

母线负荷异常数据复杂不确定性检测与 基于综合云的修正模型

尹星露,肖先勇,孙晓璐

(四川大学 电气信息学院,四川 成都 610065)

摘要: 母线负荷数据异常具有复杂不确定性,是进行母线负荷预测、确定电网运行方式和安全校核等必须解决的重要课题。用聚类分析法确定待测日负荷的相似集,基于母线负荷纵向分布规律和横向连续性,提出异常数据复杂不确定性检测方法;研究母线负荷数据的期望、熵和超熵等数学特征,提出基于综合云的异常数据修正模型。以所提方法对某电网 110 kV 母线负荷数据进行了分析和预测,结果证明了该方法的可行性、正确性和有效性。

关键词: 母线负荷; 负荷预测; 异常数据; 相似集; 不确定性分析; 数学特征; 检测与修正方法; 综合云模型; 聚类算法

中图分类号: TM 715

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2015.06.018

0 引言

异常数据检测与修正是进行母线负荷预测的重要基础,已受到国内外重视^[1]。导致母线负荷异常的原因很多,数据异常现象具有复杂不确定性,尤其是低压母线负荷的分散性进一步增加了复杂不确定性,异常数据检测困难。通常认为,改善测量、通信等硬件条件可提高数据可靠性^[2],但无法揭示母线负荷异常数据的固有规律,因此,研究母线负荷异常数据的检测与修正方法,仍是当前需要研究的重要课题。

国内外对异常数据检测与修正已开展了大量研究^[1-10]。文献[1,4]提出了基于变化率的判别方法,但对连续异常可能误判或漏判。文献[5-6]利用连续多日同时刻负荷的均值和方差进行检测,但对异常波动的检测效果不佳。文献[7]从负荷的横、纵向连续性出发,提出了基于数据密度的检测方法,但种子吸附阈值确定困难。对于异常数据修正,文献[1,7]提出了相似日同时刻负荷线性组合修正法;文献[8-10]提出了特征曲线调整法。这些方法遵循了负荷连续性原则,但对负荷变化的不确定性考虑不足。事实上,母线负荷变化同时具有连续性和不确定性,数据异常具有复杂不确定性,需针对复杂不确定性提出异常数据检测方法,并可用期望、熵、超熵等数学特征刻画母线负荷的复杂不确定性,采用李德毅院士提出的云模型进行修正。

本文基于聚类分析法,对母线负荷进行相似集划分;从母线负荷的纵向不确定性分布规律和横向连续性出发,研究异常数据的不确定特征,根据多日同时段负荷随机分布规律估算待测负荷在不同时刻的取值区间,检测异常数据,为避免疏漏,以相似日在

不同时刻的负荷变化率的正常范围为判据,对剩余待测数据进行检测,从而提出异常数据复杂不确定性检测方法;研究母线负荷的期望、熵、超熵等数学特征,综合利用负荷准周期变化规律和当前趋势,提出基于综合云模型的修正方法,并对算法进行了详细研究。以某实际电网 110 kV 母线负荷为例,验证了方法的可行性和正确性,并将该方法用于某母线负荷预测系统,证明了其工程适用性。

1 母线负荷相似集划分

1.1 相似集概念与划分依据

合理的聚类可避免由于母线负荷的类别不同而带来的异常数据检测的误判、漏判等问题,也为异常数据修正奠定了良好基础。每 15 min 采样一次,以 96 个负荷点构成的日负荷曲线总样本集进行有效聚类,本文采用改进的模糊 C 均值 FCM(Fuzzy C-Means)聚类方法^[9-11];利用减法聚类算法得到聚类数目和聚类中心,作为 FCM 算法的初始值,按最小距离原则对总样本进行划分,形成母线负荷不同类别的相似集。

1.2 相似集划分方法

a. 设 N 天总样本集为 $X=\{X_1, X_2, \dots, X_i, \dots, X_j, \dots, X_N\}$, $X_i=(X(i,1), X(i,2), \dots, X(i,96))$, 计算各样本 X_i 的密度指标 D_i :

$$D_i = \sum_{j=1}^N \exp \left[-\frac{\|X_i - X_j\|^2}{(0.5r_a)^2} \right] \quad (1)$$

$$r_a = \frac{1}{2} \min \{ \max \{ \|X_i - X_j\|, i=1, 2, \dots, N, j=1, 2, \dots, N \} \} \quad (2)$$

