

文章编号: 2095-4980(2025)07-0692-07

基于多维能耗数据的用能行为聚类分析及数据降维方法

罗帅¹, 王洋², 项添春², 周进¹, 李娜¹, 张来¹

(1. 国网天津市电力公司 经济技术研究院, 天津 300171; 2. 国网天津市电力公司, 天津 300010)

摘要: 长期的能耗数据记录和分析有助于发现能源消耗的趋势和规律, 为制定碳减排策略提供重要参考。重点碳排放监测用户的能耗数据涉及多种类型, 不仅要对用户的能耗数据聚类进行分析, 还需研究更加精确的聚类结果可视化方法。对此, 本文提出一种基于 Tent 混沌序列灰狼优化的模糊 C-均值算法(TGWO-FCM), 对用户的能耗数据进行分析。将用户能耗数据的聚类中心看作灰狼个体进行寻优, 解决 FCM 算法对初始聚类中心位置敏感、容易陷入局部最优的缺点; 采用均匀流形逼近与投影(UMAP)的数据降维方法降低能耗数据的复杂度, 将高维能耗数据映射到二维或三维空间中, 从而实现数据的直观可视化。实验结果表明, 本文方法可将具有相似能耗模式的用户归为同一类别, 不仅揭示了用户间的能耗模式差异, 还为制定针对性的节能减排政策提供了科学依据。

关键词: 混沌序列; 灰狼算法; 聚类; 多维能耗数据; 数据降维

中图分类号: TN914.42

文献标志码: A

DOI: 10.11805/TKYDA2024595

Cluster analysis of energy use behavior based on multidimensional energy consumption data and data dimensionality reduction methods

LUO Shuai¹, WANG Yang², XIANG Tianchun², ZHOU Jin¹, LI Na¹, ZHANG Lai¹

(1. Research Institute of Economy and Technology, State Grid Tianjin Electric Power Company, Tianjin 300171, China;

2. State Grid Tianjin Power Company, Tianjin 300010, China)

Abstract: Long-term energy consumption data recording and analysis help to identify the trends and patterns of energy consumption, providing important references for formulating carbon reduction strategies. The energy consumption data of key carbon emission monitoring users involves various types. It is not only necessary to conduct clustering analysis on the users' energy consumption data, but also to study more precise clustering result visualization methods. To this end, a Fuzzy C-Means algorithm based on Tent Chaotic Sequence Grey Wolf Optimization(TGWO-FCM) is proposed to analyze the users' energy consumption data. The clustering centers of users' energy consumption data are regarded as grey wolf individuals for optimization, which solves the shortcomings of the FCM algorithm being sensitive to the initial clustering center positions and easily falling into local optimum. The data dimensionality reduction method of Uniform Manifold Approximation Projection(UMAP) is adopted to reduce the complexity of energy consumption data, mapping high-dimensional energy consumption data to two-dimensional or three-dimensional spaces to achieve intuitive visualization of the data. Experimental results show that the method proposed in this paper can classify users with similar energy consumption patterns into the same category, not only revealing the differences in energy consumption patterns among

收稿日期: 2024-11-15; 修回日期: 2024-12-06

基金项目: 国网天津市电力公司科技项目资助(经研-研发 2024-02)

引用格式: 罗帅, 王洋, 项添春, 等. 基于多维能耗数据的用能行为聚类分析及数据降维方法[J]. 太赫兹科学与电子信息学报, 2025, 23(7): 692-698. DOI:10.11805/TKYDA2024595.

Citation format: LUO Shuai, WANG Yang, XIANG Tianchun, et al. Cluster analysis of energy use behavior based on multidimensional energy consumption data and data dimensionality reduction methods[J]. Journal of Terahertz Science and Electronic Information Technology, 2025, 23(7):692-698. DOI:10.11805/TKYDA2024595.

users, but also providing a scientific basis for formulating targeted energy-saving and emission-reduction policies.

