

# 加权模糊 C 均值聚类 and 主客观赋权结合的 厂用电关联特征挖掘方法

秦佳倩<sup>1</sup>, 唐海国<sup>2</sup>, 张 帝<sup>2</sup>, 张志丹<sup>2</sup>, 朱吉然<sup>2</sup>, 马 瑞<sup>1</sup>

(1. 长沙理工大学电气与信息工程学院, 湖南 长沙 410114; 2. 国网湖南省电力有限公司电力科学研究院, 湖南 长沙 410007)

**摘 要:** 厂用电是反应火电机组经济性与能效性的重要指标, 该文提出一种加权模糊 C 均值聚类 and 主客观赋权法结合的厂用电关联特征挖掘方法。首先, 针对火电机组辅机设备的历史运行数据, 利用因子分析法获取运行指标之间的强弱相关性及对厂用电目标值的影响力, 提取影响机组能耗的重要指标, 进而对提取出的运行指标进行对应权重的赋值, 再进行加权模糊 C 均值聚类; 其次, 为克服单一赋权法的缺点, 进一步利用层次分析法与熵权法相结合来修正机组能耗评估中各项指标的权重值; 最后, 通过对某火电厂 600 MW 机组的 10 个月历史数据、6 个典型负荷区间进行文中方法验证, 表明文中模型和算法正确有效。

**关 键 词:** 火电厂; 厂用电; 加权模糊 C 均值聚类算法; 主客观赋权; 关联特征

DOI:10.19781/j.issn.1673-9140.2020.04.016 中图分类号: TM621.7 文章编号: 1673-9140(2020)04-0122-06

## Auxiliary power consumption feature mining method weighted fuzzy C-means clustering and subjective and objective weighting combined

QIN Jiaqian<sup>1</sup>, TANG Haiguo<sup>2</sup>, ZHANG Di<sup>2</sup>, ZHANG Zhidan<sup>2</sup>, ZHU Jiran<sup>2</sup>, MA Rui<sup>1</sup>

(1. School of Electric & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China;

2. Electric Power Research Institute of State Grid Hunan Electric Power Co., Ltd., Changsha 410007, China)

**Abstract:** Power consumption is an important indicator of the economic and energy efficiency of thermal power units, in this paper, an improved weighted fuzzy C-means clustering algorithm is proposed. First of all, according to the historical operation data of auxiliary equipment of thermal power units, the factor analysis method is used to obtain the strong and weak correlation between the operation indicators and the influence on the target value of plant power, and extract the important indicators that affect the energy consumption of the unit, the operating indexes are assigned to the corresponding weights, and the weighted fuzzy C-means clustering is performed. Secondly, to overcome the shortcomings of the single weighting method, the combination of the analytic hierarchy process and the entropy weight method is further used to modify the indicators in the unit energy consumption assessment weight value. Finally, a six-point load interval of 10 months of historical data of a 600 MW unit in a thermal power plant is verified to that the model and algorithm are correct and effective.

**Key words:** thermal power plant; power consumption; weighted fuzzy C-means clustering algorithm; subjective and objective weighting; correlation characteristics

面对日益严峻的能源和环境问题,低碳、经济、安全、高效能源转型是必然途径,但截止目前,在全球电力系统中火电机组仍占主导地位<sup>[1-2]</sup>。因此,火电厂机组节能减排仍然是能源转型主要方面,而通过厂用电关联特征进行提取来降低综合厂用电率<sup>[3-4]</sup>,进行机组厂用电节能分析<sup>[5]</sup>至关重要。

厂用电是火电厂运行的主要能效指标,能准确确定厂用电运行目标值,可以指导机组工作人员调整控制机组在不同负荷及外部条件下的最优运行方式和参数控制<sup>[6]</sup>。文献<sup>[7]</sup>首先介绍了火电厂机组运行能耗分析方法,在得到火电厂机组耗差特性目标值的基础上对火电厂机组能耗进行进一步分析;文献<sup>[8]</sup>对实际某大型火电厂机组各运行参数数据进行研究,考虑锅炉性能而导致经济的损失和主要设备的损坏,给出最佳能源分配计划,增强设备的可靠性;文献<sup>[9]</sup>对火电厂确定机组的关联特征目标值方法进行了分析与研究;文献<sup>[10]</sup>基于现场热力试验结果,确定了机组给定工况下的优化运行目标值。

