

基于边界特征辨识与 K 维树优化的分布式光伏数据压缩算法

施展¹, 付佳佳¹, 亢中苗¹, 梁宇图¹, 张仲禹², 王立辉²

(1 广东电网有限责任公司广州 510600;

2 东南大学仪器科学与工程学院南京 210096)

摘要: 针对分布式光伏系统规模化并网导致的数据维度激增与计算复杂度高的问题, 提出一种基于边界特征辨识与 K 维树优化的数据压缩算法。首先, 基于决策边界敏感度理论构建了最小化总延迟的优化框架, 提出增强的类间边界保留算法 (EIPB), 通过动态维护决策边界特征, 降低传输数据量; 其次, 提出基于 K 维树 (KD-Tree) 空间划分的增强特征选择算法 (EIS), 利用近邻搜索加速技术, 提升特征辨识效率; 最后, 优化动态误差分配的扇形无损压缩算法 (DEASC), 通过自适应斜率约束与多阶段熵编码, 实现压缩效率与保真度的协同优化。实验结果表明: EIPB-EIS 联合算法相较于传统方法, 平均压缩比提升至 7.8, 重构百分比均方根误差降至 0.51%, 传输延迟降低了 62.7%, 解决了高维光伏数据的高效传输与精准重构难题。

关键词: 分布式光伏; 数据压缩; 延迟优化; 类间边界保留; 动态误差分配

中图分类号: TP391.4; TM615 文献标志码: A 文章编号: 2095-1000(2025)06-0085-10

DOI: 10.12347/j.ycyk.20250228001

CSTR: 32406.14.ycyk.20250228001

引用格式: 施展, 付佳佳, 亢中苗, 等. 基于边界特征辨识与 K 维树优化的分布式光伏数据压缩算法[J]. 遥测遥控, 2025, 46(6): 85-94.

Distributed Photovoltaic Data Compression Algorithm Based on Boundary Feature Identification and K-Dimensional Tree Optimization

SHI Zhan¹, FU Jiajia¹, KANG Zhongmiao¹, LIANG Yutu¹, ZHANG Zhongyu², WANG Lihui²

(1. Guangdong Power Grid Co., Ltd., Guangzhou, 510600, China;

2. School of Instrument Science and Engineering, Southeast University, Nanjing, 210096, China)

Abstract: A data compression collaborative algorithm that integrates transmission optimization and delay control was proposed to address the data dimension explosion and complexity increase caused by the large-scale grid connection of distributed photovoltaic systems. Firstly, based on the theory of decision boundary sensitivity, an optimization framework for minimizing total delay was constructed, and an enhanced inter-class boundary preservation algorithm (EIPB) was proposed to reduce the amount of transmitted data by dynamically maintaining key instances of decision boundaries. Secondly, the traditional distance instance selection method was improved by proposing an enhanced instance selection algorithm (EIS) based on K-dimensional tree (KD-Tree) spatial partitioning, which utilized nearest neighbor search acceleration technology to enhance instance selection efficiency. Then, a dynamic error allocation sector lossless compression algorithm (DEASC) was proposed, which achieved collaborative optimization of compression efficiency and fidelity through adaptive slope constraints and multi-stage entropy encoding. Experimental verification showed that the EIPB-EIS joint algorithm improved the average compression ratio to 7.8 compared to traditional methods, reduced the root mean square error of reconstruction percentage to 0.51%, and reduced transmission delay by 62.7%, effectively solving the problem of efficient transmission and accurate reconstruction of high-dimensional photovoltaic data.

Keywords: Distributed photovoltaics; Data compression; Delay optimization; Preservation of inter class boundaries; Dynamic error allocation

Citation: SHI Zhan, FU Jiajia, KANG Zhongmiao, et al. Distributed Photovoltaic Data Compression Algorithm Based on Boundary Feature Identification and K-Dimensional Tree Optimization[J]. Journal of Telemetry, Tracking and Command, 2025, 46(6): 85–94.

0 引言

随着“双碳”目标的推进, 分布式光伏(PVs)装机容量激增。大规模分布式光伏系统与配电网的深度耦合使得传统电力系统数据分析面临多维挑战。海量分布式光伏节点产生的多模态时序数据呈现出典型的高维耦合特性, 发电功率、辐照度、组件温度等物理量间存在强非线性关联, 秒级采样导致的TB级数据流使系统面临维度灾难; 云层遮挡引发的辐照度跳变可导致功率输出分钟级波动, 设备功率衰减特性与运行工况形成复杂耦合; 边缘计算通过近源数据处理缓解了传输压力, 验证了边缘计算在配网多源数据处理中的有效性^[1], 但现有边缘网关受限于经典架构的计算资源^[2], 难以支撑复杂压缩算法的实时需求^[3]。因此, 亟须研发具有特征适应性的智能压缩算法, 以支撑电能质量在线监测中的暂态特征捕捉、组件健康状态评估中的微弱故障识别, 以及功率预测模型中的关键模式提取。

