

基于自适应模糊 C 均值聚类算法的电力负荷特性分类

赵国生¹, 牛贞贞¹, 刘永光², 孙超亮²

(1. 郑州大学 电气工程学院 河南 郑州 450001; 2. 河南许继仪表有限公司 河南 许昌 461000)

摘 要: 针对传统模糊 C 均值聚类算法(FCM)存在的缺点,提出了一种自适应 FCM 算法,该算法以类内距离 MIA 和类间距离 MDC 两个聚类结果评价指标为基础,把 MDC 和 MIA 的比值 I 作为自适应函数来确定 FCM 算法的聚类数目 c ;同时,根据模糊决策的方法,利用 FCM 算法的目标函数和划分熵来共同确定最优的模糊加权指数 m 的取值.结果表明:该算法不仅能够克服 FCM 算法无法自动确定聚类数目和模糊加权指数需要凭经验给出的缺点,而且得到的聚类结果是最优的.通过算例分析也证明了该算法的正确性和有效性.

关键词: 负荷聚类; C 均值聚类算法; 负荷特性; 日负荷曲线

中图分类号: TP18 **文献标志码:** A **doi:** 10.3969/j.issn.1671-6833.2015.06.011

0 引言

基于电力用户的实际负荷曲线分类对电力需求侧管理^[1]有着重要意义.合理的电力负荷分类有助于供电部门有效地掌握用户的负荷特性并制定合理的电价政策^[2];有助于通过削峰填谷手段实现负荷曲线的整形^[3];有助于激励用户积极参与到需求侧管理项目中去.同时,精细化的负荷分类对指导电网滚动规划、实时调度及运行规划的可靠性评估等方面也具有重要意义.根据用电行业、电力负荷用电可靠性等级以及电价等进行分类的传统负荷分类方法已不能满足需求侧管理的需要.基于电力用户的实际负荷曲线分类的电力负荷特性分类方法受到越来越多的关注,成为当前负荷分类的主要方法.

目前,基于实际负荷曲线对负荷进行分类的方法有很多,常用的有 K-means 算法^[4]、层次聚类算法^[5]、模糊 C 均值聚类算法(FCM)^[6]、高斯混合模型(GMM)算法^[7]、自组织特征映射神经网络(SOM)算法^[8]、支持向量机(SVM)算法^[9]和极限学习机(ELM)算法^[10]等.这些方法中,FCM 算法在运行时间、准确度、稳定性及聚类效果等方面均表现较好,是目前应用最广泛的电力负荷特性分类算法之一.但是,FCM 算法也存在着需要人为

确定聚类数目 c 以及模糊加权指数 m 取值需要凭经验给定等问题.针对 FCM 算法存在的缺点,笔者以聚类有效性判别指标^[11]MIA 和 MDC 为基础来自动确定 FCM 算法的聚类数目,并通过模糊决策的方法来确定最优的 m 取值,以达到最优的聚类结果.

1 传统 FCM 聚类算法

1.1 传统 FCM 聚类算法

FCM 算法是一种以隶属度来确定每个数据点属于某个聚类程度的算法,该算法是传统硬聚类(HCM)算法的一种改进.FCM 把数据集 $X = [x_1, x_2, \dots, x_n]$ 分为 c 个模糊组,并求每组的聚类中心,它的模糊划分可用矩阵 $U = [u_{ij}]$ 表示,其中矩阵 U 的元素 u_{ij} 表示第 j ($j = 1, 2, \dots, n$) 个数据点属于第 i ($i = 1, 2, \dots, c$) 类的隶属度. u_{ij} 满足以下条件:

$$\begin{cases} \forall j, \sum_{i=1}^c u_{ij} = 1; \\ \forall i, j, u_{ij} \in [0, 1]; \\ \forall i, \sum_{j=1}^n u_{ij} > 0. \end{cases} \quad (1)$$

FCM 的目标函数是各点的隶属度和该点与聚类中心的欧氏距离的乘积之和,FCM 算法就是

收稿日期: 2015-06-10; 修订日期: 2015-07-19

基金项目: 国家自然科学基金资助项目(51307152)

作者简介: 赵国生(1965—)男,河南南阳人,郑州大学副教授,博士,主要从事电能质量方面的研究, E-mail: zgs410@zzu.edu.cn.

