

DOI: 10.19783/j.cnki.pspc.190486

## 基于灰狼优化聚类算法的日负荷曲线聚类分析

吴亚雄<sup>1,2</sup>, 高崇<sup>1,2</sup>, 曹华珍<sup>1,2</sup>, 陈吕鹏<sup>3</sup>, 唐俊熙<sup>1,2</sup>, 李浩<sup>1,2</sup>

(1. 广东电网有限责任公司电网规划研究中心, 广东 广州 510030; 2. 广东电网发展研究院有限责任公司, 广东 广州 510030; 3. 苏州华天国科电力科技有限公司, 江苏 苏州 215000)

**摘要:** 针对模糊 C-均值聚类算法(Fuzzy C-Means, FCM)应用于日负荷曲线聚类分析时存在易受初始聚类中心影响, 易收敛于局部最优值以及日负荷曲线的内在特性难以通过距离得到充分反映的问题, 利用日负荷特征值指标对日负荷曲线进行数据降维处理。提出了基于灰狼算法(Grey Wolf Optimizer, GWO)优化的模糊 C-均值聚类算法(GWO-FCM)。该算法利用 GWO 为 FCM 优化初始聚类中心, 结合了 GWO 的全局搜索能力和 FCM 的局部搜索能力。算例结果表明所提方法可有效提高日负荷曲线聚类效果, 算法鲁棒性好。

**关键词:** 日负荷曲线聚类分析; 灰狼优化算法; 模糊 C-均值聚类算法; 数据降维; 算法鲁棒性

### Clustering analysis of daily load curves based on GWO algorithm

WU Yaxiong<sup>1,2</sup>, GAO Chong<sup>1,2</sup>, CAO Huazhen<sup>1,2</sup>, CHEN Lüpeng<sup>3</sup>, TANG Junxi<sup>1,2</sup>, LI Hao<sup>1,2</sup>

(1. Grid Planning & Research Center, Guangdong Power Grid Co., Ltd., CSG, Guangzhou 510030, China;  
2. Guangdong Power Grid Development Research Institute Co., Ltd., Guangzhou 510030, China;  
3. Suzhou Huatian Power Technology Co., Ltd., Suzhou 215000, China)

**Abstract:** Fuzzy C-Means (FCM) is susceptible to the influence of initial clustering centers, easy to converge to local optimum values and the inherent characteristics of daily load curve can not be fully reflected through distance when applied to clustering analysis of daily load curve. In order to solve the above problems, the dimensionality of daily load curve is reduced by using the characteristic value index of daily load. A fuzzy C-means clustering algorithm based on Grey Wolf Optimizer (GWO) optimization is proposed. The algorithm uses GWO to optimize the initial clustering center for FCM, and combines the global search ability of GWO and the local search ability of FCM. The results show that the proposed method can effectively improve the clustering effect of daily load curve and has good robustness.

This work is supported by National Natural Science Foundation of China (No. 51777078) and Science and Technology Project of China Southern Power Grid Company (No. GDKJXM20172939).

**Key words:** clustering analysis of daily load curves; grey wolf optimizer; fuzzy C-means; data dimension reduction; algorithm robustness

## 0 引言

随着社会经济增长速度加快, 新兴产业蓬勃发展, 电力用户的行业类别也将日益增多。同时, 同种行业的电力用户的典型日负荷曲线特性的差异性也在逐渐增大。传统简单粗糙的以用户报装类别作为负荷分类依据的分析方法已不适合新一代电力系

统和能源互联网的发展<sup>[1-3]</sup>。精细化的日负荷曲线特性分析有助于电网进行实时优化调度<sup>[4]</sup>、制定合理的需求响应管理措施<sup>[5]</sup>以及电网规划层面的业扩报装和配电网改造等业务<sup>[6-8]</sup>。

目前, 较多聚类分析方法已被用来提取电力用户的典型日负荷特性曲线<sup>[9-11]</sup>。如: 模糊C-均值聚类算法(Fuzzy C-Means Algorithm, FCM)<sup>[12]</sup>、K-MEANS算法<sup>[13]</sup>、DBSCAN算法<sup>[14]</sup>以及自组织映射神经网络<sup>[15]</sup>等。其中FCM算法在运行效率、聚类效果以及算法的鲁棒性方面的表现较好。但传统

**基金项目:** 国家自然科学基金项目资助(51777078); 中国南方电网公司科技项目资助(GDKJXM20172939)

FCM算法直接进行日负荷曲线聚类分析往往存在以下三个缺点。

1) 使用原始日负荷曲线数据之间的距离作为相似性判据直接进行聚类分析。这将导致聚类数据样本的特性属性维度较高,进而导致聚类效果较差。此外,负荷曲线的内在特征、负荷曲线轮廓的相似性也被忽略。

2) 传统FCM算法的聚类效果易受初始点位置影响、容易陷入局部最优,导致聚类效果不佳。

3) 传统FCM算法聚类结果中往往存在少量日负荷曲线偏离聚类中心情况,出现较为明显的“误聚类”的情况,不利于精细化规划工作的推进。

针对以上存在问题,本文首先选用负荷特性指标对原始日负荷曲线数据进行数据降维,并提出一种基于灰狼优化算法(Grey Wolf Optimizer, GWO)的模糊均值聚类算法对日负荷曲线进行聚类分析。GWO算法是一种群搜索智能算法,具有全局寻优,快速收敛的特性。所提算法可利用GWO的全局搜索能力为FCM算法快速搜索最优初始聚类中心,降低FCM算法对初始聚类点的敏感程度,使其获得近似全局最优的聚类划分,提高日负荷曲线聚类分析的聚类效果。

## 1 日负荷曲线数据降维

在一般的聚类方法中,数据样本的相似性通常是利用原始样本数据之间的距离进行判断。然而作为一种时间序列数据,日负荷曲线数据受用户类型、天气、分时电价政策等因素影响较大,导致用户数据的用电特性往往无法通过距离充分体现。在高维度的原始样本下,数据间可能表现出不理想的等距离性。日负荷聚类分析不仅需要关注日负荷曲线数据的等距程度,而且还要分析其形态和轮廓的相似性,以把握用户的用电习惯和特性。

