

# 采用 M5'模型树和测量数据识别抽汽式 机组汽耗量特性

章坚民<sup>1</sup>, 刘登涛<sup>1</sup>, 吴光中<sup>2</sup>, 张云雷<sup>3</sup>

(1. 杭州电子科技大学, 浙江省 杭州市 310018; 2. 浙江省经济与信息化委员会, 浙江省 杭州市 310008;  
3. 浙江省电力公司, 浙江省 杭州市 310008)

## Working Condition Characteristics Identification for Extraction Unit By Using M5' Model Tree and Measured Data

ZHANG Jianmin<sup>1</sup>, LIU Dengtao<sup>1</sup>, WU Guangzhong<sup>2</sup>, ZHANG Yunlei<sup>3</sup>

(1. Hangzhou Dianzi University, Hangzhou 310018, Zhejiang Province, China;  
2. Zhejiang Economic and Information Technology Commission, Hangzhou 310008, Zhejiang Province, China;  
3. Zhejiang Electric Power Corporation, Hangzhou 310008, Zhejiang Province, China)

**ABSTRACT:** The working condition characteristics of steam turbine unit will be changed due to aging, so routine shut down test in regular intervals is adopted with special equipment and system in costly expense, therefore, the characteristics identification method using the historical measurement records of existed automation and information system becomes necessary. The steam turbine characteristics normally has the property of non-convex and non-continuity, and multivariate linear regression model is not able to deal with. M5' model tree, a data mining method, is good for segmental linearization of single-output multi-output system, was introduced to model the working condition characteristics of extraction unit and to identify the model structure and its parameters. Such model identification method can be used to achieve the latest characteristics by a scroll way using latest measurement data, which can replace the traditional regular characteristics test. Such method has advantages of effective and efficient in implementation, good in approximation, and special property on processing of non-convex and non-continuous function of unit characteristics, and was approved by many CHP stations which shows its good performance and better than the general practiced multi-variables linear model.

**KEY WORDS:** extraction unit; working condition characteristics; M5' model tree; non-convex function; non-continuous function; multivariate linear regression model (MLR)

**摘要:** 汽轮机特性随着机组老化而变化, 传统上采用定期

现场实测, 需停机和采用专门的设备与系统, 费用很高, 因此基于现有自动化系统历史测量数据的特性曲线识别方法十分必要。一般汽轮机汽耗量特性具有非凸和非连续等特点, 常规的多元线性回归拟合不能适应。M5'模型树算法是一种多输入单输出系统的分段线性化的数据挖掘算法。提出采用 M5'模型树的抽汽式机组汽耗量特性模型和其模型结构及参数识别算法, 用于滚动利用最新的电厂测量历史数据获取最新的汽耗量特性。该方法简单、有效, 逼近能力强, 自动化程度高, 在处理非凸形和非连续性的特性方程具有优势。通过多个热电厂的实时数据进行验证, 具有很高的预测精度, 效果优于多元线性回归拟合方程。

**关键词:** 抽汽式机组; 汽耗量特性; M5'模型树; 非凸性函数; 非连续性函数; 多元线性回归模型

## 0 引言

厂网分开, 竞价上网, 以及煤价持续上涨迫使燃煤电厂探索机组的经济运行, 即求解在一定的目标函数和约束条件下的优化问题<sup>[1-9]</sup>。其中最重要的约束条件是机组的汽耗率特性方程。火电厂凝汽机组的汽耗量模型的研究相对较成熟<sup>[5-7]</sup>, 而热电厂中大部分是抽汽式机组, 同时提供热电两种负荷, 其汽耗量特性方程更加复杂、非线性程度更高, 建立该类机组精确的汽耗量预测模型十分困难。