Keywords: chaotic sequences; grey wolf algorithm; clustering; multidimensional energy data; data downscaling

随着全球气候变化的日益严峻, 碳排放已成为国际社会普遍关注的焦点。为应对气候变化带来的挑战, 各国政府纷纷出台了一系列减排政策和措施^[1]。碳排放监测作为评估减排效果、制定科学减排策略的重要手段, 其重要性日益凸显。碳排放监测旨在通过收集、分析和报告企业或地区的碳排放数据, 为政府、企业和公众提供准确、可靠的碳排放信息^[2-5]。这些信息不仅有助于政府制定更加精准的减排政策, 还能引导企业采取更加有效的减排措施, 推动全社会形成绿色低碳的发展模式。

在应对气候变化和推动绿色发展的背景下, 对碳排放的精准监测与管理变得尤为重要。随着大数据技术的发展, 对重点碳排放监测用户的能耗数据进行深入分析成为可能^[6-7]。文献[8]以某小区居民的智能电表用电功耗数据为研究对象, 结合K-means聚类算法对其进行聚类, 分析居民的用电行为规律。文献[9]结合用户的能耗习惯和行为偏好等因素基于聚类技术设计了一种面向降低用户能耗成本和系统峰均比的需求响应方案, 之后进一步提出了一种基于等簇大小聚类和深度学习的优化方法。文献[10]综合利用油井测试、生产及设备数据, 结合抽油机生产专业知识, 分析了抽油机井能耗的影响因素, 建立了完整的抽油机井能耗指标体系, 并分别运用K均值聚类、K中心点聚类和基于密度的聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)建立了油井的总体能耗聚类模型、地面能耗聚类模型以及井下能耗聚类模型, 优选出油井的最佳聚类结果, 分析了各类油井的能耗特征。然而, 现有的能耗数据分析方法中有的只是关注电力消耗数据, 并没有综合分析用电、用热、用气和用煤等各种能耗数据; 有的虽然考虑了多种能耗数据, 但聚类效果不佳, 且没有探究更加精确的数据降维可视化方法。

针对上述问题, 本文提出一种TGWO-FCM方法对重点碳排放监测用户的能耗数据进行分析, 确定用户的能源使用模式和趋势, 将具有相似能耗模式的用户归为同一类别; 采用UMAP的数据降维方法降低能耗数据的复杂度, 将高维数据映射至二维空间中, 从而实现对聚类后的用户能耗数据的直观可视化。本文的技术方案揭示了用户间的能耗模式差异, 通过分析不同聚类中用户的能耗特点, 政府可制定更加精准有效的节能减排政策, 推动绿色低碳发展。

1 数据采集与预处理

1.1 数据采集

原始能源数据包括用电量数据(由电力公司提供)、用热数据(由城市供热公司提供)、用气数据(由城市燃气公司提供或直接接入天然气)以及特殊行业直接使用煤、焦炭、原油、汽油、煤油、柴油和燃油的数据。与现有仅统计用户用电数据不同, 本文还统计上述数据、用气数据以及使用煤、焦炭、原油、汽油、煤油、柴油和燃油数据。用热数据采用热力表获取, 用气数据采用燃气表计, 用煤、焦炭、原油、汽油、煤油、柴油和燃油数据采用统计的用户用量。而对用电量数据, 为提升计算准确度, 与现有日用电量采集或15 min用电量采集的电能表不同, 加装1 min采集超细颗粒度智能电能计量表, 获取用户的使用情况数据。

1.2 原始空缺能耗数据补全

通过获取的实际数据分析可知, 原始能源数据在大多数情况下都存在数据缺失问题, 缺失数据量从数分钟到数小时不等, 甚至数月。这种情况下, 传统采用间隔前后的数据进行简单的线性插值就无法足够精细化数据修复, 因此本文采用拉格朗日插值补全缺失的用户能耗数据。拉格朗日插值多项式的基本形式为:

$$L_n(x_j) = \sum_{k=0}^n y_k l_k(x_j) = y_j \quad (j=0, 1, \dots, n) \quad (1)$$