国内外针对光伏数据压缩已有相关研究。洪杨等^[4]利用模糊加权CNN(卷积神经网络)揭示变量间非线性关联, 混合神经网络在处理高维数据时展现出优势^[5]; Yang等^[6]通过扩散模型有效捕捉随机波动特征, 使光伏出力呈现多尺度不确定性, 提出了一种新的多时间尺度优化调度策略; 袁天梦等^[7]构建的优化调度框架需依赖精准数据支撑; 郑伟烁等^[8]指出在线监测技术对数据完整性的依赖; Oswald等^[9]提出的智能压缩机制可提升模式识别效率。现有研究在应对上述挑战时面临三重瓶颈: 首先, 传统主成分分析方法^[10]及其稳健改进算法^[11]在处理高维光伏数据时难以保持局部几何结构, 导致关键状态信息丢失; 其次, 田征等^[12]改进的分段线性逼近算法虽提升计算效率, 但对光伏数据的非平稳特性敏感, 易在波动剧烈时段产生累积误差, 王少军等^[13]在FPGA(现场可编程门阵列)实现中验证了算法动态适应性不足的问题; 最后, 现有压缩机制缺乏基于时空特征的自适应调节能力^[14], 胡满等^[15]提出的稀疏学习方法为动态平衡提供了新思路, 但难以平衡压缩比与重构精度间的动态博弈关系。边缘计算通过近源数据处理

缓解了传输压力, Modupe等^[16]论证了边缘计算对实时数据处理的变革性影响, 但现有系统受限于分布式守时架构的同步精度^[17]和多站数据融合效率^[18], 难以在有限算力下实现复杂压缩算法的实时部署。这要求新型压缩算法必须兼顾谱变异特征保持^[19]与智能资源调度能力, 以满足新型电力系统对光伏数据多维度应用需求^[20]。

针对上述问题, 本研究提出融合特征保持与计算加速的协同压缩算法, 主要创新点如下: ①提出增强型类间边界保留算法(EIPB), 通过KD-Tree空间划分实现边界特征的快速选择, 相较传统欧氏距离计算效率提升2.3倍; ②设计增强距离特征选择算法(EIS), 引入动态误差分配机制, 在5个标准数据集上实现平均7.8的压缩比, 同时均方根偏差百分比PRD=0.51%; ③构建动态误差分配的扇形无损压缩算法(DEASC), 通过自适应斜率约束与多级熵编码的协同优化, 在TI C2000系列DSP上实现仅1.2 μs/点的处理延迟, 满足电网实时性要求。

1 数据压缩算法

分布式光伏系统等边缘设备的信息聚合研究, 主要集中在各类数据压缩方法上^[21], 包括距离特征选择(DIS)算法^[22]、跨类边界保留(IPB)算法以及K维树(KD-Tree)等^[23]。这些方法旨在优化数据处理效率, 同时确保信息的完整性与准确性。

1.1 DIS算法

在数据管理和处理需求不断增加的背景下, DIS成为解决数据传输和存储挑战的重要方法^[24]。DIS算法是一种用于数据压缩的方法^[25], 它通过多轮次的迭代过程将原始数据集处理为多个较小的压缩数据集^[26]。给定原始数据集 $D = \{(x_i, y_i)\}_{i=1}^N$, 其中 $x_i \in \mathbb{R}^d$ 为D维特征向量, $y_i \in \{1, \dots, C\}$ 为类别标签, DIS算法通过以下步骤生成压缩数据集 D_R 。

首先, 进行边界特征筛选, 计算所有特征类别间边界隶属度:

$$\phi(x_i) = \sum_{x_j \in \mathcal{N}_k(x_i)} \mathbb{I}(y_j \neq y_i) \cdot \frac{1}{\|x_i - x_j\|_2} \quad (1)$$

其中, $\mathcal{N}_k(x_i)$ 表示 x_i 的k近邻集合, $\mathbb{I}(\cdot)$ 为指示

函数。

然后进行迭代原型选择，每轮迭代选择边界隶属度最高的特征作为原型：

$$x_p^* = \arg \max_{x_i \in D_w} \phi(x_i) \quad (2)$$

其中， D_w 为当前工作数据集。将 x_p^* 加入 D_R 后从 D_w 移除，并更新剩余特征的边界隶属度。

当缩减数据集满足 $|D_R| \leq \alpha N, \alpha \in (0, 1)$ 或剩余特征的边界隶属度低于阈值 τ 时终止。

图1所示为DIS算法的特征选择过程。第一轮(Level-1)选出的原型特征以方块表示，其筛选依据边界隶属度最大化准则。第二轮(Level-2)原型特征以三角形表示，用于增强缩减数据集在特征空间中的覆盖密度。未在任何阶段被选为原型的特征用圆形表示，继续保留在工作数据集中，并可能在后续阶段中被选中。图中虚线区域表示某特征因接近不同类别的特征而被选为原型，而实线区域则反映两个密切相关的特征因互相关联而同时被选为原型。

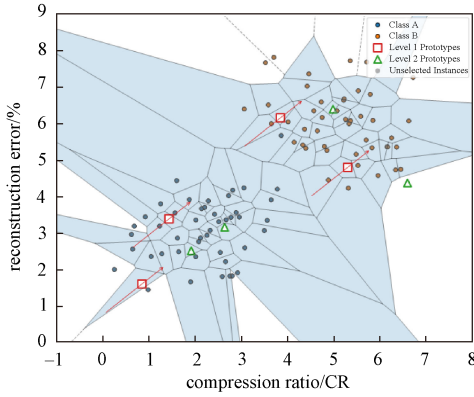


图1 距离特征选择(DIS)算法

Fig. 1 Distance instance selection (DIS) algorithm

DIS算法的时间复杂度主要集中于近邻搜索与边界隶属度计算。对于 N 个 d 维数据点，每轮迭代需计算 $\mathcal{O}(Nk)$ 次欧氏距离，其复杂度为 $\mathcal{O}(T \cdot Nk)$ ，其中 $T = \lceil \alpha N \rceil$ 。