传统的汽耗量预测模型主要采用基于传统热动力学分析方法。该类方法是对机组各子系统进行热动力学分析, 分开考虑供热、供电负荷所需的汽

耗量,如热平衡法<sup>[1]</sup>、等效焓降法<sup>[2]</sup>。这些方法需要大量的机组参数,而机组长时间运行后,机组的物理结构参数可能发生变化,从而偏离机组原出厂特性或经实测的特性参数。另外,方法没有考虑到热电负荷的关联关系,应用受到一定的限制。

热电厂信息化程度越来越高,滚动积累了海量的机组运行数据。利用这些数据,结合数据挖掘算法建立汽轮机组汽耗量模型,可滚动拟合机组汽耗量的特性曲线,以取代传统定期的现场机组特性测试。目前求解汽耗量问题常见的方法是多元线性回归方程(multivariate linear regression model, MLR)拟合数据<sup>[6-9]</sup>,其基础是汽耗量具有凸集合并且连续的曲线特性。文献[6,10-12]指出抽汽式机组汽耗量同时存在凸性与非凸形的特性,此外,由于大多数机组本身一些物理条件的限制,如锅炉压力等,使其不能处于某些出力范围内,即存在禁行区,这导致机组的出力区间有间断现象<sup>[6,13]</sup>,因此,采用多元线性回归方程并不能很好地逼近原非线性的汽耗量曲线。随着节能发电调度特别是智能调度的提出<sup>[6,14]</sup>,将机组的非凸或非连续的耗量曲线近似拟合合成凸连续曲线进而采用传统机组组合(unit commitment, UC)模型求解已不能满足要求<sup>[6]</sup>。

针对这种复杂非线性的特性曲线,最有效的方法是采用分段式线性拟合特性曲线,按给定的精度,自动优选分段数并逼近原非线性耗量曲线,这正是数据挖掘算法——M5'模型树算法<sup>[15-16]</sup>的优势。

## 1 M5' 模型树

### 1.1 算法的发展

Breiman 等人在 1984 年提出了分类回归树(Classification And Regression Tree, CART)<sup>[17]</sup>,将树结构应用到回归预测问题。CART 的输出是各个叶子节点的目标属性的平均值,只是尽可能地逼近回归的树结构。Quinlan 于 1992 年提出分段式多元线性回归树——M5 模型树<sup>[15]</sup>,用于解决预测问题,Wang 等人在 1997 年将 M5 模型树算法进行重构、改进后,形成了 M5'模型树<sup>[16]</sup>。与 CART 的零阶模型相比,M5'模型树是一阶线性模型,构建的树结构更小,预测精度更高,真正意义上实现了树结构的回归预测。

M5'模型树算法的特点:训练规则简单、有效,训练时间短,透明化程度高以及具有较好的泛化能

力。该方法在 20 世纪 90 年代提出后,已经广泛用于解决回归预测问题,如:降雨量预测<sup>[18]</sup>,洪涝灾害预测<sup>[19]</sup>,水位预测<sup>[20]</sup>,含沙量评估<sup>[21]</sup>,海浪高度评估<sup>[22]</sup>,波谱分析<sup>[23]</sup>等等。

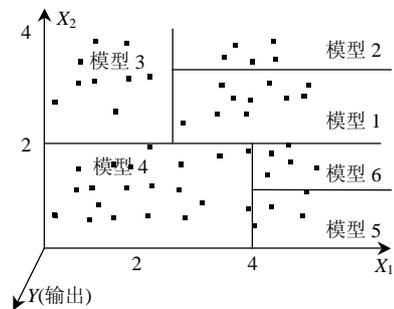
### 1.2 树的生长

M5'模型树将输入与输出参数之间非线性的关系转换为一个分段式线性关系。M5'劈分的标准是样本属性差异化原则(standard deviation reduction, SDR):