式中: y_k 为第 k 个插值基函数所对应的用户具体的能耗; $l_k(x_j)$ 为在统计的用户能耗时间范围内, 时间变量 x 取第 j 个时间的第 k 个插值基函数, 具体形式为:

$$l_k(x_j) = \frac{(x_j - x_0) \cdot \dots \cdot (x_j - x_{k-1})(x_j - x_{k+1}) \cdot \dots \cdot (x_j - x_n)}{(x_k - x_0) \cdot \dots \cdot (x_k - x_{k-1})(x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_n)} \quad (k=0, 1, \dots, n) \quad (2)$$

由以上形式可以看出, 拉格朗日插值法计算缺失的能耗数据时, 需先构建与时间变量个数相当的插值基函

数,然后将第 k 个插值基函数与第 k 个用户能耗数据值相乘,并求该式的和,其结果即为拉格朗日插值法在缺失的能耗数据处的外插值。通过数据外插,可将较小的数据量拉长,实现用户能耗数据缺失值的填充。

由于补全后的各类能耗数据颗粒度不同,为方便分析处理,将颗粒度统一为小时级。其中热力表与燃气表记录每小时消耗的数据,故用热数据与用气数据无需进行处理;煤、焦炭、原油、汽油、煤油、柴油和燃油数据通过系数折算为每小时的消耗量;频率为 1 min 的电力消耗数据则通过加总 1 h 内的 60 个消耗量得到小时级的电力消耗数据。通过以上处理,得到 1 天内每小时的各类能耗数据。

1.3 数据归一化

由于条件不同,用户之间可能在能源使用趋势上相互照应,但在全天的能耗数据上却不尽相同。归一化可对所有用户的完整数据集进行缩放,从而根据其"结构"相似性对能源数据进行聚类,由此得出的图表和比较结果反映了能源使用情况的特点。自适应标准差归一化方法能够根据能耗数据的实际情况,动态地调整归一化的参数,以适应不同种类的能耗数据变化,这种动态调整的能力使该方法在处理复杂多变的能耗数据时更加灵活和有效。相较于传统的归一化方法,自适应标准差归一化方法能够减少信息损失,特别是在处理具有复杂分布特性的能耗数据时,能够更好地保留数据的细节信息,从而有助于揭示各类能耗数据之间的潜在关系,因此采用自适应标准差归一化方法对用户能耗数据进行归一化处理。

根据所获取的用户能耗数据,计算其标准差,并设定一个目标标准差,然后调整用户能耗数据的标准差接近该目标值,得到自适应调整后的标准差 σ_{adaptive} 。使用调整后的标准差对能耗数据进行归一化处理:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma_{\text{adaptive}}} \quad (3)$$

式中: x 为用户的能耗数据点; μ 为能耗数据的平均值。

2 TGWO-FCM 聚类和数据可视化

2.1 FCM 聚类算法

在用户能耗数据中,由于各种因素的影响,数据点之间的边界可能并不清晰。FCM 算法通过引入隶属度的概念,允许一个能耗数据点以不同的隶属度属于多个簇,从而提高了聚类的准确性。通过 FCM 聚类算法,可将能耗数据划分为不同的簇,每个簇代表一种能耗模式,以帮助识别出重点碳排放监测用户主要的能耗变化趋势。通过对不同簇的能耗数据进行分析,可以更精准地监测碳排放情况,为制定有效的碳排放控制策略提供依据。