负荷特性指标可以较为全面地反映用户的用电特性。目前,电网规划部门在进行负荷特性分析时通常用负荷率、峰谷差率等特性指标对负荷特性进行描述。由于不同用户的负荷特性在峰谷平期差别较大。因此,从不同时间段的角度,可选取负荷率和峰谷差率指标对原始负荷样本数据进行降维,以增强其内部特征的表达。借鉴文献[16]的负荷特性指标选取方法,依据广东电网峰谷平期的划分方案选取5个特性指标进行数据降维。各类指标的定义以及计算方法如表1所示。

利用指标 $I_1$ — $I_5$ 可从全天、峰期、平期以及谷期的时间尺度对负荷用户的日负荷特性进行描述。各指标的侧重角度和物理意义均有不同,可以更为全

面地描述日负荷曲线的内部特征。

表1 日负荷特性指标

时段	指标名称	计算方法
全天	负荷率	$I_1=P_{av}/P_{max}$
	全日峰谷差率	$I_2=(P_{max}-P_{min})/P_{max}$
峰期	峰期负荷率	$I_3=P_{av,peak}/P_{max}$
平期	08:00—14:00	
	17:00—19:00	平期负荷率
	22:00—24:00	$I_4=P_{av,sh}/P_{max}$
谷期	00:00—08:00	谷期负荷率
		$I_5=P_{av,val}/P_{max}$

表1中: $P$ 代表负荷量;下标av、max、min分别代表均值、最大值、最小值;下标peak、val、sh分别代表峰期、谷期、平期。

## 2 基于GWO的模糊均值聚类算法

### 2.1 模糊C-均值聚类算法

FCM是一种基于划分的算法,最终实现划分到同一类的样本之间的相似度尽可能得高,而不同类之间的相似度尽可能得低。FCM是在普通C-均值聚类的基础上加以改进的,引入了模糊集的概念,进而实现柔性的模糊聚类划分。

设待聚类数据集为 $X=\{x_1, x_2, \dots, x_n\}$ ,其中 $x_j=(x_{j1}, x_{j2}, \dots, x_{jm})^T$ 为特征属性向量。每个样本对应应具有 $m$ 个特征属性; $x_{kj}$ 为特征向量 $x_k$ 的第 $j$ 维特征量。FCM算法利用模糊划分矩阵对聚类结果进行表示,即

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1j} & \cdots & u_{1n} \\ \vdots & & \vdots & & \vdots \\ u_{i1} & \cdots & u_{ij} & \cdots & u_{in} \\ \vdots & & \vdots & & \vdots \\ u_{c1} & \cdots & u_{cj} & \cdots & u_{cn} \end{bmatrix} \quad (1)$$

式中: $c$ 为拟分类数; $u_{ij}$ 表示第 $j(j=1,2,\dots,n)$ 个数据样本属于第 $i(i=1,2,\dots,c)$ 类的隶属度,且满足以下条件:对 $\forall_{i,j}, u_{ij} \in [0,1]$ ;对 $\forall_j, \sum_{i=1}^c u_{ij} = 1$ ;对 $\forall_i, 0 < \sum_{j=1}^n u_{ij} < n$ 。

FCM算法的聚类问题求解可表示求聚类目标函数最优值的优化问题,即

$$\min J(U, P) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^g d_{ij}^2 \quad (2)$$

式中: $U$ 为模糊划分矩阵; $P$ 为聚类中心矩阵; $p_i=(x_{i1}, x_{i2}, \dots, x_{im})^T$ ;  $g \in [0,2]$ 为模糊化程度系数; $d_{ij} = \|x_j - p_i\|$ 为样本 $x_j$ 与聚类中心 $p_i$ 的欧氏距离。

针对式(2)所示的优化问题, FCM 算法通过构造拉格朗日函数对其进行求解:

$$F = \sum_{i=1}^c u_{ij}^g d_{ij}^2 + \lambda (\sum_{i=1}^c u_{ij} - 1) \quad (3)$$

式中,  $\lambda$  为拉格朗日因子。根据 KKT 条件可得使式(3)取得最小值的条件为

$$u_{ij} = [\sum_{k=1}^c (d_{ij} / d_{kj})^{2/(g-1)}]^{-1} \quad (4)$$

$$P_i = \sum_{j=1}^n u_{ij}^g x_j / \sum_{j=1}^n u_{ij}^g \quad (5)$$

由于 FCM 算法在迭代计算过程中, 将不断更新模糊划分矩阵  $U$  和聚类中心矩阵  $P$ 。因此, 可以选择其中任一矩阵进行 FCM 算法的初始化, 然后利用式(4)和式(5)迭代, 反复迭代更新矩阵  $U$  和  $P$ , 搜索目标函数的最小值点, 直到算法达到最大迭代次数或其他算法终止条件。最终的模糊划分矩阵  $U$  和聚类中心矩阵  $P$  将作为聚类结果输出。每个样本对各聚类中心的隶属度可以通过模糊划分矩阵  $U$  确定。

## 2.2 灰狼优化算法

GWO 算法是 Mirjalili S 等人提出的群搜索智能算法。该算法将每个解视为狼, 并将最优解视为猎物。通过参照狼群等级、狩猎技术分别对各解进行分级、搜索方向, 从而寻找最优解<sup>[17]</sup>。GWO 算法具有较强的鲁棒性, 具有搜索个体信息反馈机制, 在迭代过程中能自适应调整收敛因子, 进而提高种群的搜索能力。由于 GWO 算法具有良好的搜索能力和收敛速度, 被广泛应用于工程之中。

### 1) 狼群等级分级

在 GWO 算法结构中, 每匹狼具有各自的社会等级。自上而下依次标记为:  $\alpha$ 、 $\beta$ 、 $\delta$ 、 $\omega$ 。其中,  $\alpha$  狼为狼群中最优个体, 其次为  $\beta$  狼和  $\delta$  狼, 剩余个体则为  $\omega$  狼。狼群内等级森严, 逐级下降。捕猎过程由  $\alpha$ 、 $\beta$ 、 $\delta$  狼决定引导完成,  $\omega$  狼服从高级别狼的命令来完成捕猎任务。