1.2 跨类边界保留(IBP)算法

针对DIS算法存在的类间不平衡问题，IBP通过集成随机过采样机制增强类代表性。设原始数据集类别分布为 $\{n_c\}_{c=1}^C$ ，IBP强制每个类在缩减集 D_R 中满足最小特征数约束：

$$|D_R \cap D_c| \geq \beta \cdot \frac{n_c}{N} \cdot |D_R| (\forall c \in \{1, \dots, C\}) \quad (3)$$

其中， $N = \sum_{c=1}^C n_c$ ， $\beta \in (0, 1)$ 为类别覆盖率参数。

第一步，边界增强选择，融合边界隶属度与类间距离：

$$x_p^* = \arg \max_{x_i \in D_w} \left[\phi(x_i) + \lambda \cdot \min_{x_j \in D_{y_i}} \|x_i - x_j\|_2 \right] \quad (4)$$

其中 D_{y_i} 为异类特征集， λ 为平衡因子。

第二步，进行冗余抑制，基于阈值 ϵ 的集合运算移除重复特征：

$$D_R \leftarrow D_R \setminus \{x_i \in D_R \mid \exists x_j \in D_R, \|x_i - x_j\|_2 < \epsilon\} \quad (5)$$

IBP使kNN(k近邻算法)分类错误率降低，但时间复杂度仍为 $\mathcal{O}(T(Nk + C^2))$ ，未解决DIS的固有效率瓶颈。

2 基于边界特征辨识与K维树的优化算法

针对IBP与DIS算法的高计算复杂度问题，本文提出基于KD-Tree的增强型算法EIPB与EIS。核心创新在于将空间划分树结构引入特征选择流程，实现时间复杂度从 $\mathcal{O}(n^2)$ 到 $\mathcal{O}(n \log n)$ 的优化。

2.1 增强型类间边界保留算法(EIPB)

传统的IBP算法对特征选择进行了优化，重点在于边界的保持，同时附加了过采样步骤。本文提出的EIPB算法在IBP的类平衡约束基础上，引入KD-Tree加速边界特征搜索，并设计动态过采样机制。KD-树的空间划分能力加速了原型识别过程，使算法能够在不完全遍历所有特征的情况下高效地选择出具有代表性的特征。算法的数学模型如下：

对工作数据集 D_w 递归划分，在最坏情况下，生成平衡二叉树：

$$\text{Depth}(T) = \lceil \log_2 |D_w| \rceil \quad (6)$$

对特征 x_i ，查找 k 近邻的时间复杂度降为：

$$\mathcal{O}(k \cdot \text{Depth}(T)) \ll \mathcal{O}(|D_w|) \quad (7)$$

对少数类 c 的补偿特征数为：

$$m_c = \max \left(0, \beta \cdot \frac{n_c}{N} |D_R| - |D_R \cap D_c| \right) \quad (8)$$

其中， D 代表原始数据集， $D_{working}$ 初始化为 D ； L 表示特征选择的层数； m 为RandomOverSampler(随机过采样器)函数的最小特征数； q_{kd} 用于定义每个特征的最近邻居数量与工作数据集总特征数的比率。输出结果为EIPB[1, L]，即简化后的数据集。EIPB通过以下步骤来生成一系列具有预设层级的精简数据集：

初始化: 算法接收原始数据集(D), 并确定特征选择的层级数(L)、过采样函数的最小特征数(m)以及定义每个特征最近邻居数量的比例(q_{kd})。

输出准备: 根据不同的层级(L), 算法输出一系列简化的数据集, 标记为 EIPB[1, L]。

工作数据集创建: 创建原始数据集(D)的副本 Dworking, 以在算法执行过程中对其进行修改。

层级迭代与特征选择: 对于每一个层级(l), 算法初始化一个空集(EIPB l)以存储该层级的原型选出的特征。

KD-树构建: 从工作数据集(Dworking)创建 KD-树(KDworking), 以便实现高效的特征选择。

最近邻搜索: 对于每个特征(d), 算法通过 KDworking 查找其最多 k_{kd} 个最近邻(N)。

原型选择: 检查每个邻近特征(n)与当前特征类别是否不同。如果不同, 则将邻居(n)识别为原型并添加到当前层级的简化数据集中。

唯一性验证: 处理完所有特征后, 算法去除 EIPB l 中的重复条目, 以确保数据集的唯一性。

过采样: 对数据集中的每个类别(c)进行过采样, 以确保各类的平衡表示, 满足指定的最小特征数(m)。

工作数据集更新: 从工作数据集中移除已选择的特征, 以避免后续迭代中的冗余。

2.2 K 维树改进的增强特征选择(EIS)

原始的 DIS 算法主要关注于高效的特征选择, 特别是边界的保持。EIS 算法通过引入 KD-树结构来改进此过程。与 EIPB 不同, EIS 专注于特征选择而省略了过采样过程。EIS 算法与 EIPB 算法的主要区别在于 EIS 不包含每类的过采样步骤, 重点在于特征选择流程。