使用 FCM 算法对重点碳排放监测用户聚类的目标函数值进行计算:

$$J_m(U, C) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2 \quad (4)$$

式中: u 为隶属度; x_i 为重点碳排放监测用户的能源消耗数据; n 为能源消耗数据的种类; c 为每种能耗模式中的聚类中心; m 为模糊指数。

对式(4)求导,得出隶属度和每种能耗模式中能耗数据聚类中心分别为:

$$u_{ij} = \left[\sum_{k=1}^c \frac{\|x_i - c_j\|^{2(m-1)}}{\|x_i - c_k\|^{2(m-1)}} \right]^{-1} \quad (5)$$

$$c_j = \frac{\sum_{i=1}^n (u_{ij})^m x_i}{\sum_{i=1}^n (u_{ij})^m} \quad (6)$$

2.2 灰狼优化算法

在对重点碳排放监测用户的能耗数据进行聚类的 FCM 聚类算法中,引入灰狼优化算法(GWO)可以加速用户能耗数据的聚类中心的收敛过程,减少迭代次数和计算时间,提高算法效率。其中,灰狼的个体位置即为用户能耗数据的最佳聚类中心,按照适应度值进行排序, α 狼作为最优的用户能耗数据聚类中心, β 狼和 δ 狼作为次优的能耗数据聚类中心,剩余的能耗数据聚类中心的候选位置为 ω 狼。针对寻优最佳的用户能耗数据聚类中心问题,假设用户的能耗数据有 N 个,则狼群数量为 N ,逼近最优的能耗数据聚类中心位置的过程,代表灰狼接近

并包围猎物的行为。该过程的数学描述如下:

$$\begin{cases} D = |CX_p(t) - X(t)| \\ X(t+1) = X_p(t) - AD \\ A = 2ar_1 - a \\ C = 2r_2 \end{cases} \quad (7)$$

式中: t 为对当前用户能耗数据进行寻找最佳聚类中心的迭代次数; D 为两种不同适应度值的能耗数据聚类中心的距离, 即灰狼与猎物的距离; X_p 为最优的能耗聚类中心位置, 即猎物的位置向量; X 为能耗聚类中心的候选位置, 即灰狼个体的位置向量; A 和 C 为系数向量; a 的值随迭代次数从 2 线性减到 0; r_1 和 r_2 为 $[0,1]$ 区间上的随机向量。

灰狼群经过包围阶段后, 将进行猎捕阶段, 由 α 狼、 β 狼、 δ 狼引导其他灰狼更新个体位置, 此时也是更新用户的能耗数据聚类中心的过程。数学描述如下:

$$\begin{cases} D_\alpha = |C_1 X_\alpha - X(t)| \\ D_\beta = |C_2 X_\beta - X(t)| \\ D_\delta = |C_3 X_\delta - X(t)| \\ X_1 = X_\alpha - A_1 D_\alpha \\ X_2 = X_\beta - A_2 D_\beta \\ X_3 = X_\delta - A_3 D_\delta \\ X(t+1) = (X_1 + X_2 + X_3) / 3 \end{cases} \quad (8)$$

最后, 灰狼群进入攻击阶段, 最终捕获猎物, 此时获得最佳的用户能耗数据聚类中心。

2.3 引入 Tent 混沌序列的灰狼优化算法

为克服在对重点碳排放监测用户聚类时 FCM 聚类算法高度依赖初始聚类中心、容易陷入局部最优的缺点, 本文引入 Tent 混沌序列优化 FCM 算法初始能耗数据聚类中心的选择, 提出 TGWO 优化算法, 提高算法的全局搜索能力。将得到的最优解作为 FCM 算法初始的能耗数据聚类中心进行重点碳排放监测用户能耗聚类分析, 解决传统 FCM 算法对初始中心敏感及陷入局部最优等问题, 能更好地将具有相似能耗模式的用户归为同一类别。

在灰狼优化 FCM 聚类算法中, 灰狼的初始种群相当于初始的用户能耗数据聚类中心, 而初始能耗数据聚类中心的好坏影响算法的全局收敛速度和解的质量, 多样性高的初始能耗数据聚类中心有利于提高算法的全局搜索能力, 提升重点碳排放监测用户的分类效果。

混沌系统产生的混沌序列能在一定范围内按其自身不重复地遍历所有状态, 因此具有遍历性。利用混沌序列搜索显然比随机序列搜索有更大的优越性, Tent 映射可在 $[0,1]$ 区间产生分布比较均匀的 Tent 混沌序列。因此, 本文利用 Tent 混沌序列初始化用户能耗数据的聚类中心, 得到在搜索空间中分布较均匀的候选能耗聚类中心, 以此提高算法的全局搜索能力。具体的步骤概括如下:

1) 使用 Tent 混沌映射的数学表达式生成混沌序列。Tent 映射是一种简单的非线性动力学系统, 其数学表达式为:

$$\begin{cases} x_{n+1} = 2rx_n, & 0 \leq x_n \leq 0.5 \\ x_{n+1} = 2r(1-x_n), & 0.5 \leq x_n \leq 1 \end{cases} \quad (9)$$

式中: x_n 为第 n 次迭代后的值; $r \in (0, 1]$, 用于控制系统的非线性程度。

2) 将生成的混沌序列转换到灰狼种群搜索空间的范围内, 假设灰狼种群搜索空间的范围为 (x_{\min}, x_{\max}) , 对应的范围具体为用户能耗数据的取值区间。设混沌序列的一个值为 z , 则可通过式(10)将 z 映射到搜索空间内:

$$X' = X_{\min} + z(X_{\max} - X_{\min}) \quad (10)$$

式中 X' 为映射后的灰狼位置, 也为候选能耗聚类中心的位置。

3) 根据转换后的混沌序列初始化用户能耗数据聚类中心的位置, 每个能耗数据聚类中心的位置由混沌序列中的一个或多个值决定。

4) 评估初始化的能耗聚类中心的适应度。定义适应度函数为:

$$f = \frac{1}{N} \sum_{i=1}^N \min_j \|x_i - c_j\|^2 \quad (11)$$

式中： N 为聚类中心所属的能耗数据点的数量； x_i 为第 i 个数据点； c_j 为第 j 个聚类中心； $\|\cdot\|$ 为欧几里得距离。

对于每个候选的能耗聚类中心，根据其位置计算适应度值。适应度值越低，表示候选能耗聚类中心在搜索空间中的表现越好。从而选择性地保留或调整候选能耗聚类中心的位置，以进一步优化多样性和搜索能力。

5) 使用灰狼优化算法进行迭代搜索，找到最优的用户能耗数据聚类中心。在搜索过程中，利用 Tent 混沌序列初始化得到的均匀分布的候选能耗聚类中心，可增加候选的能耗数据聚类中心的多样性。

2.4 基于 UMAP 的大数据降维可视化方法

由于聚类后的用户数据涉及多种能耗，属于高维数据，无法直接将聚类效果可视化，因此采用 UMAP 将用户数据降至二维。UMAP 的核心思想是利用黎曼几何和代数拓扑的理论框架，通过构建高维和低维空间中的相似度矩阵，并最小化它们之间的差异实现聚类后高维能耗数据的降维。

高维空间中，设 $X = \{x_1, x_2, \dots, x_N\}$ 为输入的用户能耗数据集， N 为能耗数据的个数，带有一个度量 $d: X \times X \rightarrow R \geq 0$ 。给定输入超参数 k ，对于每个能耗数据 x_i ，使用最近邻下降算法计算集合 $\{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$ 。

对于每个能耗数据 x_i ，定义 ρ_i 和 σ_i ：

$$\rho_i = \min \left\{ d(x_i, x_{i_j}) \mid 1 \leq j \leq k, d(x_i, x_{i_j}) > 0 \right\} \quad (12)$$

$$\sum_{j=1}^k \exp \left[\frac{-\max(0, d(x_i, x_{i_j}) - \rho_i)}{\sigma_i} \right] = \log_2 k \quad (13)$$

针对用户的能耗数据定义一个加权有向图 $G = (V, E, w)$ ， G 的顶点 V 为用户的能耗数据集 X ，则可以形成能耗数据有向边的集合 $E = \{(x_i, x_{i_j}) \mid 1 \leq j \leq k, 1 \leq i \leq N\}$ 。权重函数 w 可定义为：

$$w[(x_i, x_{i_j})] = \exp \left[\frac{-\max(0, d(x_i, x_{i_j}) - \rho_i)}{\sigma_i} \right] \quad (14)$$