采用数据挖掘技术和智能优化算法的结合进行运行目标值的确定已得到研究<sup>[11]</sup>。聚类分析法在电力系统分析中得到了广泛的应用<sup>[12-13]</sup>,模糊 C 均值聚类算法<sup>[14]</sup> (fuzzy C-means, FCM) 可以克服传统聚类算法只根据距离因素进行归类的缺点,能考虑到某一个样本点与 2 个聚类中心距离近的情况,而加权模糊 C 均值聚类算法<sup>[15]</sup> (weighted fuzzy C-means algorithm, WFCM) 中隶属度函数进行的聚类更为科学、合理。文献<sup>[16]</sup>结合火电厂机组的运行参数特性,采用数据挖掘技术对火电厂机组实际运行情况的数据进行处理和分析,最后得到火电厂机组运行参数的优化方法;文献<sup>[17]</sup>以某 300 MW 火电厂机组运行实际数据作为研究对象,对其基于模糊关联规则进行挖掘,用来进行优化目标值的确定;文献<sup>[18]</sup>在火电厂机组不同的运行方式之下,采用相应的算法对机组的能耗和排放参数进行数据挖掘与深入分析,最后得到在一定运行状况下机组的能耗排放指标最优的方案;文献<sup>[19]</sup>基于改进主成分分析法对机组能耗特征进行分析。上述文献都是对火电厂机组的运行状况进行数据挖掘,对机组优化运行的能耗目标值进行确定,用以提高火电厂机组的运行效率。关于火电机组厂用电指标与其他指标关联特征挖掘的研究相对较少。

该文主要采用多元统计分析中的聚类分析法,

在此基础上,将 WFCM 用在火电厂厂用电关联特征目标值的挖掘上。根据初始指标间的强弱相关性及所要考虑目标值的影响力,对提取出的初始指标进行对应权重的赋值,再进行聚类。最后,提出利用主客观权重相结合的方法确定机组能耗评估中各项指标的权重值,克服单一赋权法的缺点,权重的确定更为合理。采用实例结果验证该机组厂用电关联特征目标值挖掘的有效性。

## 1 WFCM

目前,已有的 WFCM 是采用不断循环进行迭代计算的方法,用时长、收敛速度慢。因此,该文提出一种改进的 WFCM 对火电厂机组厂用电的关联特征目标值进行挖掘处理,该算法在减少运算循环时间的同时,其他方面也比传统 WFCM 更具有优势。如:可以求取各关联特征指标的权重,挖掘一些难以找出指标的类似属性结构。

定义  $X = (X_1, X_2, \dots, X_n)$  为一个样本空间,其中有  $n$  个指标属性, $n$  个指标中有  $s$  个样本点,即  $X_i = (X_{i1}, X_{i2}, \dots, X_{is})$ ,得到加权模糊 C 均值聚类算法的目标函数:

$$\begin{cases} J_{\text{WFCM}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{i=1}^c \sum_{j=1}^n w_j a_{ij}^m d_{ij}^2 \\ \sum_{j=1}^s w_j = 1 \\ \sum_{i=1}^c a_{ij} = 1, 2 \leq c < n, m > 1 \end{cases} \quad (1)$$

式中  $\mathbf{A}$  为隶属度矩阵; $\mathbf{B}$  为随机选择的聚类中心的集合; $\mathbf{C}$  为权重向量; $w_j$  为第  $j$  个指标属性的权重向量; $a_{ij}$  为样本空间中第  $j$  个样本点聚类在第  $i$  个聚类的隶属度; $m$  为隶属度矩阵的模糊化指数; $d_{ij}$  为欧氏距离。