求使聚类目标函数最小化的划分矩阵 U 和聚类中心矩阵 C . 即

$$\min J_m(U, C) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m d_{ij}^2(x_j, c_i). \quad (2)$$

$$d_{ij}(x_j, c_i) = \|x_j - c_i\|. \quad (3)$$

式中: n 是样本数据集的个数; c 是聚类中心数; m 为模糊加权指数; d_{ij} 表示样本点和聚类中心之间的欧氏距离.

使得公式(2)达到最小值的两个必要条件为

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}. \quad (4)$$

$$u_{ij} = \left\{ \sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)} \right\}^{-1}. \quad (5)$$

1.2 FCM 聚类算法的实现

FCM 算法就是反复更新聚类中心矩阵 C 和隶属矩阵 U 进行迭代的过程,具体步骤如下.

step1: 给定聚类数目 c 、模糊加权指数 m 和迭代停止阈值 ε 的值,令 $k = 0$,初始化隶属度矩阵 $U^{(0)}$ 和聚类中心矩阵 $C^{(0)}$.

step2: 根据公式(6)计算更新 $U^{(k)}$.

$\forall i, j$ 若 $d_{ij}^{(k)} > 0$ 则

$$u_{ij}^{(k)} = \left\{ \sum_{r=1}^c \left(\frac{d_{ij}^{(k)}}{d_{rj}^{(k)}} \right)^{2/(m-1)} \right\}^{-1}. \quad (6)$$

若 $d_{ij}^{(k)} = 0$ 则 $u_{ij}^{(k)} = 1$ 且 $j \neq k$ $u_{ij}^{(k)} = 0$.

step3: 根据公式(7)计算更新 $C^{(k+1)}$.

$$c_i^{(k+1)} = \frac{\sum_{j=1}^n (u_{ij}^{(k)})^m x_j}{\sum_{j=1}^n (u_{ij}^{(k)})^m}. \quad (7)$$

step4: 根据公式(2)计算目标函数,如果 $\|C^{(k)} - C^{(k+1)}\| < \varepsilon$ 则停止迭代,输出聚类结果;否则 $k = k + 1$ 返回 step2.

1.3 FCM 聚类算法存在的问题

①FCM 算法不能自动确定聚类数目 c ,需要人工输入一个聚类数目初始值;②FCM 算法的模糊加权指数 m 的取值需要人为给出,不同的参数 m 代表不同的模糊度划分, m 取值的大小直接影响着 FCM 算法中目标函数的收敛性.

2 自适应 FCM 聚类算法

2.1 聚类数目 c 的确定方法

FCM 聚类算法的聚类数目不能自动确定,需要人为输入.为了克服这一缺点,笔者以类内距离

MIA 和类间距离 MDC 两个聚类效果评价指标为基础,引入聚类数目 c 的自适应函数 $I(c)$ 来确定 FCM 算法的聚类数目.理想的负荷分类结果应该是同类负荷曲线间距离最小,不同类负荷曲线间距离最大.类内距离 MIA 和类间距离 MDC 的定义如下:

假设聚类结果得到 c 个负荷分类, C_k 表示第 k 类负荷曲线中所包含的负荷曲线集合; nt_k 表示该集合中的所有负荷曲线; n_k 表示第 k 类负荷曲线中包含的单位数目; CT_k 表示第 k 类负荷曲线的聚类中心,其中 $k = 1, 2, \dots, c$.