设 A 为用户的能耗数据加权有向图 G 的加权邻接矩阵，考虑对称矩阵。

$$B = A + A^T - A \circ A^T \quad (15)$$

式中“ \circ ”表示对用户的能耗数据加权有向图的加权邻接矩阵进行逐点乘积。

然后，UMAP 通过对称从用户能耗数据的有向图 G 中定义一个无向加权图 G ，其邻接矩阵由 B 给出。

低维空间中，UMAP 使用一种力导向图布局算法，利用引力和排斥力演化出一个等价的用户能耗数据加权图 H ，它包含一组点 $\{y_i\}_{i=1,2,\dots,N}$ 。 y_i 和 y_j 处的引力和斥力为：

$$\frac{-2ab \|y_i - y_j\|_2^{2(b-1)}}{1 + \|y_i - y_j\|_2^2} w(x_i, x_j) (y_i - y_j) \quad (16)$$

$$\frac{2b}{(\epsilon + \|y_i - y_j\|_2^2)(1 + a \|y_i - y_j\|_2^2 b)} [1 - w(x_i, x_j)] (y_i - y_j) \quad (17)$$

式中： a 和 b 为超参数； ϵ 为一个小数，防止被零除。

UMAP 作为一种先进的非线性降维技术，在降低高维能耗数据中具有显著的优势和广泛的应用前景。通过 UMAP 降维，可提高数据可视化效果，降低计算复杂度，去除噪声和冗余信息，并支持后续的数据分析和建模。

3 实验分析

使用基于日期的方法组织数据，采集 2024-05-01—2024-08-31 时间范围内 53 个用户的原始能耗数据，然后使用空缺值插补、归一化等方法进行原始数据的预处理，处理后的用户能耗数据供后续分析使用。对于一个给定的重点碳排放监测用户，聚类过程从所统计用户数据的日期起止范围内开始，在这组天数里，每个用户都会产生一个平均能源消耗数据，从而将 2024-05-01—2024-08-31 范围内的数据转化为一天 24 h 内的平均能源消耗

数据。根据统计的用户所属的行业类别,通过调查分析各行业的典型能耗模式,基于这些经验和先验知识,估计一个合理的聚类个数,使用TGWO-FCM算法处理用户能耗数据,并基于肘部法则不断调整聚类个数,将重点碳排放监测用户进行分类,最终将用户的能耗分为3类。

使用基于UMAP的降维方法处理3类用户能耗数据,并将UMAP降维处理后的用户数据可视化,可视化的聚类结果如图1~图3所示。图1为第1类的聚类结果,将23个用户归为一类,在一天24 h内这些用户的能耗数据波动不大,其归一化的能耗数据基本稳定在0.3~0.5之间;图2为第2类的聚类结果,将14个用户归为一类,这些用户在白天的能耗高,在13:00左右达到顶峰,其归一化能耗数据可达0.8左右,而在夜晚能耗低,数值为0.2左右;图3为第3类的聚类结果,将16个用户归为一类,这些用户的能耗趋势在24 h内呈现出先减小后增大再减小的趋势,在5:00降至最低,数值为0.2左右,在14:00—20:00之间维持较高能耗,数值为0.5~0.7之间。该聚类算法将具有相似能耗模式的用户归为同一类别,为分析用户能源使用的规律和特征提供研究基础。