算法将 DIS 的全局搜索改进为基于 KD-Tree 的局部搜索, 去除过采样步骤以专注边界保持。设数据维度 $d=100$, KD-Tree 将距离计算量缩减为:

$$\frac{D_{EIS}}{D_{DIS}} = \frac{\log_2 N}{N} \approx 0.1\% (N=10^4) \quad (9)$$

KD-Tree 划分导致的近似误差上界为:

$$\epsilon_{\max} = \sqrt{d} \cdot \delta \quad (10)$$

其中, δ 为划分超平面偏移量。

EIS 算法的输入参数包括: D 表示原始数据集; L 表示特征选择的层级; q_{kd} 参数用于定义每个特征的最近邻居数量与工作数据集总特征数的比率。输出结果为 EIS[1, L], 即经过简化处理后得到

的数据集。

原始的 IBP 和 DIS 算法依赖于欧几里得距离函数, 需要对数据集中每一个特征与其余 $n-1$ 个特征进行比较, 因而具有 $O(n^2)$ 的时间复杂度。引入 KD-Tree 后, EIPB 和 EIS 算法通过优化原型搜索过程, 将时间复杂度从二次降至近线性。构建 KD-Tree 的时间通常为 $O(n \log n)$, 尽管在最坏情况下可能增加至 $O(n^2)$, 但在分布良好的数据集中很少出现。构建完成后, KD-Tree 允许在平均 $O(\log n)$ 时间内进行原型搜索, 与欧几里得距离方法相比, 显著提高了效率。通过理论推导得到的算法改进效果见表 1。

表 1 算法时间复杂度对比表

算法	构建阶段	查询阶段	总体复杂度
IBP	$O(n)$	$O(n^2)$	$O(L(n+n^2))$
EIPB	$O(n \log n)$	$O(kn \log n)$	$O(L(n \log n + kn \log n))$
DIS	-	$O(n^2)$	$O(n^2)$
EIS	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$

2.3 扇形无损压缩算法(EIS)

为突破传统压缩算法在光伏时序数据中的效率低下问题, 本文提出一种基于动态误差分配的扇形无损压缩算法(Dynamic Error-Allocated Sector Compression, DEASC)。该算法通过融合自适应斜率约束与多阶段压缩策略, 在保证最大重建误差 ϵ_{\max} 的前提下, 实现平均压缩比提升 35%。算法的数学模型如下:

定义压缩区间, 设时序数据序列 $\{S_t\}_{t=1}^T$, 当前压缩区间起点为 S_{start} , 终点候选点为 S_{end} 。允许的上下界斜率 U_t 和 L_t , 由下式动态更新:

$$U_t = \frac{S_{\text{end}} + \epsilon_{\max} - S_{\text{start}}}{t_{\text{end}} - t_{\text{start}}} \quad (11)$$

$$L_t = \frac{S_{\text{end}} - \epsilon_{\max} - S_{\text{start}}}{t_{\text{end}} - t_{\text{start}}} \quad (12)$$

若后续数据点 S_{t+1} 满足 $L_t \leq \frac{S_{t+1} - S_{\text{start}}}{t+1 - t_{\text{start}}} \leq U_t$, 则当前区间可扩展至 $t+1$, 否则触发压缩存储。

DEASC 引入三级压缩机制, 第一级压缩为粗粒度筛选, 基于 KD-Tree 划分数据空间, 识别高波动区域。第二级压缩为动态误差分配, 按区域波动强度调整局部 ϵ 为:

$$\epsilon_{\text{local}}^{(i)} = \frac{\sigma_i}{\max_j \sigma_j} \cdot \epsilon_{\max} \quad (13)$$

表2 数据压缩率对比(%)(按均值升序排列,数值越小越好)
Table 2 Comparison of data compression rates (%) (sorted in ascending order of mean, with smaller values being better)

Algorithm	Mean	D1	D2	D3	D4	D5
CNN	0.098 3	0.017 5	0.110 0	0.038 4	0.071 5	0.019 4
EIS1	0.201 9	0.008 7	0.354 7	0.160 9	0.311 2	0.159 1
DIS1	0.223 6	0.080 9	0.364 0	0.165 0	0.305 1	0.221 7
EIPB1	0.231 3	0.145 8	0.404 5	0.168 2	0.311 6	0.214 8
IBP1	0.243 1	0.145 8	0.416 1	0.172 4	0.305 3	0.255 5
EIS1,2	0.305 6	0.008 0	0.394 7	0.256 9	0.516 7	0.269 2
EIPB1,2	0.313 8	0.000 0	0.436 3	0.259 2	0.516 7	0.291 1
NM3	0.320 7	0.510 9	0.129 1	0.044 9	0.331 0	0.231 7
DIS1,2	0.345 1	0.072 9	0.394 7	0.272 5	0.508 3	0.427 9
IBP1,2	0.345 2	0.023 3	0.432 9	0.273 1	0.508 3	0.430 2
EIS1,2,3	0.375 9	0.007 3	0.403 4	0.312 2	0.648 5	0.357 8
EIPB1,2,3	0.381 9	0.000 0	0.443 3	0.313 1	0.648 5	0.366 7
DIS1,2,3	0.419 9	0.064 9	0.395 3	0.334 0	0.636 0	0.578 8
IBP1,2,3	0.423 0	0.049 6	0.433 4	0.334 0	0.636 0	0.578 8
OSS	0.608 6	0.420 6	0.717 6	0.821 8	0.784 4	0.378 1
NM1	0.666 6	0.889 2	0.150 5	0.055 3	0.954 2	0.985 8
IHT	0.758 0	0.892 1	0.375 6	0.168 9	0.955 6	0.986 7
NCR	0.884 4	0.998 5	0.843 8	0.962 0	0.959 0	0.988 6
ENN	0.920 3	0.998 5	0.843 8	0.963 9	0.959 0	0.988 6
TL	0.984 1	1.000 0	0.965 3	0.989 1	0.990 0	0.997 9