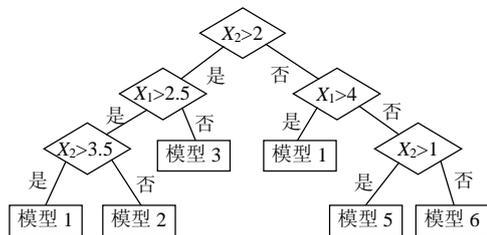
$$S_{DR} = sd(T) - \frac{|T_1|}{|T|}sd(T_1) - \frac{|T_2|}{|T|}sd(T_2) \quad (1)$$

式中:  $T$  为总样本空间,其样本数为  $|T|$ ;  $T$  被劈为 2 个子空间,分别为  $T_1$ 、 $T_2$ ,分别具有样本数为  $|T_1|$ 、 $|T_2|$ ;  $sd(T)$  是总目标属性标准差,  $sd(T_1)$ 、 $sd(T_2)$  表示劈分后 2 个子空间的目标属性标准差。

以图 1 的 2 维输入为例,可以看出其细化模型的空间划分是一个递归空间 2 分化问题:先要制定 2 分化规则,即  $X_1$ 、 $X_2$  平面是先按  $X_2$ ,还是以  $X_2$  进行劈分;其次要决定在何处劈分。一旦这个问题解决,余下的就是被劈分的空间如何再次 2 分化,可以继续按此规则递归,一直到满足停止条件。



(a) 细化模型的空间划分



(b) 按输入划分的模型树

图 1 M5'模型树(样本空间分割图)

Fig. 1 M5' model tree (splitting the input space by M5' model tree algorithm)

设共有  $n$  个离散的样本,即  $X_1$ 、 $X_2$  均具有  $n$  个值;对  $X_1$  空间按其样本值排序,得到序列

$$A_{X_1} = \{x_1^1, x_1^2, x_1^3, \dots, x_1^n\} \quad (2)$$

其中  $x_1^1 \leq x_1^2 \leq x_1^3 \dots \leq x_1^n$ 。

依次对  $A_{X_1}$  按照  $x_1^2, x_1^3, \dots, x_1^{n-1}$  进行 2 分, 共有  $n-2$  中分法, 按式(1)计算可获得  $n-2$  个  $S_{DR}$  值, 设其最大值为  $S_{DR\max}(A_{X_1})$ , 其对应的分隔点为  $x_1^A$ 。

同理, 对  $X_2$  空间按其样本值排序, 也可获得其最大值为  $S_{DR\max}(A_{X_2})$ , 其对应的分隔点为  $x_2^A$ 。

若  $S_{DR\max}(A_{X_1}) \geq S_{DR\max}(A_{X_2})$ , 则首先选择  $X_1$  进行劈分, 劈分点为  $x_1^A$ ; 否则, 则首先选择  $X_2$  进行劈分, 劈分点为  $x_2^A$ 。

这样就可以获得如图 1(a)的第一个空间分割线  $X_2=2$ , 及图 1(b)模型树的顶级结点或跟结点, 即  $X_2>2$  的节点。对分割后的平面继续按上述规则, 就可以实现不断的平面分割, 或模型树的不断生成。

按以上步骤, 从根结点开始, 对每个结点递归劈分。树停止生长的条件有 2 个: 1 是分割结点样本数少于一定数量停止分割, 本文取为 4; 2 是达到结点的样本目标属性的标准差与总体样本目标属性标准差的比例小于某个限定值时, 树停止生长。本文取值的比值为 5%, 即

$$\frac{S_{DR}}{sd(T)} < 5\% \quad (3)$$

由上述分析可知, 样本集的在某结点处劈分需要确定 2 个参数: 劈分的输入分量属性和劈分阈值。以下是劈分的具体步骤:

1) 设结点处的训练数据集  $T$  共有  $n$  个样例, 每个样本为  $t_i$ ,  $i=1,2,3,\dots,n$ ; 输入属性表示为  $X_i$ ,  $i=1,2,\dots,m$ ; 目标属性为  $Y$ ; 计算训练数据集  $T$  的标准差为  $S_D$ 。