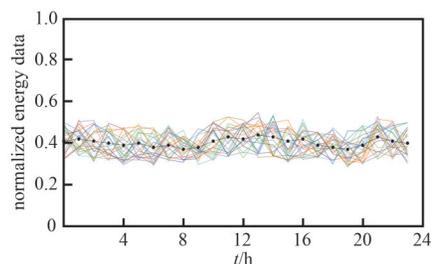


Fig.1 Visualization of clustering results I
图1 第1类可视化的聚类结果

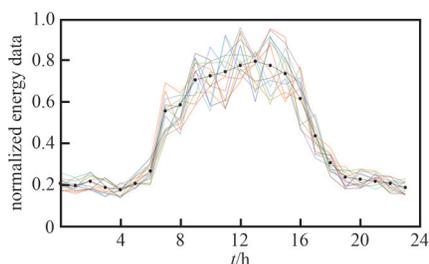


Fig.2 Visualization of clustering results II
图2 第2类可视化的聚类结果

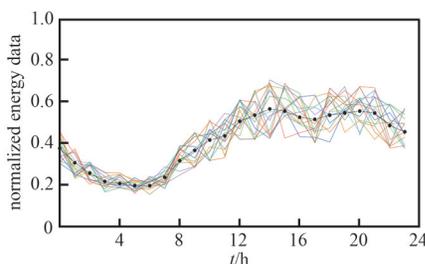


Fig.3 Visualization of clustering results III
图3 第3类可视化的聚类结果

此外,为衡量聚类效果,引入均方根误差(Root Mean Square Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE)评价聚类的准确性。图1~图3中黑色曲线的数据点代表了聚类的中心,根据相同时间下的聚类中心数据点与用户数据点,计算每个用户的RMSE和MAE。二者的值越小,聚类的效果越好,准确度越高。计算结果如图4所示,可以看出,3种聚类结果中每个用户的RMSE大致在0.05~0.08之间,MAE则大致在0.04~0.06之间,聚类效果较好。

4 结论

本文针对现有的方法仅关注电力数据或虽综合多类能耗数据但聚类效果不佳等问题,提出一种融合Tent混沌序列灰狼优化算法的FCM聚类方法对重点碳排放监测用户的能耗数据进行分析。将用户能耗数据的聚类中心看作灰狼个体进行寻优,为算法跳出局部最优提供了机制,提高了算法的局部寻优能力。该算法确定了用户的能源使用模式和趋势,将具有相似能耗模式的用户归为同一类;由于用户数据为包含多种能耗类型的高维数据,采用UMAP的大数据降维方法降低能耗数据的复杂度,在降维过程中能够保留用户能耗数据的局部和全局结构,使降维后的数据在可视化时具有更好的表现。本文的方法不仅揭示了用户间的能耗模式差异,还为制定针对性的节能减排政策提供了科学依据,为重点碳排放监测用户的能电数据时空特性分析研究提供参考。

参考文献:

[1] 招景明,李经儒,潘峰,等. 电力碳排放计量技术现状及展望[J]. 电测与仪表, 2023,60(3):1-8. (ZHAO Jingming, LI Jingru, PAN