### 2) 包围猎物行为

狼群包围猎物过程中,  $\alpha$ 、 $\beta$ 、 $\delta$  狼的位置很大程度影响接下来的移动方向, 狼群包围策略可表示为

$$D = |C \cdot X_p(l) - X(l)| \quad (6)$$

$$X(l+1) = X_p(l) - A \cdot D \quad (7)$$

$$A = 2a \cdot r_1 - a \quad (8)$$

$$C = 2 \cdot r_2 \quad (9)$$

$$a = 2 - l \times \left( \frac{2}{l_{\max}} \right) \quad (10)$$

式中:  $l$  为当前迭代代数;  $l_{\max}$  为最大迭代代数;  $A$  和  $C$  为协调系数矩阵;  $X_p(l)$  为猎物的位置向量;  $X(l)$  和  $X(l+1)$  分别为受  $X_p(l)$  影响前后, 各狼的位置向量, 即解的位置向量;  $a$  收敛因子, 其值在迭代过程中由 2 逐渐下降为 0;  $r_1$  和  $r_2$  是为范围在  $[0,1]$  的随机向量。

### 3) 灰狼进攻行为

在每次迭代中获得的最佳三个解将被保留并分成  $\alpha$ 、 $\beta$ 、 $\delta$  狼, 其他灰狼根据  $\alpha$ 、 $\beta$ 、 $\delta$  狼更改自身位置, 具体表达式为

$$D_\alpha = |C_1 \cdot X_\alpha(l) - X(l)| \quad (11)$$

$$D_\beta = |C_2 \cdot X_\beta(l) - X(l)| \quad (12)$$

$$D_\delta = |C_3 \cdot X_\delta(l) - X(l)| \quad (13)$$

$$X_1 = X_\alpha(l) - A_1 \cdot D_\alpha \quad (14)$$

$$X_2 = X_\beta(l) - A_2 \cdot D_\beta \quad (15)$$

$$X_3 = X_\delta(l) - A_3 \cdot D_\delta \quad (16)$$

$$X(l+1) = (X_1 + X_2 + X_3) / 3 \quad (17)$$

式中,  $X_1$ 、 $X_2$ 、 $X_3$  分别为各狼受  $\alpha$ 、 $\beta$ 、 $\delta$  狼影响后的移动方位, 由式(17)可综合判断出猎物的移动方向, 通过不断迭代搜索, 寻找全局最优解。

图 1 为 GWO 算法位置更新示意图。

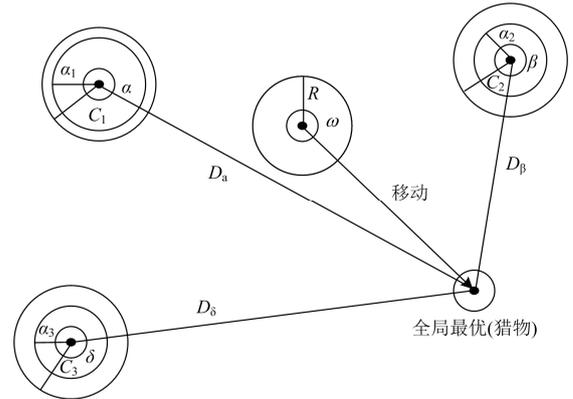


图 1 GWO 算法位置更新示意图

Fig. 1 Diagram of GWO algorithm location updating

## 2.3 GWO-FCM 聚类算法

作为基于梯度下降的搜索算法, FCM 聚类算法具有收敛快、局部搜索能力强的特点。传统 FCM 算法由初始聚类中心矩阵或模糊划分矩阵开始进行迭代计算, 由于初始矩阵是随机生成的, 若初始值选取不同, 则聚类结果也可能不同。因此若初始值选取不恰当, 则聚类结果有可能会陷入局部最优。此问题在样本数量较大的时候显得尤为突出。而 GWO 算法具有较强的全局搜索能力, 收敛速度快, 且不易陷入局部最优等优点。因此, 考虑将 GWO

算法和 FCM 算法进行结合, 充分利用 GWO 算法全局寻优和 FCM 算法局部收敛性强的优势, 降低 FCM 算法对初始值选择的依赖程度, 进而获取更好的聚类效果。

### 2.3.1 灰狼粒子的编码

从上述分析中可知, FCM 算法可选择对模糊划分矩阵  $U$  或聚类中心矩阵  $P$  进行初始化。针对电力用户日负荷曲线聚类分析问题, 通常需要对大量用户的负荷曲线数据进行处理分析, 这将导致数据样本集的数据条数  $n$  过大。若采用模糊划分矩阵  $U$  中的元素构成灰狼的编码, 则编码维度则达到  $c \times n$  维, 维数过高将影响 GWO 算法寻优的效率。而对原始数据进行了数据降维操作之后, 单条数据的特征量  $m$  往往远小于其数据条数  $n$ 。因此, 本文选择聚类中心矩阵  $P$  的元素结构来作为灰狼粒子的编码, 粒子维度可降为  $c \times m$  维。则粒子编码可表示为

$$P_i = (p_{i1}, \dots, p_{im}, \dots, p_{c1}, \dots, p_{cm}) \quad (18)$$

### 2.3.2 适应度函数

FCM 算法需要构造适合的适应度函数来对每个灰狼粒子的好坏程度评估, 定义的适应度函数如下:

$$f(p_i) = \frac{1}{J(U, P)} \quad (19)$$

对某一灰狼个体, 其对应目标值  $J(U, P)$  越小, 适应度值  $f(p_i)$  越高, 表示其聚类效果更优。

狼群中每个灰狼的位置均代表解问题空间中的一个可行解, 表示一种初始聚类中心的选取方法, 因此, 在 GWO 算法每次迭代过程中,  $\alpha$ 、 $\beta$ 、 $\delta$  狼的位置信息更新只需根据适应度值具体对应到每一种初始聚类中心即可。

### 2.3.3 GWO 算法与 FCM 算法转换条件

GWO-FCM 算法主要分为两个部分进行: 第一部分利用 GWO 算法在可行域范围内全面快速地搜寻较好的初始聚类中心; 第二阶段利用 FCM 算法的局部搜索能力, 在已得到的最优初始聚类中心的基础上, 迭代计算进行局部寻优, 完成聚类分析。