2) 对所有输入分量  $X_i$ , 分别按其分量值排序, 生成数据集  $T_i$ ,  $T_i = \{t_i \in T | x_i^1 \leq x_i^2 \leq x_i^3 \dots \leq x_i^n\}$ 。

3) 对 2) 的数据集  $T_i$ , 顺序取  $x_i^k$  值 ( $k=2,\dots,n-1$ ), 可以将  $T_i$  劈分成 2 个子集  $T_i^1 = \{t_i \in T | \forall x_i^j, x_i^j \leq x_i^k\}$ ,  $T_i^2 = \{t_i \in T | \forall x_i^j, x_i^j > x_i^k\}$ , 共有  $n-2$  中劈分组合, 分别按式(1)计算  $S_{DR}(T_i, x_i^k)$ 。计算获得最大的  $S_{DR\max}(T_i) = \max[S_{DR}(T_i, x_i^k)]$ 。

4) 对 3) 所有输入分量  $X_i$  对应  $S_{DR\max}(T_i)$  中选择最大值, 其对应的输入分量将作为该结点的劈分分量, 对应的该分量的样本值即该劈分的阈值。

该树停止生长的结点称为叶子结点, 对叶子结点样本, 利用线性回归算法产生一个多元回归方程, 得到线性模型。

可以看出, 树的生长就是寻找最大的空间分割差异分界, 式(1)给出的是按有限样本的标准差预测空间的离散度, 并以此作为差异性空间分割的依据。同时可以看出, 空间分割与分割后的线性方程没有直接关系。

### 1.3 树的剪切

剪枝即是对某些子树进行归并而以叶子结点取代, 以提高整个模型的简洁性和效率。

在剪切之前, 首先利用线性回归方法遍历树的各个结点, 拟合出每个结点的多元线性方程。剪切的原则为预测误差减少量, 即

$$E_R = |N|R_{MSE} - |N_l|R_{MSEl} - |N_r|R_{MSEr} \quad (4)$$

式中:  $R_{MSE}$  为该结点(包括其下属分支和叶子的所有样本)出拟合方程预测的均方根误差;  $R_{MSEl}, R_{MSEr}$  为该结点劈分后的左、右叶子节点预测的均方根误差。当  $E_R$  大于 0 时, 该子树保留, 否则, 将该子树转变成一个叶子结点。剪切过程也是一个递归的过程。

### 1.4 树的平滑

剪切后, 树的相邻叶子结点处可能会出现一定程度的不连续性。因此, Quinlan 提出树叶子结点的平滑方法<sup>[15, 22]</sup>。每个结点都存在一个按线性回归方法拟合的多元线性方程, 并且具有相同的参数, 因此子结点与其父结点的 2 个方程合并为 1 个新的线性方程即

$$f_{\text{new}} = \frac{nf_{\text{child}} + kf_{\text{parent}}}{n+k} \quad (5)$$

式中:  $f_{\text{parent}}$  为叶子上级父结点拟合方程;  $f_{\text{child}}$  为叶子结点拟合方程;  $n$  为到达本叶子结点的训练样本数目;  $k$  为一个常数(通常取值为 15);  $f_{\text{new}}$  为合并的方程; 若子结点采用新函数后的  $R_{MSE}$  变化小于一定的阈值, 则将取代子结点的线性方程, 否则不进行平滑处理。

## 2 数据特性及数据处理

### 2.1 数据处理

实测数据中不可避免地含有不确定的随机误差, 这些误差最终会导致 M5' 模型的偏差。因此, 在利用这些实测数据进行数据挖掘之前, 必须对实测数据进行预处理, 以减小采集、传输过程中产生的各种误差。具体包括:

1) 数据收集与抽取。根据分析问题的要求,

通过条件设置，从源数据集中抽取出样本数据集。

2) 数据采样。按数据挖掘的不同需要，从样本数据集中按照提供的采样方法抽取分析数据集。

3) 野值的识别、剔除。用7点二阶前推插值算法对野值进行识别和剔除。

4) 数据补正。经过以上步骤后，数据存在不连续现象，选用7点二阶多项式中心插值拟合算法对剔除数据进行补正。

2.2 数据特性及 M5'模型识别结果

浙江省已建立了“以热定电”为目的的全省集中建设的地方热电厂数据采集与管理信息系统<sup>[24]</sup>，对将近110家电厂进行时间间隔为5min的热电生产过程数据采集。本文对其中数据完备的热电机组汽耗量特性方程进行了研究。

以某热电厂型号为C12-4.9/0.88-HQ单抽机组为例，选取2008-05-01至2009-04-31跨度为一年的历史数据。每小时提取1个样本，共8760个样本，并对样本数据进行预处理。要识别机组进汽量F与机组输出功率P和蒸汽抽汽量D<sub>1</sub>的关系。

利用M5'算法训练完成后，该模型树的叶子节点多达39个，如图2所示；时间仅仅花费9.45s。

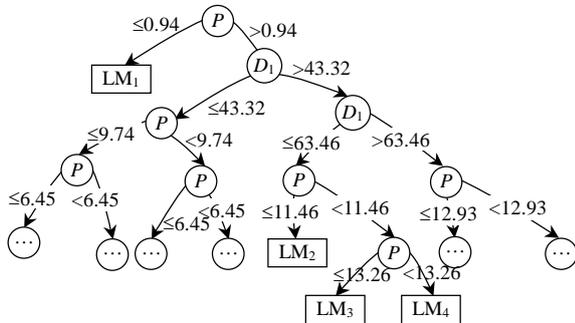


图2 M5'模型树枝的树状图

Fig. 2 Tree diagram of sections of M5's model

从图2中看出，M5'模型树利用输出功率P与低压抽汽量D<sub>1</sub>这2个属性将机组的汽耗量特性分成了多个多元线性方程，其中LM<sub>1</sub>~LM<sub>4</sub>只是其中的几个分支：

当P≤0.94时，汽耗量特性方程为LM<sub>1</sub>：

$$LM_1: F=0.0062D_1+0.0376P-0.8675 \quad (6)$$

当0.94<P≤11.46且43.32<D<sub>1</sub>≤63.46时，汽耗量特性方程为LM<sub>2</sub>：

$$LM_2: F=0.7118D_1+4.8575P+1.7386 \quad (7)$$

当11.46<P≤13.26且43.32<D<sub>1</sub>≤63.46时，汽耗量特性方程为LM<sub>3</sub>：

$$LM_3: F=0.6543D_1+3.2108P+24.5892 \quad (8)$$

当13.26<P且43.32<D<sub>1</sub>≤63.46时，汽耗量特性方程为LM<sub>4</sub>：

$$LM_4: F=0.764D_1+2.9338P+24.1138 \quad (9)$$

2.3 基于多元线性回归的模型的识别结果

根据样本数据，采用多元线性回归方法对型号C12-4.9/0.88-HQ单抽机组进行拟合方程，根据样本多元线性拟合汽耗特性，其特性方程为

$$F=0.7406D_1+4.5196P+4.028 \quad (10)$$

2.4 模型对比

本文采用交叉检验的方式评估模型的误差率。模型评估指标为关联度、绝对差值率、均方根误差率。

关联度：

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}} \quad (11)$$

绝对差值：

$$M_{AE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (12)$$

均方根误差：

$$R_{MSE} = \sqrt{\frac{1}{n} \sum(x_i - y_i)^2} \quad (13)$$

式中：x<sub>i</sub>为测试样本集的进汽量数据；y<sub>i</sub>为模型预测进汽量数据； $\bar{x}$ 为测试样本进汽量平均值； $\bar{y}$ 为模型预测进汽量平均值。