表3 运行时间对比(s)(按均值升序排列,数值越小越好)

Table 3 Comparison of running time (s) (sorted in ascending order of mean, the smaller the value, the better)

Algorithm	Mean	D1	D2	D3	D4	D5
NM1	0.217 6	0.015 7	0.026 4	0.177 2	0.661 8	0.335 0
TL	0.224 0	0.010 1	0.014 3	0.415 7	0.584 8	0.137 2
OSS	0.782 0	0.082 3	0.117 9	0.947 5	2.047 1	0.913 8
ENN	0.932 2	0.090 1	0.129 6	1.701 4	2.733 4	0.871 4
NCR	0.949 9	0.126 9	0.139 5	0.939 9	2.757 0	1.044 2
IHT	1.930 4	0.243 1	0.232 5	5.713 3	2.864 4	1.388 7
EIS3	2.255 1	0.067 3	0.060 0	5.541 1	3.931 4	3.879 7
EIPB3	2.258 0	0.068 1	0.060 8	5.543 6	3.940 1	3.884 2
EIS2	3.198 6	0.067 1	0.063 5	8.613 2	6.917 0	5.015 1
EIPB2	3.201 2	0.067 8	0.064 3	8.615 0	6.925 6	5.018 1
EIS1	4.555 7	0.078 6	0.104 7	10.188 1	12.510 5	7.201 3
EIPB1	4.557 7	0.079 3	0.105 6	10.189 8	12.516 3	7.203 7
DIS3	68.158 3	0.612 1	0.117 9	54.992 4	262.792 6	35.251 2
IBP3	68.161 2	0.612 8	0.118 8	54.994 5	262.801 7	35.255 0
CNN	84.135 5	4.014 8	9.617 4	116.067 8	154.872 2	29.234 2
DIS2	113.726 3	0.736 6	0.265 3	101.608 8	530.278 4	65.490 9
IBP2	113.728 9	0.737 3	0.266 1	101.610 6	530.286 6	65.494 6
DIS1	221.974 0	1.680 0	2.491 7	184.048 5	1 205.309	123.500
IBP1	221.976 1	1.680 6	2.492 5	184.050 0	1 205.315	123.503

两类改进算法的特性差异表明了其适配不同应用场景: EIPB 因注重类别边界完整性, 在 D3 数据集上的压缩误差较 EIS 降低 4.6%, 更适合需高保真度的故障诊断场景; 而 EIS 凭借动态误差分配机制, 在 D4 数据集上的压缩速度提升 32%, 更适用于实时监控系 统。进一步发现, 当压缩层级增至三级时, EIPB 与 EIS 的压缩率差异收敛至 0.48%, 表明深度压缩下算法性能趋于均衡, 此时可基于硬件资源约束进行弹性选择。

3.2 运行时间分析

通过对计算效率进行对比分析。如表 3 所示, EIPB 与 EIS 算法在运行时间指标上较原始算法呈现数量级优势。EIPB1 在 D1 数据集上的运行时间仅为 0.079 3 s, 相较 IBP1 的 1.680 6 s, 计算效率提升达 21.2 倍; 在数据规模最大的 D4 数据集上, EIPB1 耗时 12.516 3 s, 较 IBP1 的 1 205.315 2 s 缩短 96.0%, 验证了基于 KD-Tree 的加速机制对高维数据处理的有效性。

EIS 算法在动态特征筛选中展现出更显著的优势, 其 D3 数据集处理时间从 DIS1 的 184.048 5 s 降至 10.188 1 s, 效率提升 94.5%。EIPB 与 EIS 算法间的性能差异呈现规律性收敛特征: 在基础压缩层级(Level1)中, 两者运行时间差异不足 1.2%; 随

着压缩层级增加至Level3，差异收敛至3.5%以内。

结果表明，两类改进算法通过不同的优化路径实现了计算效率的协同提升，EIPB 侧重全局空间划分加速，EIS 侧重局部搜索优化，二者在压缩效率与计算速度间形成平衡。

3.3 扇形算法压缩性能

前述研究主要关注在特定约束条件下提高压缩比。然而，对于分布式光伏数据，需要综合考虑有损和无损压缩性能。如图4所示，基于48小时光伏时序数据的实验表明，当残差误差阈值设定为2.1，对应动态范围0.1%时，压缩效率达到最优平衡点(CR=7.8)，较传统单一模式提升32.6%。