定义狼群适应度方差为

$$\sigma^2 = \frac{1}{z} \sum_{i=1}^z [f(p_i) - f_{avg}]^2 \quad (20)$$

式中:  $z$  为狼群规模;  $f_{avg}$  为所有灰狼个体的平均值。适应度方差  $\sigma^2$  的大小表征灰狼个体的收敛程度。当  $\sigma^2$  的值较小时, 狼群的适应度值的离散程度不高, GWO 算法趋向收敛, 其全局搜索能力开始下降。因此, 当  $\sigma^2$  的值小于设定的阈值  $\zeta$  时, 则 GWO-FCM 算法由第一阶段转为第二阶段, 即利用 FCM 进行

聚类分析, 使得后期收敛更快。

### 2.3.4 GWO-FCM 算法流程

GWO-FCM 算法流程图如图 2 所示。

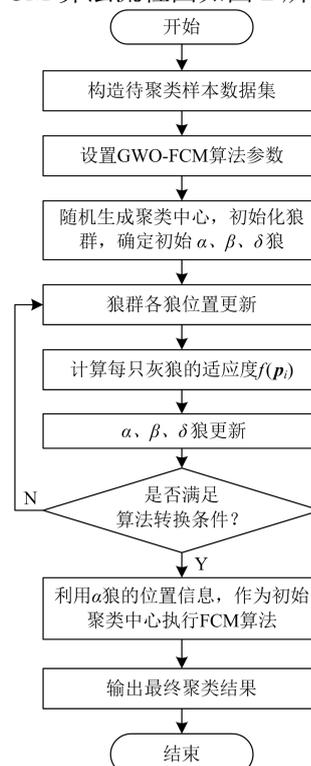


图 2 GWO-FCM 算法流程图

Fig. 2 Flow chart of GWO-FCM algorithm

GWO-FCM 算法描述如下:

- 1) 狼群初始化。设定所需的聚类簇个数  $c$ , 随机生成  $c$  个聚类中心对灰狼位置进行编码, 进而初始化一只灰狼的位置; 重复  $S$  次以获得由  $S$  只初始灰狼组成的初始狼群。然后计算狼群中每只狼的适应度数, 挑选出排名前三位的灰狼个体, 按顺序记为  $\alpha$ 、 $\beta$ 、 $\delta$  狼;
- 2) 狼群位置更新。根据式(6)一式(17)更新狼群中各狼的位置;
- 3) 计算每只灰狼的适应度值  $f(p_i)$ ;
- 4) 狼群等级更新。选择具有最佳适应值的三只灰狼, 并按顺序更新  $\alpha$ 、 $\beta$ 、 $\delta$  狼的位置;
- 5) 转换条件判断。若狼群适应度方差  $\sigma^2$  小于阈值  $\zeta$  或算法循环迭代次数达到最大迭代次数, 则停止 GWO 迭代全局搜索阶段, 并把  $\alpha$  狼位置设为 FCM 算法的初始聚类中心; 否则转步骤 2);
- 6) 采用步骤 5) 输出的结果作为初始聚类中心执行 FCM 算法。输出最终的聚类中心  $P$  以及模糊划分矩阵  $U$ 。依据最大隶属度原则, 输出结果。

### 3 基于 GWO-FCM 算法的日负荷曲线聚类分析

#### 3.1 输入数据集的构建

随着智能电网中智能采集终端配置日益完善,获取海量用户负荷数据的难度逐步降低。但负荷数据中依然存在坏数据及异常数据无法在数据采集过程中进行清洗和剔除。除此之外,还需要构建标准的负荷数据输入样本集作为聚类算法的输入。

1) 数据清洗。首先,剔除异常或缺失数据量较大的负荷曲线数据。然后,可利用平滑公式或其他插值算法对存在少量异常或缺失数据的日负荷曲线数据进行更改或填补。

2) 输入样本集的构造。利用表 1 中所列的日负荷特性指标对已完成数据清洗的日负荷曲线数据进行数据降维计算。现有计量自动化系统中负荷数据通常为一天 96 个采样点,采样间隔为 15 min。设已完成步骤 1)后,剩余日负荷数据条数为  $N$ 。则原始数据样本集的维度通过数据降维,从原来的  $N \times 96$  维下降到  $N \times 5$  维。

#### 3.2 聚类有效性评价指标

日负荷曲线聚类分析问题归属于数据挖掘中的无监督学习(Unsupervised learning)的范畴。由于绝大多数聚类方法是基于一种试探性的行为进行的,无论所给定的待聚类数据集和给定的簇的个数是否合理,均能利用聚类算法求得其簇结构。所以,聚类结果是否有效,仍待进一步评价<sup>[18]</sup>。

可利用轮廓系数(Silhouette coefficient)指标来体现类的紧密性以及可分性,该指标在聚类数未知时能确定最优聚类数和评价聚类效果。对于样本  $x$  的轮廓系数特性指标定义如式(21)一式(23)所示。

$$SIL(x) = \frac{D_a(x) - D_b(x)}{\max(D_a(x), D_b(x))} \quad (21)$$

$$D_a(x) = \frac{\sum_{s \in C_k, s \neq x} d(x, s)}{|C_k| - 1} \quad (22)$$

$$D_b(x) = \min_{C_j, 1 \leq j \leq c, j \neq k} \left\{ \frac{\sum_{s \in C_j} d(x, s)}{|C_j|} \right\} \quad (23)$$

式中:  $SIL(x)$ 代表样本  $x$  的轮廓系数;  $x$  代表属于  $C_k$  类的样本;  $D_a(x)$ 和  $D_b(x)$ 分别代表  $x$  与  $C_k$  类内部剩余对象的平均距离和  $x$  与非  $C_k$  类对象距离的最小平均距离。

轮廓系数的变动范围为  $[-1, 1]$ 。当轮廓系数越接近 1,则表明  $x$  所属的  $C_k$  类紧密性和类可分性越好,

聚类效果好。若轮廓系数小于 0,则聚类失效。

采用所有数据样本轮廓系数的均值评价整体聚类的有效程度。其表示如式(24)。

$$SILMEAN = \frac{1}{N} \sum_{i=1}^N SIL(x_i) \quad (24)$$

$SILMEAN$  越大,则整体聚类效果更好。使得  $SILMEAN$  最大的聚类数  $c$  即为最优聚类个数。

#### 3.3 整体算法流程

整体算法包含负荷数据预处理以及确定 GWO-FCM 算法的最佳聚类数。

具体流程描述如下:

- 1) 采用 3.1 节所述方法对原始日负荷曲线数据进行负荷数据预处理;
- 2) 设定和最小聚类数  $c_{\min}$  以及最大聚类数  $c_{\max}$ , GWO-FCM 相关参数;
- 3) 执行 GWO-FCM 算法进行聚类分析;
- 4) 使用轮廓系数评价聚类的有效程度;
- 5) 判断当前  $c$  是否小于最小聚类数  $c_{\min}$ ,若是,转至步骤 6)。若否,  $c=c-1$ , 转至步骤 3);
- 6) 输出使轮廓系数均值最大时,对应的聚类结果以及聚类数。

整体算法流程图如图 3 所示。

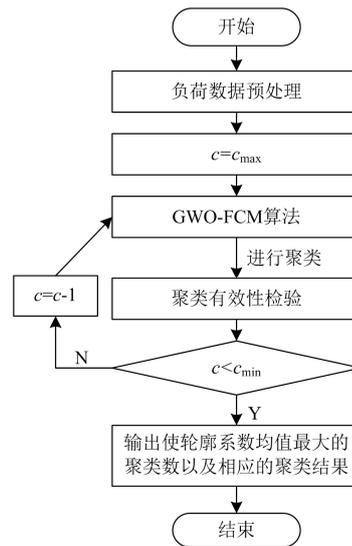


图 3 整体算法流程图

Fig. 3 Flow chart of the whole algorithm

## 4 实验结果与分析

### 4.1 日负荷曲线数据聚类

聚类研究对象选用广东地区某市 2016 年 1 月某正常工作日中 1 083 个用户的实际日负荷曲线数据。数据采集上传间隔为 15 min,每个用户的单日负荷

数据量为 96 个点。采用 3.1 节介绍的输入数据集的构建方法对原始日负荷曲线数据进行预处理。经处理后, 最终算例包含数据条数为 951 条, 待聚类矩阵由 951×96 维下降到 951×5 维。利用所提 GWO-FCM 算法对降维处理后的待聚类矩阵进行聚类分析, 并将聚类结果与利用粒子群优化 FCM 聚类算法(PSO-FCM)<sup>[19]</sup>、萤火虫群优化 FCM 聚类算法(GSO-FCM)<sup>[20]</sup>、蚁群优化 FCM 聚类算法(ACO-FCM)<sup>[21]</sup>以及以 96 个采样点的数据为输入, 经最大值归一化后直接利用传统 FCM 算法的聚类结果作对比。

GWO-FCM 算法和其他群优化 FCM 算法的具体参数设置如下: 最大聚类数  $c_{max}=10$ ; 最小聚类数  $c_{min}=2$ ; 模糊化程度系数  $g=2$ ; 种群规模均取 30; 最大迭代次数取  $t_{max}=200$ 。传统 FCM 算法具体参数如下: 模糊化程度系数  $g=2$ ; 最大迭代次数取  $t_{max}=100$ , 目标函数阈值为  $10^{-7}$ 。

表 2 展示了在设置不同聚类数的情况下, 五种方法的聚类有效性指标计算结果, 图 4 展示了五种方法在不同的聚类数下, 轮廓系数均值的变化趋势。结合表 2 和图 4 可知, 五种算法的 SILMEAN 均值都在聚类数为 6 时达到最大。因此, 五种算法的最优聚类数为均为 6。同时, 采用 GWO-FCM 聚类算法结果的 SILMEAN 均值指标在聚类数取任何其他值时, 均大于传统 FCM 方法得出的 SILMEAN 均值, 聚类效果更优。这也说明了所提方法结合 GWO 算法的全局搜索能力和 FCM 算法的局部搜索能力, 降低聚类结果收敛于局部最优的可能性。

表 2 聚类有效性指标对比

Table 2 Calculation results of clustering effectiveness index

聚类数 $c$	GWO-FCM	GSO-FCM	ACO-FCM	PSO-FCM	传统 FCM
2	0.321 6	0.321 6	0.321 6	0.321 6	0.321 6
3	0.342 4	0.339 1	0.340 5	0.338 4	0.334 7
4	0.359 7	0.350 8	0.352 6	0.351 7	0.347 1
5	0.367 5	0.358 7	0.360 9	0.360 4	0.354 6
6	<b>0.410 8</b>	<b>0.398 7</b>	<b>0.401 6</b>	<b>0.402 4</b>	<b>0.384 2</b>
7	0.381 4	0.379 3	0.378 4	0.376 8	0.368 4
8	0.334 1	0.326 9	0.325 8	0.321 5	0.311 6
9	0.287 6	0.263 2	0.265 8	0.257 4	0.238 1
10	0.225 6	0.214 2	0.216 2	0.215 6	0.184 9

图 5—图 9 显示了五种聚类算法的日负荷曲线聚类结果, 其中红色曲线是该类负荷整体的平均值。由于所提算法具有数据降维的过程, 因此, 只可求出各降维指标的聚类中心, 而无法直接获得各类负荷的典型负荷曲线。本文认为被归为同一类负荷