基于 WFCM 对火电机组厂用电的关联特征目标值进行提取,其中,对每一个指标属性进行权重向量的计算。在一个有很多数据指标属性的样本空间中,很多情况下各样本点的典型程度不能进行具体准确的确定。如果某一样本点的附近有其他聚类的样本点,就会导致此样本点的密度偏大,对最后聚类分析结果产生一定的偏差。定义样本空间中各样本点的密度函数为

$$z_i = \sum_{j=1, j \neq i}^n \frac{1}{D_{ij}^\alpha}, 1 \leq i \leq n \quad (2)$$

$$D_{ij} = \|x_i - x_j\|, 1 \leq i, j \leq n \quad (3)$$

式中  $D_{ij}$  为样本空间某个样本点中  $x_i, x_j$  之间的欧氏距离;  $\alpha$  为一个大于等于 1 的参数。如果该样本点附近的  $x_i$  越多, 那么此样本点的密度函数  $z_i$  也会越大; 同理, 如果该样本点附近的  $x_i$  越少, 那么此样本点的密度函数  $z_i$  就会越小。基于此, 对  $z_i$  样本进行标准化处理, 求出加权矩阵为

$$w_i = \frac{z_i}{\sum_{j=1}^n z_j}, 1 \leq i \leq n \quad (4)$$

由拉格朗日算法求取  $a_{ij}, b_{ij}$ , 即

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

$$b_{ij} = \frac{\sum_{j=1}^n w_j a_{ij}^m x_j}{\sum_{j=1}^n w_j a_{ij}^m} \quad (6)$$

## 2 基于 WFCM 的火电机组厂用电关联目标值模型

该文旨在研究由机组负荷变化而产生影响的火电机组厂用电关联特征目标值。基于 WFCM 的火电机组厂用电关联特征目标值挖掘的主要步骤如下。

1) 对影响火电厂厂用电的指标进行确定, 根据实际负荷运行情况, 分割出  $N$  个典型负荷领域区间。

首先, 对影响发电厂厂用电的机组辅机设备运行数据能否适合使用因子分析法进行检验; 然后, 采用因子分析法的降维作用, 对影响机组运行的辅机设备能耗特性特征进行提取(该文只对影响发电厂厂用电机组运行的辅机设备主要因子进行分析)。该文确定送风机、磨煤机、循环水泵为火电厂机组厂用电的 3 个指标, 进行目标值的挖掘。其中, 送风机主要与锅炉风烟系统密切相关; 磨煤机主要与制粉系统的优化运行相关, 在一定程度上可以减少火电机组的厂用电; 循环水泵主要与循环冷却系统的优

化运行密切相关。该文主要选取与发电厂厂用电率密切相关的运行可控参数, 将这 3 个主要指标作为火电机组厂用电关联特征的主要目标值的挖掘, 建立一个总的样本点, 在一个总的样本空间内针对不一样的样本点进行聚类分析。

2) 在总样本空间里计算各样本点的密度  $z_i$ 。

3) 对聚类分析的参数进行确定。

4) 计算特征权值  $w_i$ 。

5) 计算隶属度矩阵  $A$ 。

6) 计算聚类中心。

7) 对目标函数值进行判断是否小于某个确定的阈值  $\epsilon$ 。若目标函数值小于一迭代阈值, 则算法停止, 否则继续步骤 5, 再次进行模糊隶属度矩阵的计算。

基于 WFCM 的火电机组厂用电关联特征目标值挖掘的具体流程如图 1 所示。

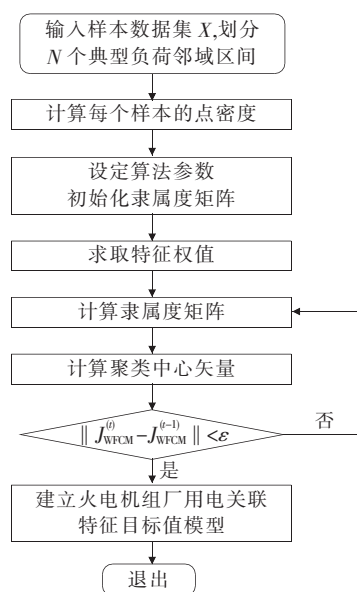


图 1 基于 WFCM 的火电机组厂用电关联特征目标挖掘流程

Figure 1 Flow chart of WFCM-based auxiliary characteristics target of thermal power unit