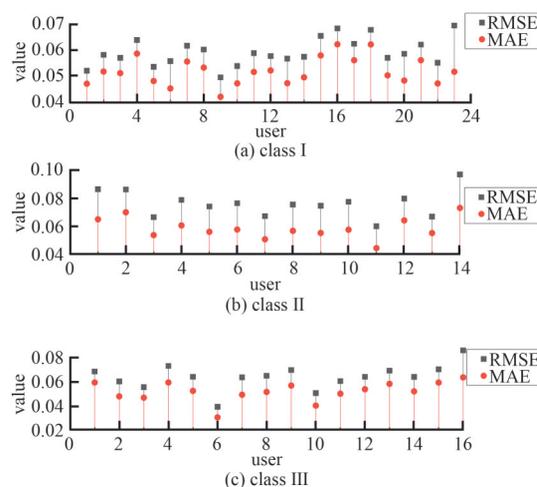


Fig.4 Evaluation of clustering results
图4 聚类结果评价

- Feng, et al. Current status and future prospects of electricity carbon emission measurement technology[J]. *Electrical Measurement & Instrumentation*, 2023,60(3):1-8.) DOI:10.19753/j.issn1001-1390.2023.03.001.
- [2] 宋迎春,杨文昉,周耀鑫. 制造企业碳排放成本核算与披露研究[J]. *财会通讯*, 2020(3):88-91. (SONG Yingchun, YANG Wendie, ZHOU Yaoyi. Research on carbon emission cost accounting and disclosure in manufacturing enterprises[J]. *Communication of Finance and Accounting*, 2020(3):88-91.)
- [3] 张舒涵,陈晖,王彬,等. 基于水泥企业电-碳关系的碳排放监测[J]. *中国环境科学*, 2023,43(7):3787-3795. (ZHANG Shuhan, CHEN Hui, WANG Bin, et al. Carbon emission monitoring based on analysis from "electricity-carbon" relationship of cement enterprises[J]. *China Environmental Science*, 2023,43(7):3787-3795.) DOI:10.3969/j.issn.1000-6923.2023.07.054.
- [4] 叶臻芳,钟志鹏,郑仁广,等. 基于碳电强度的碳排放监测方法[J]. *能源与环境*, 2023(1):40-44. (YE Liufang, ZHONG Zhipeng, ZHENG Renguang, et al. Carbon emission monitoring method based on carbon electrical intensity[J]. *Energy and Environment*, 2023(1):40-44.) DOI:10.3969/j.issn.1672-9064.2023.01.012.
- [5] 李姚旺,刘昱良,杨晓斌,等. 计及电量交易信息的用电碳计量方法[J]. *中国电机工程学报*, 2024,44(2):439-451. (LI Yaowang, LIU Yuliang, YANG Xiaobin, et al. Electricity carbon metering method considering electricity transaction information[J]. *Proceedings of the CSEE*, 2024,44(2):439-451.) DOI:10.13334/j.0258-8013.pcsee.222323.
- [6] YU Bolin, FANG Debin, XIAO Kun, et al. Drivers of renewable energy penetration and its role in power sector's deep decarbonization towards carbon peak[J]. *Renewable and Sustainable Energy Reviews*, 2023(178): 113247. DOI: 10.1016/j.rser.2023.113247.
- [7] LIU Yuan, TAN Qinliang, HAN Jian, et al. Energy-water-carbon nexus optimization for the path of achieving carbon emission peak in China considering multiple uncertainties: a case study in Inner Mongolia[J]. *Energies*, 2021,14(4):1067. DOI:10.3390/en14041067.
- [8] 赵万旗,孙皓月. 基于K-means聚类的居民用电行为分析[J]. *电脑知识与技术*, 2022,18(7):26-28. (ZHAO Wanqi, SUN Haoyue. Analysis of residential electricity consumption behavior based on K-means clustering[J]. *Computer Knowledge and Technology*, 2022,18(7):26-28.)
- [9] 戴碧坚. 智能电网中基于用户聚类的需求响应方法[D]. 南京:南京航空航天大学, 2020. (DAI Bijian. User clustering based demand response method in smart grid[D]. Nanjing, China: Nanjing University of Aeronautics and Astronautics, 2020.) DOI: 10.27239/d.cnki.gnhhu.2020.000380.
- [10] 王辉萍. 基于抽油机井能耗数据的聚类分析方法的研究及应用[D]. 北京:中国石油大学(北京), 2017. (WANG Huiping. Research and application of cluster analysis method based on pumping well energy consumption data[D]. Beijing: China University of Petroleum(Beijing), 2017.) DOI:CNKI:CDMD:2.1019.808162.

作者简介:

罗 帅(1991-), 男, 博士, 高级工程师, 主要研究方向为能源经济、能源大数据、电力供需. email: shuai-luo@outlook.com.

王 洋(1989-), 男, 硕士, 高级工程师, 主要研究方向为能源大数据、双碳.

项添春(1979-), 男, 硕士, 教授级高级工程师, 主要研究方向为能源大数据、新型电力系统.

周 进(1979-), 女, 硕士, 高级工程师, 主要研究方向为能源发展规划、电力发展规划.

李 娜(1985-), 女, 硕士, 正高级经济师, 主要研究方向为电力经济、能源发展.

张 来(1983-), 男, 硕士, 高级工程师, 主要研究方向为新型电力系统.