的日负荷曲线均值为该类负荷的典型日负荷曲线。表 3 对比了不同算法的聚类结果属于每个类别的曲线数。

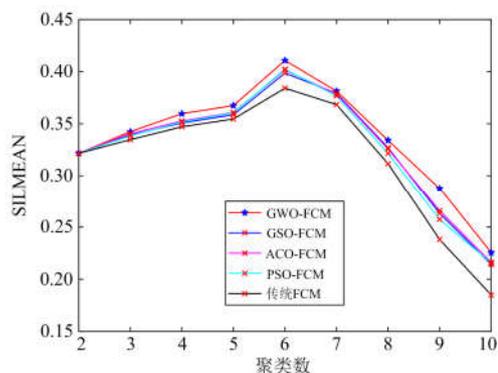


图 4 聚类有效性评价指标变化趋势

Fig. 4 Variation trend of clustering effectiveness index

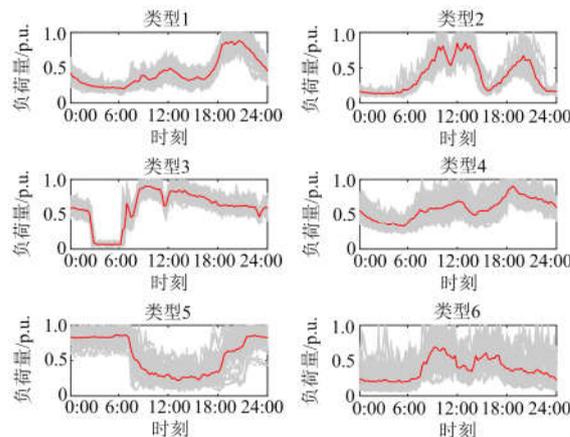


图 5 基于 GWO-FCM 日负荷曲线聚类结果

Fig. 5 Daily load curve clustering result based on the GWO-FCM algorithm

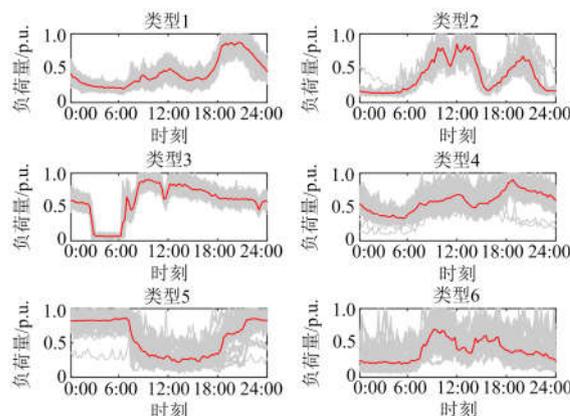


图 6 基于 GSO-FCM 日负荷曲线聚类结果

Fig. 6 Daily load curve clustering result based on the GSO-FCM algorithm

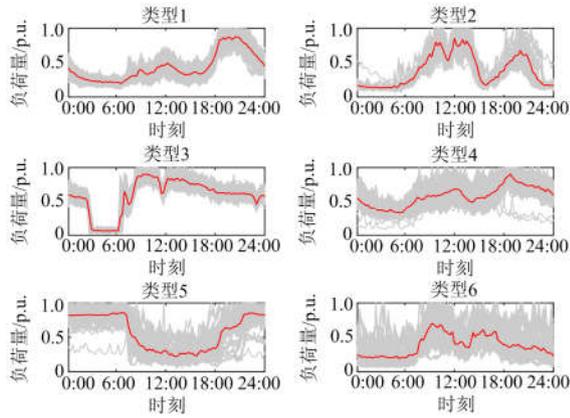


图 7 基于 ACO-FCM 日负荷曲线聚类结果

Fig. 7 Daily load curve clustering result based on the ACO-FCM algorithm

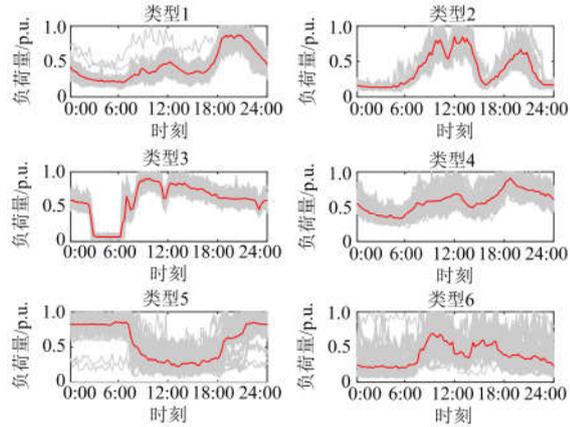


图 8 基于 PSO-FCM 算法日负荷曲线聚类结果

Fig. 8 Daily load curve clustering result based on the PSO-FCM algorithm

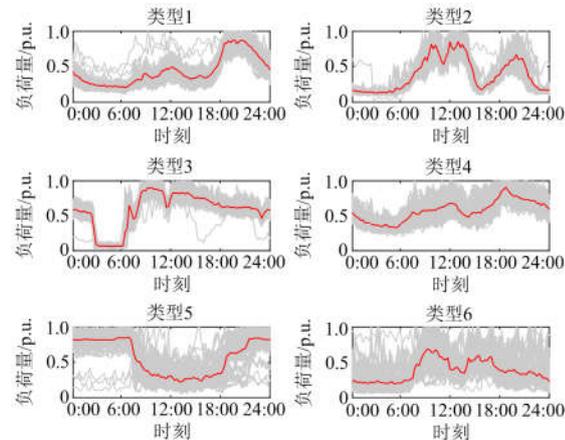


图 9 基于传统 FCM 算法日负荷曲线聚类结果

Fig. 9 Daily load curve clustering result based on the FCM algorithm

表 3 不同算法的聚类曲线数目对比  
Table 3 Comparing the number of clustering curves of different algorithms

算法名称	类型 1	类型 2	类型 3	类型 4	类型 5	类型 6
GWO-FCM	177	99	148	179	171	177
GSO-FCM	176	100	148	181	172	174
ACO-FCM	174	99	149	180	171	178
PSO-FCM	179	99	148	177	173	175
传统 FCM	182	102	145	174	169	179

图 10 为利用五种算法计算所得到的日负荷特征曲线对比, 整体差别不大。且聚类所得到的典型日负荷曲线的负荷特性差异明显, 可满足工程应用需要。但综合图 5—图 9 的结果来看, 对于单条负荷曲线的归类, GWO-FCM 算法的聚类结果中, 归为同一类的负荷曲线之间较为紧密, 而其他试验算法均存在少量日负荷曲线偏离聚类中心(即典型日负荷曲线)较远的情况发生。因此, GWO-FCM 算法聚类效果更优。