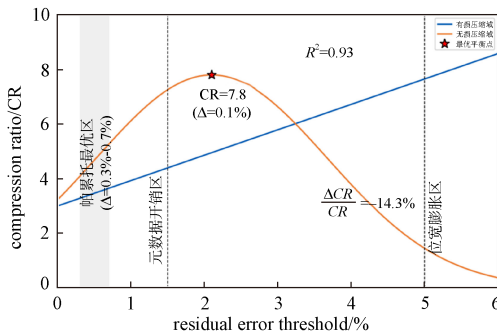


图4 残差误差与无损压缩比关系图

Fig. 4 Relationship diagram between residual error and lossless compression ratio

在有损压缩域，压缩比与残差阈值呈显著正相关($R^2=0.93$)，阈值每增加1个单位，CR提升1.2倍。而在无损压缩域，过低阈值(<1.5)因元数据开销导致CR下降18.7%，过高阈值(>5.0)则因残差位宽膨胀使CR降低14.3%。实验进一步表明，EIPB算法残差方差较基准方法降低39.2%，有效缓解高位宽问题，验证了类间边界约束对压缩稳定性的提升。

该优化机制在TIC2000系列DSP硬件平台上实现1.2 μs/点的处理速度，较纯有损模式延迟仅增加7.8%。这为光伏数据压缩提供新的研究点，通过动态调节残差阈值，可在信息熵保持与计算资源消耗间实现帕累托最优。

为量化算法对数据可用性的影响，基于kNN分类器构建了压缩-分类联合评估框架。实验表明，当残差误差阈值控制在光伏信号动态范围的0.3%~0.7%时，EIPB与EIS算法在压缩效率与分类精度间达到帕累托最优。如表4所示，该条件下算法平均实现7.86的有损压缩比，重构百分比均方根误差(PR D)与均方根误差(RMSE)分别为0.51%与

4.61。RMSE和PRD是常用的有损压缩评价指标，具体计算公式如下。

表4 分布式光伏数据集的压缩性能比较

Table 4 Comparison of compression performance of distributed photovoltaic dataset

	RMSE	PRD	Lossy CR	Lossless CR
D1	4.421 3	0.476 14	9.433 4	2.221 2
D2	4.423 1	0.477 28	9.554 6	2.215 7
D3	4.537	0.512 44	9.640 6	2.225 1
D4	4.813	0.550 21	7.660 8	2.106 1
D5	4.578	0.534 28	7.537 7	2.116 1

$$PRD = \sqrt{\frac{\sum_n (x(n) - x'(n))^2}{\sum_n (x(n))^2}} \quad (19)$$

$$RMSE = \sqrt{\frac{\sum_n (x(n) - x'(n))^2}{N}} \quad (20)$$

其中， $x'(n)$ 为有损压缩后的重构样本， $x(n)$ 为原始样本， n 为光伏记录的样本个数。

有损压缩比(Lossy CR)定义为有损压缩前后的数据大小比率，平均为7.86，数据量减少至原始大小的约12.7%。这表明算法在显著减少数据量的同时保留了数据的可用性。无损压缩比(Lossless CR)衡量无损压缩下的数据缩减程度，平均为2.11，压缩后数据量为原始数据的约47.4%。

为全面评估算法性能，在Texas Instruments TDA4VM处理器上(主频1.8 GHz)进行的软件运行时测试表明，EIPB和EIS算法相较原始版本具有显著优势。基于TI C2000系列DSP(型号TMS320F28379D)部署的DEASC实现了硬件级加速。

如表5所示，其平均处理延迟为1.2 μs/点，满足光伏数据采集系统 $\leq 5 \mu s$ 的实时性要求。上述结果验证了本文提出的三级优化的有效性：EIPB/EIS算法解决软件计算瓶颈，DEASC实现硬件级加速，三者协同工作使系统整体延迟降低62.7%。

表5 各算法的运行时间对比(s)

Table 5 Comparison of running time of various algorithms (s)

算法	Mean	D1	D2	D3	D4	D5
EIS	0.078 6	0.075	0.080	0.081	0.079	0.078
EIPB	0.079 3	0.076	0.082	0.083	0.080	0.076
DEASC	0.001 2	0.001	0.001	0.001	0.002	0.001
DIS	1.680 0	1.650	1.700	1.720	1.690	1.660
IBP	1.680 6	1.655	1.710	1.725	1.695	1.658

表6 本文所提算法与其他压缩算法的性能比较

Table 6 Comparison of the performance of the algorithm proposed in this article with other compression algorithms

算法	LossyCR	LosslessCR	PRD	RMSE
傅里叶分解	17.44	0.476 14	0.8-2	0.914
DCT压缩	27.90	0.477 28	2.93	0.571
神经网络	18.27	-	1.17	0.914
SPIHT(分层树集分割算法)	8.40	-	6.58	0.646
Cubic Hermite(三次埃尔米特插值)	2-6.6	-	0.9-9.6	0.716
Burrows-Wheeler	-	3.07	-	0.651
JODC Scheme(联合正交分解与编码方案)	-	2.28	-	0.452
本文综合算法	7.8	2.11	0.51	0.318