类内距离 MIA 表示各聚类中心和其对应聚类中所有负荷曲线数据的距离平均值,定义如下:

$$MIA(c) = \sqrt{\frac{1}{c} \sum_{k=1}^c d^2(CT_k, C_k)}. \quad (8)$$

其中,

$$d(CT_k, C_k) = \sqrt{\frac{1}{n_k} \sum_{n=1}^{n_k} d^2(CT_k, nt_k^n)}. \quad (9)$$

类间距离 MDC 表示不同类的聚类中心负荷曲线间距离的平均值,定义如下:

$$MDC(c) = \text{mean}(d(CT_k, CT_k)). \quad (10)$$

式中: mean 为求平均值的函数.

定义类间距离 MDC 与类内距离 MIA 的比值 I 为

$$I(c) = MDC(c) / MIA(c). \quad (11)$$

由于 $MIA(c)$ 表示类内距离; $MDC(c)$ 表示类间距离;类间距离与类内距离的比值 $I(c)$ 越大表明聚类效果越好,即 $I(c)$ 取得最大值时对应的 c 为最佳聚类数目.

2.2 模糊加权指数 m 的确定

FCM 聚类算法与传统的硬聚类算法的区别就在于引入了模糊加权指数 m ,而参数 m 取值的大小是根据经验确定的.不同的 m 值对 FCM 算法的分类结果有着很大的影响. m 的值越大,聚类结果越模糊,所得到的 FCM 分类结果中每类之间的区别越不明显.也就是说 m 的取值影响着 FCM 算法的聚类效果. m 的取值区间为 $[1, +\infty]$,在不作特殊要求的时候, m 的值取 2. Pal 等人从聚类有效性方面入手,通过实验得出 m 的最优选择区间为 $[1.5, 2.5]$.

选取 m 的最优取值是为了 FCM 聚类算法能够得出合理有效的聚类结果,所以,笔者以 FCM 算法的目标函数为基础,引入 m 的自适应函数.根

据文献[12]中的划分熵 $H_m(U, C)$ 和 FCM 算法的目标函数 $J_m(U, C)$ 共同约束来求得最优的 m 。其中 $H_m(U, C)$ 定义如下:

$$H_m(U, C) = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^c u_{ij} \cdot \ln u_{ij}. \quad (12)$$

其中, 当 $u_{ij} = 0$ 时 $u_{ij} \cdot \ln u_{ij} = 0$; $H_m(U, C)$ 越大, 聚类越模糊。模糊加权指数 m 的最佳值为 m^* , 表示如下:

$$m^* = \arg \min_{\forall m} \left\{ \min(J_m(U^*, C^*)); \right. \\ \left. \text{s. t. } H_m(U^*, C^*) \rightarrow \min. \right\} \quad (13)$$

将 m^* 确定转化为一个带约束的非线性规划问题, 用模糊决策的方法来确定 m^* 的值, 则 $J_m(U, C)$ 和 $H_m(U, C)$ 分别定义成如下形式:

$$\mu_c(m) = \exp \left(-\alpha \cdot \frac{J_m(U^*, C^*)}{\max_{\forall m} (J_m(U^*, C^*))} \right). \quad (14)$$

$$\mu_c(m) = \frac{1}{1 + \beta \left(\frac{H_m(U^*, C^*)}{\max_{\forall m} (H_m(U^*, C^*))} \right)}. \quad (15)$$

其中: 参数 $\alpha > 1$, 一般 $\alpha = 1.5$; β 为较大的正常数, 一般 $\beta = 10$ 。根据式(14)、(15), 利用模糊决策的方法可以得到 m^* 。

$$m^* = \arg \min_{\forall m} (\max(\min(\mu_c(m), \mu_c(m)))) \quad (16)$$

即以较大的隶属度来通过同时极小化 $J_m(U, C)$ 和 $H_m(U, C)$, 由此得到的聚类结果不但分类清晰, 而且类内的元素相似性也比较大。

2.3 自适应 FCM 聚类算法的实现

笔者所提出的自适应 FCM 算法是通过模糊决策的方法自动确定模糊加权指数 m 的值, 并通过类间距离与类内距离的比值形成的自适应函数 $I(c)$ 来自动确定聚类数目。自适应 FCM 算法流程图如图 1 所示。自适应 FCM 聚类算法的具体实现步骤如下。