其中, r_a 为一正数,定义为样本 X_i 的一个邻域。

将式(1)中密度指标最高的样本 X_i 作为第一个初始聚类中心 c_1 ,对应的密度指标为 D_c^1 。

b. 根据式(3)进行修正,从剩余样本中选取计算结果最大的样本作为下一个聚类中心。对应密度指标为 D_c^k 。

$$D_i^k = D_i^{k-1} - D_c^{k-1} \exp \left[\frac{-\|X_i - c_{k-1}\|^2}{(0.5r_b)^2} \right] \quad (3)$$

通常 $r_b = 1.5r_a$, 可避免出现距离过近的聚类中心。

c. 判断 $\frac{D_c^k}{D_c^1} < \delta$ 是否成立,若不成立,则转到步骤

b;若成立,算法结束。相关研究证实^[11-12], $\delta \geq 0.5$ 时,可得到合理的聚类数,且 $\delta = 0.5$ 时聚类数目 k 最大,故当 D_c^k 与 D_c^1 之比小于 0.5 时停止计算,并记下合理聚类数目 k 和聚类中心 $c_1, c_2, \dots, c_k$ 。

d. 记 $g=1$,将减法聚类得到的聚类中心 $c_1(g), c_2(g), \dots, c_k(g)$ 作为 FCM 算法的初始聚类中心。

e. 计算所有样本与各聚类中心的距离:

$$d(X, c_j(g)) = \sum_{i=1}^N \sum_{j=1}^k (\mu_{ij})^m \|X_i - c_j(g)\|^2 \quad (4)$$

其中, μ_{ij} 为样本 X_i 属于第 j 类的隶属度,定义为式(5)。

$$\mu_{ij} = \left[\sum_{l=1}^k \left(\frac{d_{il}}{d_{il}} \right)^{2/(m-1)} \right]^{-1} \quad (5)$$

其中, d_{ij} 为样本 X_i 与第 j 类的距离; d_{il} 为样本 X_i 与每一类的距离; m 一般取 2。按最小距离原则将总样本 X 进行聚类。

f. 重新计算聚类中心:

$$c_j(g+1) = \frac{\sum_{i=1}^N [(\mu_{ij})^m c_j(g)]}{\sum_{j=1}^k (\mu_{ij})^m} \quad (6)$$

若存在 $j \in \{1, 2, \dots, k\}$, 有 $c_j(g+1) \neq c_j(g)$, 则 $g=g+1$, 转至步骤 e, 否则聚类结束。最终形成母线负荷不同类别的聚类中心 $c_1, c_2, \dots, c_k$ 和相似集 $G_1, G_2, \dots, G_k$ 。

2 异常数据复杂不确定性检测

2.1 异常数据检测基本思路

基于母线负荷纵向分布规律和横向连续性特点,提出异常数据复杂不确定性检测方法。基本思路为:按最小距离原则判别待测日曲线属于哪一个相似集,选取该相似集的 n 天样本作为待测日的相似日,并提取该相似集的聚类中心作为特征曲线。根据多日同时段负荷随机分布规律估算待测负荷在不同时刻的取值区间,检测得到 r 个异常数据,为避免疏漏,以相似日在不同时刻的负荷变化率的正常范围为判断,对剩余待测数据进行检测。

2.2 复杂不确定性刻画方法

根据 D 检验法^[13]可证明多日同一时刻的负荷样本 $Y = \{Y(i, t) | i=1, 2, \dots, n\}$ 近似服从正态分布,以此

为本文异常数据检测方法提供依据。假设检验问题为: H_0 样本服从正态分布, H_1 样本不服从正态分布。

将 n 天同一时刻负荷 $\{Y(i, t) | i=1, 2, \dots, n\}$, 由小到大排列,其均值 $\bar{Y}$ 为:

$$\bar{Y} = \frac{\sum_{i=1}^n Y(i, t)}{n} \quad (7)$$

检验统计量 $C(Y)$ 为:

$$C(Y) = \frac{\sum_{i=1}^{n'} (Y(n+1-i, t) - Y(i, t))^2 (n+1-2i)}{\sum_{i=1}^n (Y(i, t) - \bar{Y})^2 n} \quad (8)$$

$$n' = \begin{cases} \frac{n}{2} & n \text{ 为偶数} \\ \frac{n+1}{2} & n \text{ 为奇数} \end{cases}$$

当 $C(Y) > C(\alpha, n)$ (可查统计量分布表^[13], α 为显著性水平)时,接受 H_0 ;反之,接受 H_1 。由此刻画母线负荷的随机不确定性。