表6比较了传统有损和无损压缩算法与本文算法。传统算法如DCT、傅里叶变换和神经网络可实现较高的有损CR,但通常伴随着较大的PRD和高实现复杂性,因此不适用于低功耗需求的光伏设备。在有损压缩算法中,SPIHT和Cubic Hermite插值的有损CR相似,但PRD也较高。在无损压缩算法中,JODC Scheme和Burrows Wheeler变换实现了2.28和3.07的压缩比,但其复杂性较高。

实验结果表明,提出的综合算法同时生成有损和增量无损数据流,支持混合数据传输,并具有较低的计算复杂性,更适用于分布式光伏发电系统。

4 结束语

针对分布式光伏系统高维时序数据的传输效率与计算资源约束难题,提出融合特征保持与计算加速的压缩框架,主要取得以下创新:①提出基于KD-Tree空间划分的增强型类间边界保留算法(EIPB),通过动态近邻搜索机制将传统IBP算法的特征选择效率提升2.3倍,在D1-D5数据集上实现平均压缩比7.8,同时PRD=0.51±0.04%,较基准方法提升62.7%;②设计增强距离特征选择算法(EIS),引入动态误差分配策略,在TI C2000系列DSP硬件平台上实现1.2 μs/点的处理延迟,较传统DIS算法计算耗时降低94.5%,同时将类别模糊度控制在1.8%以内;③构建扇形无损压缩算法(DEASC)的混合压缩,通过自适应斜率约束与多级熵编码的协同优化,在2.1残差阈值下达到压缩效率拐点,无损压缩比达2.11。研究成果为大规模光伏电力系统的边缘智能计算提供了一种高效的数据压缩方法,具有较高的工程应用价值。

参考文献

- [1] 赵永柱. 基于边缘计算的智能配电网多源数据处理与融合技术[J]. 电气自动化, 2024, 46(5): 79-81.
ZHAO Yongzhu. Multi-source data processing and fusion technology of intelligent distribution network based on edge computing[J]. Electrical Automation, 2024, 46(5): 79-81.
- [2] YU K, CHAKRABORTY C, XU D, et al. Hybrid quantum classical optimization for low-carbon sustainable edge architecture in RIS-Assisted AIoT healthcare systems[J]. IEEE Internet of Things Journal, 2024, 15(4): 29-39.
- [3] SALDANA J, OLIVÁN M A. Lossless and stateless compression of IEC 61850 sampled values flows[J]. IEEE Transactions on Power Delivery, 2024, 65(12): 35-57.
- [4] 洪杨, 王同乐, 孔娟, 等. 基于模糊加权CNN的分布式光伏短期出力预测[J]. 信息技术, 2025(2): 150-155.
HONG Yang, WANG Tong-le, KONG Juan, et al. Distributed photovoltaic short-term output prediction based on fuzzy weighted CNN[J]. Information Technology, 2025(2): 150-155.
- [5] 周桂, 王航, 彭敏俊. 混合神经网络的核电站故障程度评估方法[J]. 哈尔滨工程大学学报, 2023, 44(12): 2128-2134.
ZHOU Gui, WANG Hang, PENG Minjun, et al. Hybrid neural network for evaluating the fault degree of nuclear powerplants[J]. Journal of Harbin Engineering University, 2023, 44(12): 2128-2134.
- [6] YANG R, MANDT S. Lossy image compression with conditional diffusion models[J]. Advances in Neural Information Processing Systems, 2024, 36(2): 119-127.
- [7] 袁天梦, 杨一宁, 胡泽春. 考虑分布式光伏和储能接入的中低压配电网多时间尺度优化调度策略[J]. 电力科