## 3 基于主客观权重的火电厂机组厂用电节能状态模糊综合评估

模糊综合评估是利用数学模糊关系的原理, 将一些不容易直接评价的不确定性的问题进行定量处理, 以便进行综合性评估。其主要步骤如下。

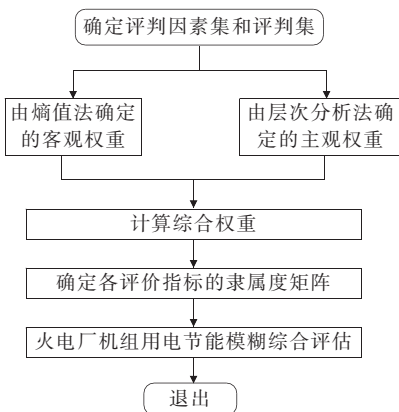
1) 确定评价对象的因素集。 $U = \{u_1, u_2, \dots, u_i, \dots, u_n\}$ , 其中,  $U$  为因素集,  $n$  为评价因素种类数。每一种评价对象的因素集又可以设置附属于这个因素的下一级因素集, 如:  $u_i = \{u_{i1}, u_{i2}, \dots, u_{ik}\}$ 。

2) 确定评价对象的评判集。 $V = \{v_1, v_2, \dots, v_i, \dots, v_m\}$ , 其中,  $V$  为评判集;  $m$  为评价结果数。根据实际情况, 评价结果数一般划分为 3~5 个层次。

- 3) 确定权重向量。
- 4) 确定隶属度函数。
- 5) 确定评判矩阵。
- 6) 计算综合评估结果。

在进行模糊综合评估过程中, 权重的确定对最后模糊综合评估的结果起关键作用, 获取火电厂机组厂用电关联特征评判因素指标后, 该文基于层次分析法(analytic hierarchy process, AHP)的主观赋权法与客观赋权法相结合的方法确定总的综合权重。所得综合权重具有一定的科学性, 避免了传统上由单一主观性确定的权重对结果进行分析; 其次, 在确定评判因素指标隶属度矩阵时采用两级模糊综合评估; 最后, 对火电厂机组厂用电的关联特征利用主客观权重与主观权重所得模糊综合评估得分进行对比。

由主客观赋权法结合确定的综合权重对火电厂机组厂用电关联特征模糊综合评估的流程如图 2 所示。



## 4 算例分析

该文基于 WFCM 对某火电厂 600 MW 机组厂用电关联特征目标值进行挖掘。以机组辅机设备运行状态最佳、厂用电率最低时的运行参数作为辅机设备运行优化的目标值。以某火电厂 600 MW 机组 10 个月的辅机设备运行状态数据及机组运行分析的部分参数为研究对象, 进行数据统计分析, 利用 WFCM 对厂用电关联特征目标值进行挖掘。

首先, 对辅机设备运行状态及机组运行所需要的数据进行筛选整理; 然后, 根据历史数据及实际情况进行负荷预测, 机组在正常运行状态之下的负荷保持为 300~601.5 MW, 选取 300、360、420、480、540、600 MW 为典型负荷, 进行聚类分析。

该文采用 WFCM 确定火电厂机组厂用电关联特征目标值样本及样本分类, 如表 1 所示。在确定的典型负荷中, 聚类数是根据样本数的分类来决定的, 样本所处区间的差别会导致聚类数目不同。由表 1 可知, 该文所确定的 3 个辅机设备的主要指标目标值在不同的典型负荷中有所变化, 根据相关文献的查阅, 表中所表现出来的特点与相关理论知识基本相符, 即送风机、磨煤机、循环水泵这 3 个目标值随着负荷的上升, 所对应的功率也会相应的增加。

样本点在负荷为 360 MW 时的聚类分析结果如表 2 所示, 分析结果与预估结果基本一致, 验证了把 WFCM 应用于机组监控参数基准值的挖掘方法是可行的。

表 1 不同负荷领域区间的样本数分类和目标值样本点的详细结果

Table 1 Detailed results of sample number classification and sample points of target value in different load domain intervals

典型负荷/ MW	样本数	聚类数	送风机/ (MW·h)	磨煤机/ (MW·h)	循环水泵/ (MW·h)	综合厂用电率/%
300	264	3	3 555.34	5 991.18	3 925.01	5.41
360	228	3	3 559.28	5 998.51	3 936.21	5.32
420	207	3	3 563.12	6 009.01	3 949.25	5.15
480	190	3	3 568.92	6 017.19	3 958.98	4.97
540	333	6	3 572.13	6 022.27	3 969.14	4.78
600	347	6	3 574.04	6 029.40	3 978.42	4.68

图 2 基于主客观权重相结合的模糊综合评估流程

Figure 2 Fuzzy comprehensive evaluation process based on the combination of subjective and objective weights