2.3 异常数据检测方法

按最小距离原则判别待测日曲线属于母线负荷哪一个相似集,选取该相似集的 n 天样本数据作为待测日的相似日,并提取该相似集的聚类中心作为特征曲线,算法如下。

a. 统计 n 天相似日在不同时刻的负荷均值 $\bar{Y}(t)$ 和方差 s_t^2 。

b. 利用多日同时段负荷近似正态分布,但总体 $N(\mu_t, \sigma_t^2)$ 的 σ_t^2 未知的性质,用相似日的方差 s_t^2 作 σ_t^2 的无偏估计。

c. 结合实际确定置信水平 $1-\alpha$,估算待测负荷 $\{Y_0(t) | t=1, 2, \dots, 96\}$ 在不同时刻可能的取值区间为:

$$\left[\bar{Y}(t) - \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1), \bar{Y}(t) + \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1) \right] \quad (9)$$

其中, s 为标准方差; $t_{\alpha/2}(n-1)$ 可通过 t 值表查询。

若 $Y_0(t)$ 不在对应时刻可能的取值区间,则判定该时刻的负荷为异常数据。步骤 a—c 基于母线负荷纵向分布规律,主要针对异常明显的缺失数据和极大、极小值进行检测,为避免其他异常数据疏漏,利用母线负荷横向连续性特点,对余下的 $96-r$ (r 为步骤 c 已检测出的异常数据个数)个待测数据进行检测。

d. 统计 n 天相似日在余下数据所对应时刻的负荷变化率的正常范围:

$$\left[\min \left(\frac{Y(i, t) - Y(i, t-1)}{Y(i, t-1)} \right), \max \left(\frac{Y(i, t) - Y(i, t-1)}{Y(i, t-1)} \right) \right] \quad (10)$$

其中, $i=1, 2, \dots, n; t=2, \dots, 96-r$ 。

对于计算 $t=1$ 时刻的负荷变化率时,可利用前一天最后一时刻的负荷值 $Y(i-1, 96)$ 进行负荷变化率运算。

e. 计算待测负荷值 $Y_0(t)$ 与特征曲线所对应的 $Y_c(t-1)$ 的负荷变化率 $\delta(t)$:

$$\delta(t)=\frac{Y_0(t)-Y_c(t-1)}{Y_c(t-1)} \tag{11}$$

用 $Y_c(t-1)$ 代替 $Y_0(t-1)$ 的目的是避免 $Y_0(t-1)$ 是异常数据而带来的误检。

f. 判断 $\delta(t)$ 是否在式 (10) 的正常范围,若不在,则说明该时刻的数据为异常数据。

3 母线负荷的数学特征与异常数据修正模型

3.1 母线负荷的数学特征

基于母线负荷变化的不确定性,本文采用云模型进行异常数据修正。云模型是用语言值表示某定性概念与其定量表示之间的不确定性转换模型。其数学特征为:期望 E_x 、熵 E_n 、超熵 H_e 。期望 E_x 刻画云的重心,熵 E_n 反映分布不确定性和可接受范围,超熵 H_e 刻画云滴的凝聚度,反映离散程度和云的厚度。

由 D 验证法确定多日同时段负荷的近似正态分布后,可利用普适性最强的正态云模型^[14]刻画负荷变化规律,由此确定期望曲线方程:

$$C_T(x)=e^{-(x-E_x)^2/(2E_n^2)} \tag{12}$$

3.2 异常数据修正的综合云模型

母线负荷曲线以 96 个负荷点为准周期,将 n 天相似日的负荷 $\{L(a_i, b_i)\}$ (其中 a_i 为时刻, b_i 为对应负荷) 作为历史数据集 D_H ,从整体反映负荷准周期性规律,依据 D_H 中的数据可挖掘得到修正规则集 $\{A_1 \rightarrow B_1, A_2 \rightarrow B_2, \dots, A_l \rightarrow B_l\}$ ^[15],如表 1 所示。其中 A_i, B_i 分别为 a_i, b_i 论域上的定性概念,可由其数学特征(E_x, E_n, H_e)来表征。

表 1 修正规则集
Table 1 Set of correction rules

时间 A_i	负荷 B_i	时间 A_i	负荷 B_i	时间 A_i	负荷 B_i
下降期 1	中 1	下降期 2	中 3	下降期 3	中 5
低谷期 1	低 1	低谷期 2	低 2	低谷期 3	低 3
上升期 1	中 2	上升期 2	中 4	上升期 3	中 6
高峰期 1	高 1	高峰期 2	高 2	高峰期 3	高 3

修正规则集中的 A_i 和 B_i 分别表示规则前件和后件语言变量的原子概念,通过判定待修正负荷的时刻 a_i 属于前件语言变量中的哪个原子来激活相应规则,令 $\mu=a_i$,方法如下。