关联度R表示模型与真实关联函数的相似程度，越接近1表明模型越接近真实关联函数；M<sub>AE</sub>用于评判预测值与实际值之间的差异度，注重反映个体样本的预测效果程度，越小越好。R<sub>MSE</sub>与M<sub>AE</sub>类似，但R<sub>MSE</sub>注重反映样本总体的预测效果程度，值越小越好。

由表1可以看出，2个模型的关联度均很高，说明利用M5'和多元线性拟合的方法分析机组的汽耗量均非常有效。从绝对差值率和均方根误差率2个指标看，M5'的预测效果明显优于多元线性方程

表1 M5'与多元线性方程的模型比较

Tab. 1 Model comparison of M5' and MLR

模型名称	关联度 R	绝对	均方根
		差值率 M <sub>AE</sub> /%	误差率 R <sub>MSE</sub> /%
M5'	0.998 9	3.217 7	4.624 1
多元线性回归模型	0.996 0	8.571 9	8.891 9

拟合的效果。

从图 3 可以看出, M5' 模型预测点比多元线性拟合模型预测点更多地分布在对角线附近, 尤其是大值。

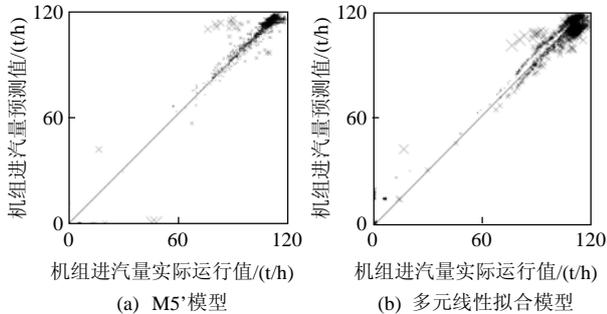
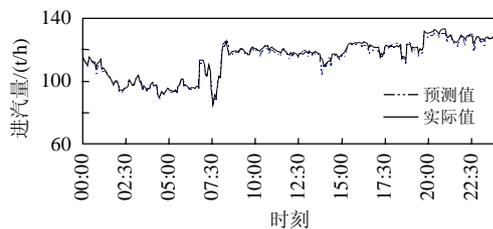


图 3 M5'、多元线性拟合模型预测值-实际值的对比图

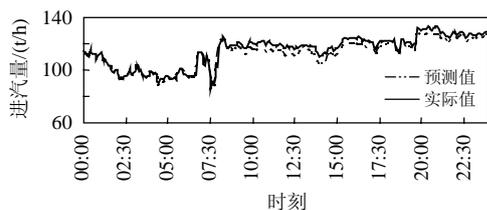
Fig. 3 Comparison of M5' and

#### MLR prediction values with real measured

图 4(a)、(b)分别为采用 M5'和多元线性模型进行进汽量预测的一个片段, 可以看出, 在 00:00—8:30 时段, 两模型预测效果基本相同, 从 5:00—23:55, M5'模型预测曲线与实际的运行曲线基本吻合, 而多元线性模型预测曲线严重偏离了实际运行曲线。



(a) M5'模型进汽量预测与实测曲线



(b) 多元线性模型进汽量预测与实测曲线图

图 4 M5'、多元线性模型预测与实测曲线对比

Fig. 4 Comparison of M5' and

#### MLR prediction values with real measured

### 2.5 结果分析

M5'模型之所以更好地实现了对汽耗量的拟合及其预测是因为 M5'算法有一个强大的数据处理和数据拟合的建树体系, 是一种精巧和高度原则化的学习方法。

1) M5'树结构植根于  $S_{DR}$  最大化原则, 树的大小由训练数据本身的离散化程度决定。因此, M5'