表 2 360 MW 负荷的聚类分析结果

Table 2 Results of clustering analysis for 360 MW load domain

类序号	送风机/ (MW·h)	磨煤机/ (MW·h)	循环水泵/ (MW·h)	综合厂用 电率/%
C <sub>1</sub>	3 563.17	5 999.45	3 937.35	5.34
C <sub>2</sub>	3 566.12	6 000.12	3 938.46	5.36
C <sub>3</sub>	3 561.01	5 995.23	3 935.43	5.28

火电厂机组厂用电主要与火电机组运行的辅机设备运行有关,辅机设备运行指标可在很大程度上体现机组厂用电水平。该文把辅机设备的运行指标作为评判因素集。

#### 1) 客观权重的计算。

首先得出火电厂机组辅机设备评判因素的决策矩阵,据此可以得到机组辅机设备指标的熵(0.45,0.86,0.52)和机组辅机设备的特征向量(0.46,0.14,0.40)。

#### 2) 主观权重的计算。

由专家经验进行打分和机组运行的辅机设备评判因素集的判断矩阵,可以得到机组辅机设备的特征向量(0.40,0.20,0.40)。

#### 3) 综合权重的计算。

采用乘法综合赋权法,综合权重可以确定为(0.49,0.08,0.43)。

基于主客观赋权法确定的综合权重计算得到的模糊综合评估的得分与由 AHP 确定的主观权重计算得到的模糊综合评估得分的对比如图 3 所示,可以看出,2 种方法获取结果对应的分数趋势相同。1—3 月和 4—7 月的机组厂用电水平有所上升,3—4 月和 9—10 月的机组厂用电水平呈现下降趋势。由此,对比这 2 种方法进行模糊综合评估的最终结果,基于主客观结合的赋权方法所确定权重获取的结果更能明显清晰地体现机组厂用电水平的趋势,而由层次分析法所确定的主干权重虽然能从总体趋势上把握机组厂用电水平,但不能获取更精确的变化趋势。总的来说,算例分析结果表明,基于主客观权重的火电厂机组厂用电节能状态模糊综合评估的结果更为合理、可靠。

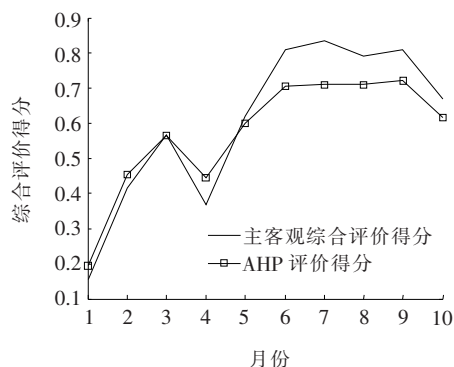


图 3 基于主客观综合权重和 AHP 的模糊综合评估得分对比

Figure 3 Comparison of fuzzy synthesis evaluation scores based on the comprehensive weights and AHP

## 5 结语

该文提出基于改进的模糊 C 均值聚类算法,较为准确地对火电厂机组厂用电关联特征目标值进行挖掘,并提出利用主客观赋权法确定的综合权重进行模糊综合评估。得到以下结论:

1) 火电厂辅机设备中耗电量高的系统主要为烟风、制粉和循环水系统,主要设备有送风机、磨煤机和循环水泵。

2) 重点对 3 个系统和设备进行技术经济改造,可以明显减少火电厂的综合厂用电率,对发电机组厂用电的节能分析具有一定的指导意义。

3) 主客观权重相结合的方法更能真实地反应机组能耗的实际运行水平。

### 参考文献:

- [1] 刘振亚. 中国电力与能源[M]. 北京:中国电力出版社, 2012.
- [2] 刘振亚. 全球能源互联网[M]. 北京:中国电力出版社, 2015.
- [3] 苏鹏,王文君,杨光,等. 提升火电机组灵活性改造技术方案研究[J]. 中国电力,2018,51(5):87-94.  
SUN Peng, WANG Wenjun, YANG Guang, et al. Research on the technology to improve the flexibility of thermal power plants[J]. Electric Power, 2018, 51(5): 87-94.
- [4] 蒋敏华,黄斌. 燃煤发电技术发展展望[J]. 中国电机工程学报,2012,32(29):1-8.  
JIANG Minhua, HUANG Bin. Prospects for the devel-