由参数 $A_1(E_{x1}, E_{n1}, H_{e1}), \dots, A_l(E_{xl}, E_{nl}, H_{el})$ 分别构造 μ -条件下正态云发生器 $CG_1, CG_2, \dots, CG_l$,将 μ 分别输入云发生器得一系列输出 $\mu_1, \mu_2, \dots, \mu_l$,分别反映了 μ 对 $A_1, A_2, \dots, A_l$ 的隶属程度。从中找出最大值 $\mu_{\max}$,则 $\mu=a_i$ 属于 A_i 。说明规则 $A_i \rightarrow B_i$ 最能反映时刻 a_i 的准周期规律,其后件 B_i 相应(E_{xB}, E_{nB}, H_{eB})作为修正信息,可称为历史修正云 B_i 。历史修正云推

理器的原理如图 1 所示。

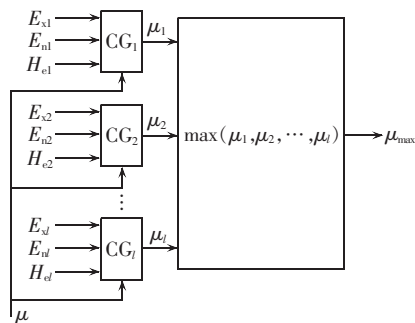


图 1 历史修正云推理器

Fig.1 Reasoning engine of historic data correction cloud

以待修正数据的 n 天相似日同时刻 a_i 的负荷 $L=\{L(i, a_i)|i=1, 2, \dots, n\}$ 作为当前时刻数据集 D_C ,根据 D_C 中的数据分布,用无需确定度信息的逆向云算法^[16]得到当前时刻趋势——当前修正云 I_i ,该信息既有不确定性又遵循当前时刻数据的分布规律,其算法如下:

- a. $E_{xI}=\text{mean}(L)$;
- b. 一阶绝对中心距 $M_1=\frac{1}{n} \sum_{i=1}^n |L(i, a_i)-E_{xI}|$;
- c. 二阶平方距 $M_2=\frac{1}{n-1} \sum_{i=1}^n (L(i, a_i)-E_{xI})^2$;
- d. $E_{nI}=\sqrt{\frac{\pi}{2}} M_1$;
- e. $H_{eI}=\sqrt{M_2-E_{nI}^2}$ 。

结合历史修正云 B_i 和当前修正云 I_i ,形成综合云修正模型 $S(E_x, E_n, H_e)$:

$$\begin{cases} E_x = \frac{E_{xB}E_{nB} + E_{xI}E_{nI}}{E_{nB} + E_{nI}} \\ E_n = E_{nB} + E_{nI} \\ H_e = \frac{H_{eB}E_{nB} + H_{eI}E_{nI}}{E_{nB} + E_{nI}} \end{cases} \tag{13}$$

基于修正综合云模型 S 所得的修正数据 $L(a_i)$,同时包含了负荷整体特性和当前时刻数据分布规律,具体算法如下。

a. 产生一个期望为 E_n 、方差为 H_e 的正态随机数 $E'_n=\text{NORM}(E_n, H_e)$ 。

b. 以 E_x 为期望、 E'_n 为方差得到一个正态随机修正数据 $L(a_i)=\text{NORM}(E_x, E'_n)$ 。

c. 检验所得修正值 $L(a_i)$ 是否能满足横向连续性和纵向分布规律,计算公式如下:

$$\begin{cases} \delta(a_i) \in \left[\min \left(\frac{L(i, a_i) - L(i, a_{i-1})}{L(i, a_{i-1})}, \max \left(\frac{L(i, a_i) - L(i, a_{i-1})}{L(i, a_{i-1})} \right) \right) \right] \\ L(a_i) \in \left[\bar{L}(a_i) - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1), \bar{L}(a_i) + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) \right] \end{cases} \tag{14}$$

若不满足,转步骤**b**;若满足,将所得修正值 $L(a_i)$ 作为最终的修正结果。

4 异常数据检测与修正算法流程

异常数据检测与修正流程如图 2 所示。

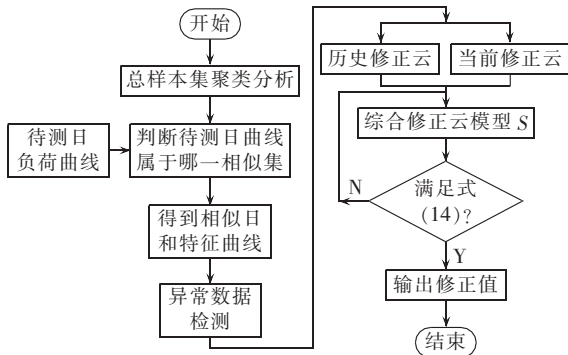


图 2 异常数据检测与修正流程

Fig.2 Flowchart of abnormal data detection and correction

5 实例分析

以某地区电网某 110 kV 母线为例,以 2012 年 7 月到 9 月的日负荷为数据源,人为丢失多个数据,并加入连续突变量。用本文方法进行异常数据检测与修正。上述算法均由 MATLAB 程序实现。