可以根据系统的非线性程度来构造树结构和大小, 更好地逼近目标特性曲线。

2) M5'叶子节点不是一个单一数据值, 而经过劈分后由若干个相似度最高的数据样本拟合线性方程, 这样大大增加了 M5'算法的泛化能力。

M5'采用对劈法自动分段搜索拟合。M5'的叶子节点的拟合线性方程没有直接联系, 在剪切处理若左右叶子的特性不相同, 该两叶子不会被剪切掉; 平滑处理中, 要判断平滑后的函数是否能保持精度, 若不能保证, 则不进行平滑处理, 因此 M5'在拟合非线性特性方程, 尤其是非凸形和非连续性的特性方程具有优势。

### 3 结论

本文提出了采用一种数据挖掘算法——M5'模型树的抽汽式机组汽耗量特性模型和其模型结构及参数识别算法, 滚动利用最新的电厂测量历史数据获取最新的汽耗量特性, 以取代传统定期的现场机组特性测试。该方法简单、有效, 逼近能力强, 自动化程度高, 在处理非凸形和非连续性的特性方程具有很强的优势。通过多个热电厂的实时数据进行验证, 具有很高的预测精度, 效果优于多元线性回归拟合方程。

### 参考文献

- [1] Padhy N P. Unit commitment: a bibliographical survey[J]. IEEE Trans on Power Systems, 2004, 19(2): 1196-1205.
- [2] 李蔚, 刘长东, 盛德仁, 等. 基于免疫遗传算法的机组负荷优化分配研究[J]. 中国电机工程学报, 2004, 24(7): 241-245.  
Li Wei, Liu Changdong, Sheng Deren, et al. Research on optimization of unit commitment based on immune genetic algorithm[J]. Proceedings of the CSEE, 2004, 24(7): 241-245(in Chinese).
- [3] 温志刚, 王勇, 骆贵兵, 等. 模拟退火算法在供热机组负荷分配中的应用研究[J]. 热力发电, 2003, 32(7): 18-20.  
Wen Zhigang, Wang Yong, Luo Guibing, et al. Research on optimization of unit commitment based on simulated annealing algorithm in the heat-supply units[J]. Power Engineering, 2003, 32(7): 18-20(in Chinese).
- [4] Aiyong Rong, Henri Hakonen, Risto Lahdelma. A variant of the dynamic programming algorithm for unit commitment of combined heat and power systems [J]. European Journal of Operational Research,

