edition), 2008, 25(4):273-275.)
- [13] CHANG T P, KO H H, LIU F J, et al. Fractal dimension of wind speed time series[J]. Applied Energy, 2012, 93:742-749.
- [14] 张俊伟,李春,李倩倩,等. 基于风速时间序列分形与传统插值对比研究[J]. 热能动力工程, 2018, 33(7):122-127, 48. (ZHANG J W, LI C, LI Q Q, et al. The contrastive study of fractal and traditional interpolation based on wind speed time series [J]. Journal of Engineering for Thermal Energy and Power, 2018, 33(7):122-127, 48.)
- [15] 李洋,李春,李倩倩,等. 分段分形插值在风速时间序列中的应用[J]. 热能动力工程, 2018, 33(12):118-124, 153. (LI Y, LI C, LI Q Q, et al. Application of segmental fractal interpolation in wind speed time series [J]. Journal of Engineering for Thermal Energy and Power, 2018, 33(12):118-124, 153.)
- [16] 孙洪泉. 分形几何与分形插值[M]. 北京:科学出版社,2011:99-116, 197. (SUN H Q. Fractal Geometry and Fractal Interpolation[M]. Beijing: Science Press, 2011:99-116,197.)
- [17] 李倩倩,李春,杨阳. 自相似性和分形维数在风场分析中的应用[J]. 动力工程学报, 2016, 36(11):914-919, 926. (LI Q Q, LI C, YANG Y. Application of wind speed self-similarity and fractal dimension in wind field analysis [J]. Journal of Chinese Society of Power Engineering, 2016, 36(11):914-919, 926.)
- [18] 周海鹏,高芹,蒋丰千,等. 自适应混沌量子粒子群算法及其在WSN覆盖优化中的应用[J]. 计算机应用, 2018, 38(4):1064-1071. (ZHOU H P, GAO Q, JIANG F Q, et al. Application of self-adaptive chaotic quantum particle swarm algorithm in coverage optimization of wireless sensor network [J]. Journal of Computer Applications, 2018, 38(4):1064-1071.)
- [19] 郑庆新,顾晓辉,张洪铭. 基于SQP和自适应搜索的混沌粒子群算法[J]. 计算机工程与应用, 2018, 54(13):131-136. (ZHENG Q X, GU X H, ZHANG H M. Chaotic particle swarm optimization algorithm based on SQP and adaptive search [J]. Computer Engineering and Applications, 2018, 54(13):131-136.)
- [20] 范君,王新,徐慧. 粒子群优化混合核极限学习机的构造煤厚度预测方法[J]. 计算机应用, 2018, 38(6):1820-1825. (FAN J, WANG X, XU H. Prediction method of tectonic coal thickness based on particle swarm optimized hybrid kernel extreme learning machine [J]. Journal of Computer Applications, 2018, 38(6):1820-1825.)

This work is partially supported by the First-class China Postdoctoral Science Foundation (2014M560371), the Project of Rose Willow Distinguished Young Scientists of Lanzhou University of Technology (J201304).

GUO Xiuting, born in 1982, Ph. D. candidate, lecturer. Her research interests include data mining, time series prediction.

ZHU Changsheng, born in 1972, Ph. D., professor. His research interests include high performance computing, big data.

ZHANG Shengcai, born in 1984, Ph. D. candidate, associate professor. His research interests include big data and cloud computing.

ZHAO Kuipeng, born in 1992, M. S. candidate. His research interests include big data and cloud computing.