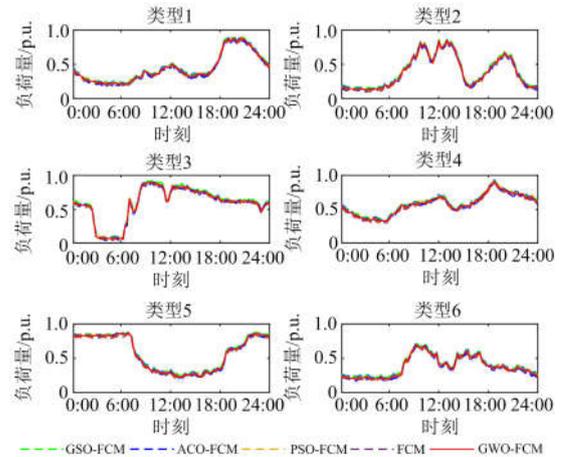


图 10 不同算法提取的典型日负荷曲线比较

Fig. 10 Comparison of typical daily load curves extracted by different algorithms

#### 4.2 算法鲁棒性检验

选用文献[22]中已知的 8 类日负荷曲线, 在已知其正确聚类结果的基础上, 将比例为  $r$  的随机噪声分别添加到每类典型曲线的每个数据点上, 从而获得的分属 8 类、每类 200 条总共 1 600 条的日负荷曲线作为聚类样本, 验证所提聚类算法的鲁棒性。

利用最佳聚类数、聚类有效性指标和准确率三种指标对算法的鲁棒性进行分析。将正确分类的负荷曲线数据总数与所有负荷曲线数目的比值作为准确率。鲁棒性检验结果对比结果如表 4 所示。

从表 4 可以看出, 当增加噪声比  $r$  时, 五种聚

类算法的最佳聚类数均发生改变, 聚类有效性指标以及准确率均有所下降。GWO-FCM 算法在噪声比  $r$  小于 25% 时, 最佳聚类数总是 8, 准确率接近 100%。而对于其他群优化 FCM 算法和传统 FCM 算法, 当噪声比  $r$  分别增加到 20% 和 10% 时, 最佳聚

类数发生改变, 分类准确率持续下降, 鲁棒性较差。当噪声比  $r$  大于 25% 时, 五种算法的最佳聚类数都不是 8, 但 GWO-FCM 的聚类有效性指标和准确率均高于其他四种聚类算法。因此, 所提方法的鲁棒性优于其他群优化 FCM 算法和传统 FCM 算法。

表 4 算法鲁棒性比较

Table 4 Comparison of algorithm robustness

噪声比/%	GWO-FCM			GSO-FCM			ACO-FCM			PSO-FCM			传统 FCM		
	最佳聚类数	聚类有效性指标	准确率/%	最佳聚类数	聚类有效性指标	准确率/%									
5	8	0.843 2	100	8	0.843 2	100	8	0.843 2	100	8	0.843 2	100	8	0.843 2	100
10	8	0.723 2	100	8	0.714 6	99.9	8	0.713 7	99.8	8	0.710 1	99.8	7	0.692 7	87.5
15	8	0.634 1	100	8	0.620 3	99.4	8	0.618 7	99.3	8	0.614 2	99.2	7	0.427 4	87.5
20	8	0.548 4	99.9	7	0.432 8	87.5	7	0.429 6	87.5	7	0.410 6	87.4	7	0.368 8	87.3
25	8	0.456 9	99.8	7	0.381 2	87.4	7	0.378 9	87.3	7	0.371 9	87.2	7	0.364 9	87
30	7	0.386 4	87.1	6	0.363 5	74.5	5	0.358 6	62.4	5	0.351 6	62.3	3	0.307 8	37.5
35	3	0.272 9	37.5	3	0.268 3	37.5	3	0.264 3	37.5	3	0.261 4	37.5	3	0.246 4	37.5
40	3	0.229 5	37.5	3	0.223 4	37.3	3	0.219 8	37.2	3	0.218 4	37.3	3	0.213 1	37.2

## 5 结论

本文提出了一种基于灰狼优化的模糊均值聚类算法的日负荷曲线降维聚类方法。首先选用负荷特性指标对原始日负荷曲线数据进行数据降维, 并采用 GWO-FCM 算法进行日负荷曲线降维聚类分析。算例表明, 该方法可提高传统 FCM 聚类算法的全局搜寻能力, 降低聚类结果陷入局部最优的可能性, 且其聚类效果较其他群优化 FCM 算法也更具优势, 具有工程可行性, 聚类鲁棒性好。

随着智能电网的不断发展, 精细化电网规划和电网运行管理要求不断提高, 如何结合多源数据融合, 进行典型负荷特性指标体系的选取, 建立更加精细化的聚类方法是今后的研究方向。

## 参考文献

- [1] 别朝红, 王旭, 胡源. 能源互联网规划研究综述及展望[J]. 中国电机工程学报, 2017, 37(22): 4-21, 316.  
BIE Zhaohong, WANG Xu, HU Yuan. Review and prospect of planning of energy internet[J]. Proceedings of the CSEE, 2017, 37(22): 4-21, 316.
- [2] 周孝信, 陈树勇, 鲁宗相, 等. 能源转型中我国新一代电力系统的技术特征[J]. 中国电机工程学报, 2018, 38(7): 1893-1904.  
ZHOU Xiaoxin, CHEN Shuyong, LU Zongxiang, et al. Technology features of the new generation power system in china[J]. Proceedings of the CSEE, 2018, 38(7): 1893-1904.
- [3] 李婷, 胥威汀, 刘向龙, 等. 含高比例可再生能源的交

直流混联电网规划技术研究综述[J]. 电力系统保护与控制, 2019, 47(12): 177-187.

LI Ting, XU Weiting, LIU Xianglong, et al. Review on planning technology of AC/DC hybrid system with high proportion of renewable energy[J]. Power System Protection and Control, 2019, 47(12): 177-187.