step1: 输入初始聚类数目 $c = 2$, $I(0) = 0$ 。

step2: 给定 $m = 1.1$ 。

step3: 由 1.2 中的 FCM(m, c) 进行聚类, 求出目标函数 $J_m(U^*, C^*)$ 和划分熵 $H_m(U^*, C^*)$ 。

step4: 判断模糊加权指数 m 的大小, 如果 $m < 2.5$ 则 $m = m + 0.1$ 返回 step3。

step5: 根据公式(12), 利用模糊决策的方法得到最优的模糊加权指数的 m_c^* 。

step6: 由 1.2 中的 FCM(m_c^*, c) 进行聚类, 求出 $I(c)$ 。

step7: 当 $2 < c < n$ 时, 如果 $I(c-1) > I(c-2)$

且 $I(c-1) > I(c)$ 则聚类过程结束; 否则 $c = c + 1$, 返回 step2。

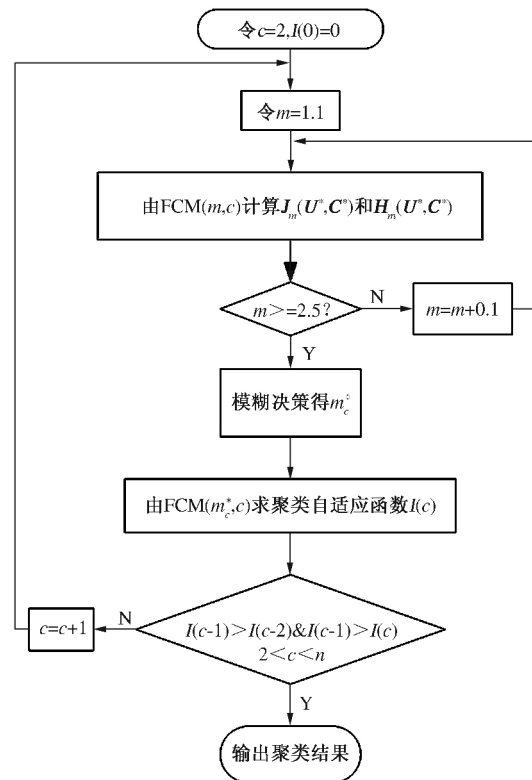


图 1 自适应 FCM 算法流程图

Fig. 1 The flow chart of adaptive FCM algorithm

3 算例分析

3.1 日负荷曲线数据的选取和预处理

算例所采用的数据为北方某省电力公司的用电信息采集系统所采集到的包括商业楼宇、公共机构、高耗电行业等不同行业的 98 个电力用户的 96 点典型日负荷曲线。为了降低其他因素的影响, 笔者采用连续多个工作日的 96 点日负荷曲线数据取平均值的方法来获取典型的日负荷曲线数据。

现取连续一个月的 96 点日负荷曲线数据来获取该类负荷的典型日负荷曲线, 则该负荷的典型日负荷曲线为一个月日负荷曲线数据的平均值。第 i 条负荷曲线的典型负荷曲线表示为 $x_i = [x_{i1}, x_{i2}, \dots, x_{i96}]$, 由于不同类型的负荷数据变化很大, 不便于进行数据比较, 为去除负荷数量级对聚类分析的影响, 在进行聚类分析之前已经对电力用户的典型日负荷曲线数据进行归一化处理, 这里采用极值序列归一化, 即

$$x'_{ij} = \frac{x_{ij}}{\max [x_{i1}, x_{i2}, \dots, x_{i96}]} \quad (16)$$

归一化之后所有的数值都在 $[0, 1]$, 式中

$\max [x_{i1}, x_{i2}, \dots, x_{i96}]$ 为第 i 条典型日负荷曲线数据 x_i 中数据点的最大值。

3.2 自适应 FCM 聚类算法分类结果

根据算例所使用的典型日负荷曲线数据,对其进行极值序列归一化处理后,得到归一化的 98 条典型日负荷曲线数据。按照第 2 节中的自适应 FCM 聚类算法对 98 条负荷曲线进行分类。当聚类数目 $c=8$ 且最优模糊加权指数 $m_c^*=1.9$ 时算法停止,最终的聚类结果如图 2 所示。