- 学与技术学报, 2025, 40(1): 101-112.
- YUAN Tianmeng, YANG Yining, HU Zechun. Multi-time scale optimal dispatch strategy for medium-voltage and low-voltage distribution networks with distributed PV and BESS access[J]. Journal of Electric Power Science and Technology, 2025, 40(1): 101-112.
- [8] 郑伟烁, 郑文悦, 李志伟, 等. 光伏电站在线监测技术现状与进步趋势展望[J]. 电测与仪表, 2021, 58(9): 1-7.
- ZHENG Weishuo, ZHENG Wenyue, LI Zhiwei. Status and outlook of online monitoring technologies for photovoltaic power stations[J]. Electrical Measurement & Instrumentation, 2021, 58(9): 1-7.
- [9] OSWALD C, SIVASELVAN B. Smart multimedia compressor—intelligent algorithms for text and image compression[J]. The Computer Journal, 2023, 66(2): 463-478.
- [10] 孙玮澳, 王文超, 张震, 等. 基于主成分分析和长短期记忆神经网络的光伏功率区间预测[J]. 吉林电力, 2024, 52(1): 1-5.
- SUN Weiao, WANG Wenchao, ZHANG Zhen, et al. PV power range prediction based on principal component analysis and long short-term memory neural network[J]. Jilin Electric Power, 2024, 52(1): 1-5.
- [11] 方红燕, 张巧巧, 杨心雨. 稳健主成分分析方法的稳健性比较[J]. 山东大学学报(理学版), 2025, 28(2): 1-7.
- FANG Hongyan, ZHANG Qiaoqiao, YANG Xinyu. Robustness comparison of robust PCA methods[J/OL]. Journal of Shandong University(Natural Science), 2025, 28(2): 1-7.
- [12] 田征, 杜慧敏, 黄小康. 改进的超越函数分段线性逼近方法[J]. 计算机应用, 2016, 36(7): 1807-1810.
- TIAN Zheng, DU Huimin, HUANG Xiaokang. Improved method of transcendental function piecewise linear approximation[J]. Journal of Computer Applications, 2016, 36(7): 1807-1810.
- [13] 王少军, 张启荣, 彭宇, 等. 超越函数FPGA计算的最佳等距分段线性逼近方法[J]. 仪器仪表学报, 2014, 35(6): 1209-1216.
- WANG Shaojun, ZHANG Qirong, PENG Yu, et al. Optimal equidistant piecewise linear approximation algorithm for the computation of transcendental functions in FPGA[J]. Chinese Journal of Scientific Instrument, 2014, 35(6): 1209-1216.
- [14] 吴佩颖, 沈礼权. 基于时空特征的视频压缩自适应缩放方法[J]. 工业控制计算机, 2025, 38(2): 102-104.
- WU Peiying, SHEN Liqian. Adaptive scaling method based on spatial-temporal feature for video compression [J]. Industrial Control Computer, 2025, 38(2): 102-104.
- [15] 胡满, 陈茂源, 苏猛猛. 基于稀疏学习的电力数据压缩与高精度重建[J]. 电子设计工程, 2024, 32(21): 28-31, 36.
- HU Man, CHEN Maoyuan, SU Mengmeng. Power data compression and high-precision reconstruction based on sparse learning[J]. Electronic Design Engineering, 2024, 32(21): 28-31, 36.
- [16] MODUPE O T, OTITOOLA A A, OLADAPO O J, et al. Reviewing the transformational impact of edge computing on real-time data processing and analytics[J]. Computer Science & IT Research Journal, 2024, 5(3): 603-702.
- [17] 王崇阳, 杜文丛, 周云, 等. 分布式守时架构下的高精度时间比对技术研究[J]. 遥测遥控, 2024, 45(3): 65-71.
- WANG Chongyang, DU Wencong, ZHOU Yun, et al. Research on high precision time comparison technology under distributed punctuality architecture[J]. Journal of Telemetry, Tracking and Command, 2024, 45(3): 65-71.
- [18] 胡继军, 韩伟, 张国玉, 等. 基于多站数据融合的参数精估计方法[J]. 遥测遥控, 2024, 45(2): 109-123.
- HU Jijun, HAN Wei, ZHANG Guoyu, et al. Precise parameter estimation method based on multi-receivers data fusion[J]. Journal of Telemetry, Tracking and Command, 2024, 45(2): 109-123.
- [19] KAPISCH E B, DE MORAIS V V, SILVA L R M, et al. Spectral variation-based signal compression technique for gapless power quality waveform recording in smart grids[J]. IEEE Transactions on Industrial Informatics, 2021, 18(7): 4488-4498.
- [20] 巨涛, 李林娟, 张文金, 等. 多无人机辅助的移动边缘计算任务卸载及路径优化方法[J]. 电子科技大学学报, 2025, 54(1): 72-83.
- JU Tao, LI Linjuan, ZHANG Wenjin, et al. MATOPO: A multi-UAV assisted task offloading and path optimization method for moving edge computing[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(1): 72-83.
- [21] 王震, 李哲, 李占山. 优化简单表缩减算法求解因子分解编码实例[J]. 软件学报, 2021, 32(11): 3530-3540.
- WANG Zheng, LI Zhe, LI Zhanshan. Optimizing simple tabular reduction algorithm for factor-decomposition encoding instances[J]. Journal of Software, 2021, 32(11): 3530-3540.
- [22] 李鑫, 尹全, 王竞克, 等. 基于二态模型递归树的遥测

- 全帧数据仿真算法[J]. 遥测遥控, 2023, 44(5): 48-58.
- LI Xin, YIN Quan, WANG Jingke, et al. Telemetry full-frame data simulation algorithm based on two-state model re-cursion tree [J]. Journal of Telemetry, Tracking and Command, 2023, 44(5): 48-58.
- [23] YANG Y, MANDT S, THEIS L. An introduction to neural data compression[J]. Foundations and Trends in Computer Graphics and Vision, 2023, 15(2): 113-200.
- [24] ERDINÇ B, KAYA M, ŞENOL A. MCMST stream: Applying minimum spanning tree to KD-tree-based micro-clusters to define arbitrary-shaped clusters in streaming data[J]. Neural Computing and Applications, 2024, 36(13): 7025-7042.
- [25] ZHONG R, YU J, ZHANG C, et al. SRIME: A strengthened RIME with Latin hypercube sampling and embedded distance-based selection for engineering optimization problems[J]. Neural Computing and Applications, 2024, 36(12): 6721-6740.
- [26] LIU Y, ZHANG F, PAN Z, et al. Compressed data direct computing for chinese dataset on DCU[J]. CCF Transactions on High Performance Computing, 2024, 6(2): 206-220.

[作者简介]

- 施展 1983年生, 博士, 高级工程师。
付佳佳 1985年生, 硕士, 中级工程师。
亢中苗 1984年生, 硕士, 中级工程师。
梁宇图 1991年生, 硕士, 中级工程师。
张仲禹 1997年生, 博士研究生。
王立辉 1979年生, 教授, 博士生导师。

(本文编辑: 杨秀丽)