- [4] 贾艳芳, 易灵芝, 李胜兵. 基于多目标分子动理论的楼宇负荷用电调度优化[J]. 电网技术, 2018, 42(5): 1549-1555.  
JIA Yanfang, YI Lingzhi, LI Shengbing. Optimal scheduling of building load electricity consumption based on multi-objective molecular motion theory[J]. Power System Technology, 2018, 42(5): 1549-1555.
- [5] WANG Jianxiao, ZHONG Haiwang, MA Ziming, et al. Review and prospect of integrated demand response in the multi-energy system[J]. Applied Energy, 2017, 202: 772-782.
- [6] 骆柏锋, 穆云飞, 贾宏杰, 等. 基于负荷特征库的大用户供电接入决策方法[J]. 电力系统自动化, 2018, 42(6): 66-72.  
LUO Bofeng, MU Yunfei, JIA Hongjie, et al. Decision method of power supply access for large consumers based on load feature library[J]. Automation of Electric Power Systems, 2018, 42(6): 66-72.
- [7] 祝锦舟, 张焰, 华月申, 等. 基于马尔科夫链的电力负荷优化组合及应用[J]. 电网技术, 2016, 40(8): 2437-2443.  
ZHU Jinzhou, ZHANG Yan, HUA Yueshen, et al. Power loads optimal combination based on Markov chain and its

- application[J]. Power System Technology, 2016, 40(8): 2437-2443.
- [8] AREFI A, ABEYGUNAWARDANA A, LEDWICH G. A new risk-managed planning of electric distribution network incorporating customer engagement and temporary solutions[J]. IEEE Transactions on Sustainable Energy, 2017, 7(4): 1646-1661.
- [9] AL-OTAIBI R, JIN N, WILCOX T, et al. Feature construction and calibration for clustering daily load curves from smart-meter data[J]. IEEE Transactions on Industrial Informatics, 2017, 12(2): 645-654.
- [10] RHODES J D, COLE W J, UPSHAW C R, et al. Clustering analysis of residential electricity demand profiles[J]. Applied Energy, 2014, 135: 461-471.
- [11] JIANG Zigui, LIN Rongheng, YANG Fangchun, et al. A fused load curve clustering algorithm based on wavelet transform[J]. IEEE Transactions on Industrial Informatics, 2018, 14(5): 1856-1865.
- [12] 刘永光, 孙超亮, 牛贞贞, 等. 改进型模糊 C 均值聚类算法的电力负荷特性分类技术研究[J]. 电测与仪表, 2014, 51(18): 5-9.
- LIU Yongguang, SUN Chaoliang, NIU Zhenzhen, et al. Research on the improved fuzzy c-means clustering algorithm based power load characteristic classification technology[J]. Electrical Measurement & Instrumentation, 2014, 51(18): 5-9.
- [13] 赵文清, 龚亚强. 基于 Kernel K-means 的负荷曲线聚类[J]. 电力自动化设备, 2016, 36(6): 203-207.
- ZHAO Wenqing, GONG Yaqiang. Load curve clustering based on Kernel K-means[J]. Electric Power Automation Equipment, 2016, 36(6): 203-207.
- [14] 王桂兰, 周国亮, 赵洪山, 等. 大规模用电数据流的快速聚类和异常检测技术[J]. 电力系统自动化, 2016, 40(24): 27-33.
- WANG Guilang, ZHOU Guoliang, ZHAO Hongshan, et al. Fast clustering and anomaly detection technique for large-scale power data stream[J]. Automation of Electric Power Systems, 2016, 40(24): 27-33.
- [15] 王文生, 王进, 王科文. SOM 神经网络和 C-均值法在负荷分类中的应用[J]. 电力系统及其自动化学报, 2011, 23(4): 36-39.
- WANG Wensheng, WANG Jin, WANG Kewen. Application of SOM neural network and C-means method in load classification[J]. Proceedings of the CSU-EPSA, 2011, 23(4): 36-39.
- [16] 刘思, 李林芝, 吴浩, 等. 基于特性指标降维的日负荷曲线聚类分析[J]. 电网技术, 2016, 40(3): 797-803.
- LIU Si, LI Linzhi, WU Hao, et al. Cluster analysis of daily load curves using load pattern indexes to reduce dimensions[J]. Power System Technology, 2016, 40(3): 797-803.
- [17] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69(3): 46-61.
- [18] 耿嘉艺, 钱雪忠, 周世兵. 新模糊聚类有效性指标[J]. 计算机应用研究, 2019, 36(4): 1001-1005.
- GENG Jiayi, QIAN Xuezhong, ZHOU Shibing. New fuzzy clustering validity index[J]. Application Research of Computers, 2019, 36(4): 1001-1005.
- [19] WU Ziheng, WU Zhongcheng, ZHANG Jun. An improved FCM algorithm with adaptive weights based on SA-PSO[J]. Neural Computing and Applications, 2017, 28: 3113-3118.
- [20] 胡博, 王昕, 郑益慧, 等. 基于萤火虫优化算法的分布式发电设备容量分配及配电网孤岛划分[J]. 电力系统保护与控制, 2018, 46(13): 21-26.
- HU Bo, WANG Xin, ZHENG Yihui, et al. Calculation of isolated island partition and distributed generator capacity based on firefly algorithm[J]. Power System Protection and Control, 2018, 46(13): 21-26.
- [21] 李泓泽, 郭森, 王宝. 基于遗传改进蚁群聚类算法的电力客户价值评价[J]. 电网技术, 2012, 36(12): 256-261.
- LI Hongze, GUO Sen, WANG Bao. Evaluation on power customer value based on ants colony clustering algorithm optimized by genetic algorithm[J]. Power System Technology, 2012, 36(12): 256-261.
- [22] 黄宇腾, 侯芳, 周勤, 等. 一种面向需求侧管理的用户负荷形态组合分析方法[J]. 电力系统保护与控制, 2013, 41(13): 20-25.
- HUANG Yuteng, HOU Fang, ZHOU Qin, et al. A new combinational electrical load analysis method for demand side management[J]. Power System Protection and Control, 2013, 41(13): 20-25.

收稿日期: 2019-08-11; 修回日期: 2019-09-30

作者简介:

吴亚雄(1990—), 男, 工学硕士, 工程师, 从事电力系统规划与运行研究; E-mail: jxpxlxwyx@163.com

高崇(1983—), 男, 工学硕士, 高级工程师, 从事电力系统规划与运行研究; E-mail: 25903257@qq.com

曹华珍(1974—), 女, 工学硕士, 高级工程师, 从事电力系统规划与运行研究。E-mail: 729875469@qq.com

(编辑 姜新丽)