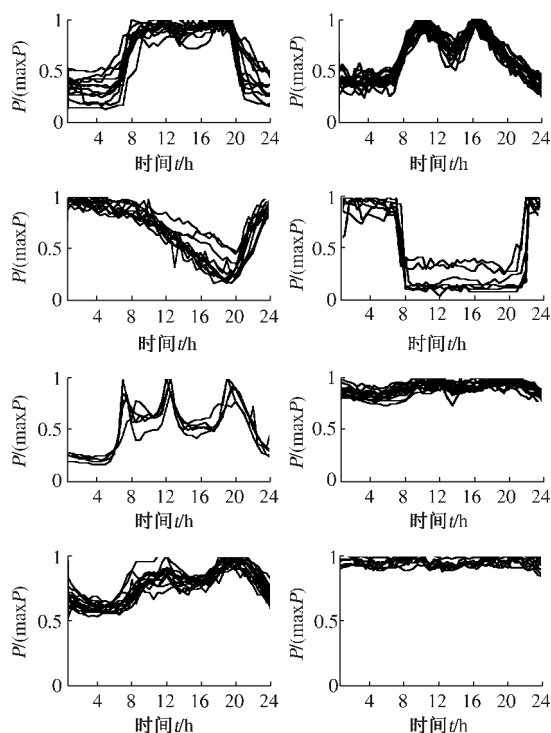


图2 自适应 FCM 算法分类结果
Fig.2 The classification result of adaptive FCM algorithm

由图 2 的最终分类结果可以看出,该分类结果中每个分类能明显地代表一类典型负荷的负荷特征,分类结果比较理想。与传统的 FCM 算法相比,笔者所用自适应 FCM 算法的聚类数目 c 的确定,是根据聚类有效性指标的类间距离与类内聚类的比值来确定的;而模糊加权指数 m 值是以 FCM 算法的目标函数为基础,利用模糊决策的方法确定的。该算法克服了传统 FCM 算法的不足,并且优于那些单纯确定聚类数目 c 的自适应 FCM 算法。

文献[8]是将 SOM 算法与 FCM 算法相结合,利用 SOM 算法得到聚类数目和初始聚类中心,并作为 FCM 算法的初始输入,以克服 FCM 算法存在的缺点,即相当于进行了两次聚类,实现起来比较复杂,且聚类结果的优劣仍需要进行衡量。与文

献[8]相比,笔者所采用的自适应 FCM 算法,是基于聚类结果的评价指标确定自适应函数以确定聚类数目 c ,且以 FCM 算法的目标函数为基础来确定模糊加权指数 m ,均是以聚类结果最优为基础对 FCM 算法本身进行的改进。此方法不仅克服了 FCM 算法存在的缺点,且得到的聚类结果是最优的,实现起来也比较简单。

4 结论

针对 FCM 聚类算法存在需要人为确定聚类数目的缺点,笔者通过计算聚类算法中的有效性判别指标 $MDC(c)$ 和 $MIA(c)$ 的比值 $I(c)$,选取 $I(c)$ 取得最大值时所对应的聚类数目 c 为最优的聚类数目,并且通过模糊决策的方法来优选模糊加权指数 m ,以 FCM 算法的目标函数 $J_m(U^*, C^*)$ 为模糊决策方法的目标函数和划分熵 $H_m(U^*, C^*)$ 为模糊决策方法的约束条件,以较大的隶属度来通过同时极小化 $J_m(U, C)$ 和 $H_m(U, C)$ 以求得最优的 m 。通过算例分析可知,该算法不仅克服了 FCM 算法存在的缺点,而且所得到的聚类结果能够分别代表一类典型负荷的负荷特性。