- opment of coal-fired power generation technology[J]. Proceedings of the CSEE, 2012, 32(29): 1-8.
- [5] 王军,李永华,阎维平.一种燃煤发电节能减排综合评价指数[J].中国电机工程学报,2011,31(12):144-148.  
WANG Jun, LI Yonghua, YAN Weiping. A comprehensive evaluation of energy saving and emission reduction for coal-fired power generation price index[J]. Proceedings of the CSEE, 2011, 31(12): 144-148.
- [6] 王宁玲,杨勇平,杨志平.多变边界条件下火电机组能耗基准状态诊断[J].中国电机工程学报,2013,33(26): 1-7.  
WANG Ningling, YANG Yongping, YANG Zhiping. Energy-consumption benchmark diagnosis of thermal power units under varying operation boundary[J]. Proceedings of the CSEE, 2013, 33(26): 1-7.
- [7] 胡文杰.火电机组耗差分析与软件系统开发[D].北京:华北电力大学,2015.
- [8] Nurkhat Zhakiyev, Yerbol Akhmetbekov, Javier Silvente, et al. Optimal energy dispatch and maintenance of an industrial coal-fired combined heat and power plant in kazakhstan[J]. Energy Procedia, 2017, 12(9): 2485-2490.
- [9] 王玮.火电厂耗差分析优化目标值确定方法的研究[D].保定:华北电力大学,2011.
- [10] 王惠杰,张春发,宋之平.火电机组运行参数能耗敏感性分析[J].中国电机工程学报,2008,28(29):6-10.  
WANG Huijie, ZHANG Chunfa, SONG Zhiping. Sensitive analysis of energy consumption of operating parameters for coal-fired unit[J]. Proceedings of the CSEE, 2008, 28(29): 6-10.
- [11] 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): 58-63.  
ZHOU Kaile, YANG Shanlin. An improved fuzzy C-means algorithm for power load characteristics classification[J]. Power System Protection and Control, 2012, 40(22): 58-63.
- [12] 刘建华,王进,杨洪春,等.基于ACO-PAM综合算法的电力负荷聚类分析[J].电力科学与技术学报,2011, 26(4):94-99.  
LIU Jianhua, WANG Jin, YANG Hongchun, et al. Power load clustering analysis based on ACO-PAM synthesis algorithm[J]. Journal of Electric Power Science and Technology, 2011, 26(4): 94-99.
- [13] 杜明洋,毕大平,王树亮,等.基于模糊聚类的群分离与合并跟踪算法[J].探测与控制学报,2018,40(6):121-127.  
DU Mingyang, BI Daping, WANG Shuliang, et al. Group separation and merge tracking algorithm based on fuzzy clustering[J]. Journal of Detection and Control, 2018, 40(6): 121-127.
- [14] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping[J]. Energy, 2012, 42(1): 68-80.
- [15] 薛艳锋,刘继华,高永强,等.一种加权模糊C均值聚类算法及其在图像分割中的应用[J].计算机应用与软件,2016,33(7):273-277.  
XUE Yanfeng, LIU Jihua, GAO Yongqiang, et al. A weighted fuzzy algorithm C-means clustering algorithm and its application in image segmentation[J]. Computer Application and Software, 2016, 33(7): 273-277.
- [16] 吴雨浓.基于数据挖掘技术的火电机组运行参数优化方法研究[D].北京:华北电力大学,2013.
- [17] 李建强,刘吉臻,张栾英,等.基于数据挖掘的电站运行优化应用研究[J].中国电机工程学报,2006,26(20) 118-123.  
LI Jianqiang, LIU Jizhen, ZHANG Luanying, et al. The research and application of fuzzy association rule mining in power plant operation optimization[J]. Proceeding of the CSEE, 2006, 26(20): 118-123.
- [18] Kusiak A, Song Z. Clustering-based performance optimization of the boiler-turbine system[J]. IEEE Transactions on Energy Conversion, 2008, 23(2): 651-657.
- [19] 马瑞,康仁,罗斌,等.基于改进主成分分析法的火电机组能耗特征识别方法[J].电网技术,2013,37(5): 1196-1201.  
MA Rui, KANG Ren, LUO Bin, et al. An improved principal component analysis based recognition method for energy consumption characteristics of thermal generation unit[J]. Power System Technology, 2013, 37(5): 1196-1201.