- 2008(190): 741-755.
- [5] 王友, 马晓茜, 刘翱. 自动发电控制下的火电厂厂级负荷优化分配[J]. 中国电机工程学报, 2008, 28(14): 103-107.  
Wang You, Ma Xiaoqian, Liu Ao. Study on plant-level optimal load distribution based on automatic generation control[J]. Proceedings of the CSEE, 2008, 28(14): 103-107(in Chinese).
- [6] 徐帆, 姚建国, 耿建, 等. 机组耗量特性的混合整数模型建立与分析[J]. 电力系统自动化, 2010, 34(10): 45-50.  
Xu Fan, Yao Jianguo, Geng Jian, et al. Modeling and analysis of unit I/O characteristics based on mixed-integer programming[J]. Automation of Electric Power Systems, 2010, 34(10): 45-50(in Chinese).
- [7] 王治国, 刘吉臻, 谭文, 等. 基于快速性与经济性多目标优化的火电厂厂级负荷分配研究[J]. 中国电机工程学报, 2006, 26(19): 86-92.  
Wang Zhiguo, Liu Jizhen, Tan Wen, et al. Multi-objective optimal load distribution based on speediness and economy in power plants[J]. Proceedings of the CSEE, 2006, 26(19): 86-92(in Chinese).
- [8] 盛德仁, 陈坚红, 李蔚, 等. 供热机组间热负荷优化分配的研究[J]. 动力工程, 2001, 31(6): 1560-1563.  
Sheng Deren, Chen Jianhong, Li Wei, et al. Research on optimized distribution of thermal and electric load in the heat-supply units[J]. Power Engineering, 2001, 31(6): 1560-1563(in Chinese).
- [9] 朱红霞, 沈炯, 缪国均. 基于改进型免疫遗传算法的母管制机组负荷优化分配[J]. 热力发电, 2009, 38(8): 1-5.  
Zhu Hongxia, Shen Jiong, Miao Guojun. Study on load optimization distribution among unites with header system based on improved immune genetic algorithm[J]. Thermal Power Generation, 2009, 38(8): 1-5(in Chinese).
- [10] Makkonen S. Decision modeling tools for utilities in the deregulated energy market[D]. Finland: Helsinki University of Technology, 2005.
- [11] Rong A, Lahdelma R. An efficient linear programming model and optimization algorithm for tri-generation[J]. Applied Energy, 2005(82): 40-63.
- [12] Rong A, Hakonen H, Lahdelma R. An efficient linear model and optimisation algorithm for multi-site combined heat and power production[J]. European Journal of Operational Research, 2006, 168(2), 612-632.
- [13] Tveit T, Savola T, Fogelholm C J. Modeling of steam turbines for mixed integer nonlinear programming (MINLP) in design and off-design conditions of CHP plants[EB/OL]. SIMS'05 46th Conference on Simulation and Modeling. Trondheim, Norway, 2005, [http://www.scansims.org/sims2005/SIMS2005\\_65.pdf](http://www.scansims.org/sims2005/SIMS2005_65.pdf).
- [14] 陈树勇, 宋书芳, 李兰欣, 等. 智能电网技术综述[J]. 电网技术, 2009, 33(8): 1-7.  
Chen Shuyong, Song Shufang, Li Lanxin, et al. Survey on smart grid technology[J]. Power System Technology, 2009, 33(8): 1-7(in Chinese).
- [15] Quinlan R J. Learning with continuous classes[C]//5th Australian Joint Conference on Artificial Intelligence, Singapore, 1992: 343-348.
- [16] Wang Y, Witten I H. Induction of model trees for predicting continuous classes[C]//Poster Papers of the 9th European Conference on Machine Learning, 1997.
- [17] Breiman L, Friedman J H, Olshen R A, et al. Classification and regression trees[M]. Wadsworth Statistical Press, Belmont, CA, 1984.
- [18] Solomatine D P, Dulal K N. Model tree as an alternative to neural network in rainfall-runoff modeling[J]. Hydrological Sciences Journal, 2003, 48(3), 399-411.
- [19] Solomatine D P, Yunpeng X. M5 model trees and neural networks: application to flood forecasting in the upper reach of the huai river in china[J]. Journal of Hydrologic Engineering, 2004, 9(6): 491-501.
- [20] Bhattacharya B, Solomatine D P. Neural networks and M5 model trees in modelling water level-discharge relationship[J]. Neurocomputing, 2005, 63(1): 381-396.
- [21] Bhattacharya B, Price R K, Solomatine D P. Machine learning approach to modeling sediment transport[J]. Journal of Hydraulic Engineering, 2007, 133(4): 440-450.
- [22] Etemad-Shahidi A, Mahjoobi J. Comparison between M5' model tree and neural networks for prediction of significant wave height in lake superior[J]. Ocean Engineering, 2009, 36(15): 1175-1181.
- [23] Sakhare S, Deo M C. Derivation of wave spectrum using data driven methods[J]. Marine Structures, 2009, 22(3): 594-609.
- [24] Zhang Jianmin, Zhang Chenlai, Zhang Yunlei, et al. Data acquisition and analysis system for distributed cogeneration heat and electrical power plant in zhejiang province[C]//IEEE Asia-Pacific Power and Energy Engineering Conference, Chengdu, 2010.



章坚民

收稿日期: 2010-11-01。

作者简介:

章坚民(1962), 男, 教授, 研究方向为电力及能源系统自动化、信息化的建模和系统开发等, [zhangjm@tom.com](mailto:zhangjm@tom.com);

刘登涛(1986), 男, 硕士研究生, 研究方向为系统集成技术及应用。

(责任编辑 张媛媛)