参考文献:

- [1] 王冬利. 电力需求侧管理实用技术[M]. 北京: 中国电力出版社, 2005: 7-14.
- [2] 黄永皓, 康重庆, 夏清, 等. 用户分类电价决策方法的研究[J]. 中国电力, 2004, 37(1): 24-28.
- [3] 徐明. 基于负荷特性分析的错峰方案研究[D]. 广州: 华南理工大学电力学院, 2012: 17-19.
- [4] 楼佳, 王小华. 一种分列式的 k -means 聚类算法[J]. 杭州电子科技大学学报, 2009, 29(4): 54-57.
- [5] 郭晓娟, 刘晓霞, 李晓玲. 层次聚类算法的改进及分析[J]. 计算机应用与软件, 2008, 25(6): 243-246.
- [6] 周开乐, 杨善林. 基于改进模糊 C 均值算法的电力负荷特性分类[J]. 电力系统保护与控制, 2012, 40(22): 58-62.
- [7] 岳佳, 王士同. 高斯混合模型聚类中 EM 算法及初始化的研究[J]. 微计算机信息, 2006, 11(22): 244-247.
- [8] 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.
- [9] 冯晓蒲. 基于实际负荷曲线的电力用户分类技术研究[D]. 保定: 华北电力大学电气与电子工程学院, 2011: 22-27.
- [10] NAGI J, YAP K S, TIONG S K, et al. Nontechnical loss

- detection for metered customers in power utility using support vector machines [J]. IEEE Transactions on Power Delivery 2010 25(2): 1162 – 1171.
- [11] CHICCO G ,NAPOLI R ,PETAL P. Customer characterization options for improving the tariff offer [J]. IEEE Trans on Power Systems 2003 ,18(1): 381 – 387.
- [12] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms [M]. NewYork: Plenum Press , 1981: 100 – 136.

Power Load Characteristic Classification Technology Research Based on an Optimal Fuzzy C-means Clustering Algorithm

ZHAO Guosheng¹ , NIU Zhenzhen¹ , LIU Yongguang² , SUN Chaoliang²

(1. School of Electrical Engineering , Zhengzhou University , Zhengzhou 450001 ,China; 2. Henan Xu Ji Instrument Co. Ltd , Xuchang 461000 , China)

Abstract: In view of the disadvantages of the traditional Fuzzy C-means clustering algorithm , the author proposes an adaptive FCM algorithm. This algorithm is based on two clustering results evaluation index of within the class distance MIA and between the class distance MDC. The ratio of MDC and MIA , defined as I , is an adaptive function to determine the clustering number c of FCM algorithm. At the same time , according to the fuzzy decision method , we use the objective function and partition entropy of FCM algorithm together to determine the value of optimal fuzzy weighted m . This algorithm not only overcomes the FCM algorithm disadvantage of not being able to determine the clustering number automatically and fuzzy weighted index needs to be given by experience , but also the clustering result is optimal. Finally , the correctness and effectiveness of the algorithm were proved through example analysis.

Key words: load clustering; FCM; load characteristic; daily load curve

(上接第 55 页)

Study on Optimal Allocation Algorithm of the Visible Light Communication LED Array

YE Huiying , WANG Li , LIU Jinliang

(School of Information Engineering Zhengzhou University Zhengzhou 450001 ,China)

Abstract: In view of the limitation of the visible light communication light source layout optimization , this paper presents a universal optimization layout of light source. The simulation results show that , using the light source layout of the algorithm not only can make the average area of the room reaches the maximum spectral efficiency and uniform distribution of illumination. Illuminance level is between 400 lx and 1500 lx which can meet the demand for lighting. Therefore , this algorithm can be used as an optimization algorithm for the layout of the universal light source.

Key words: visible light communication; LED array layout; optimization algorithm; universal