表 2 为随机抽取负荷数据的检测结果,可见,本文方法的平均检测正确率为 87.63%,高于改进横向法^[1]的平均正确率 83.28%。

表 2 随机抽取负荷数据的检测结果
Table 2 Results of detection for randomly extracted load data

日期	异常数据数量	漏检数		错检数	
		本文方法	改进横向法	本文方法	改进横向法
8 月 13 日	18	2	1	1	1
8 月 17 日	24	1	2	2	3
8 月 24 日	16	2	3	0	0
9 月 3 日	13	0	2	1	0
9 月 4 日	40	3	4	2	3

分别用加权平均法、特征曲线调整法^[11]、综合云修正模型对异常数据进行修正。

加权平均法:选取待修正日的相似日的同一时刻的历史负荷数据作为参考数据,则修正值如式(15)所示。

$$\begin{cases} L'(d,t)=\lambda_1L(d-1,t)+\lambda_2L(d-2,t)+\cdots+ \\ \lambda_qL(d-q,t) \\ \sum_{j=1}^q\lambda_j=1 \end{cases} \quad (15)$$

其中, λ_j 表征第 $d-j$ 天 t 时刻负荷对待修正值的影响程度,本文取 $\lambda_1=\lambda_2=\cdots=\lambda_5=0.2$ 。

特征曲线调整法:若检测出 t_1 到 t_2 为坏数据,则

修正值如式(16)所示。

$$L'(d,t)=L_c(d,t)\left[\frac{L(d,t_1-1)}{2L_c(d,t_1-1)}+\frac{L(d,t_2+1)}{2L_c(d,t_2+1)}\right] \quad (16)$$

其中, $t=t_1,t_1+1,\cdots,t_2$; $L_c(d,t)$ 、 $L_c(d,t_1-1)$ 、 $L_c(d,t_2+1)$ 为特征曲线上对应的负荷值; $L(d,t_1-1)$ 、 $L(d,t_2+1)$ 为待修正日曲线对应的负荷值。

以 2012 年 9 月 3 日的负荷曲线为例,修正结果和相对误差分别如表 3、表 4 所示。

表 3 异常负荷数据修正结果

Table 3 Results of abnormal load data correction

异常数据点	实际值/kW	异常数据值/kW	修正值/kW		
			方法 1	方法 2	方法 3
2	11739.2	0	10746.6	10964.7	11333.4
5	13596.0	33596.0	12276.0	12545.0	12594.2
15	12390.4	2390.4	10093.6	10534.5	11234.4
23	12196.8	0	10658.6	10964.2	11286.7
32	15605.6	35605.6	16933.8	16032.4	15234.9
58	18928.8	8928.8	17540.8	17962.8	18039.2
59	20134.4	30134.4	18658.7	18936.1	18963.4
75	24824.8	14824.8	22506.8	22763.3	23593.0
84	18057.6	28057.6	15602.8	15840.0	17655.1
87	14819.2	4819.2	13928.7	14132.1	14224.0
91	11976.8	0	13492.2	13675.5	11820.1
92	13860.0	0	14233.1	14566.0	12340.2
93	15232.8	0	14444.5	14563.5	13968.0

注:方法 1、2、3 分别表示加权平均法、特征曲线调整法、云模型修正法,后同。

表 4 各方法相对误差

Table 4 Relative errors of different methods

异常数据点	相对误差/%		
	方法 1	方法 2	方法 3
2	8.46	6.60	3.46
5	9.71	7.73	7.37
15	18.54	14.98	9.33
23	12.61	10.11	7.46
32	8.51	2.73	2.38
58	7.33	5.10	4.70
59	7.33	5.95	5.82
75	9.34	8.30	4.96
84	13.59	12.28	2.23
87	6.01	4.64	4.02
91	12.65	14.18	1.31
92	2.69	5.09	10.97
93	5.18	4.39	8.30
平均相对误差/%	9.38	7.85	5.56

从表 3 可知,本文方法所得修正值更接近实际。由表 4 可见,由于母线负荷的基数相对于系统负荷小,某些点相对误差偏大,用综合云模型修正后的相对误差基本在 10% 以内,平均相对误差比其他 2 种方法分别低 3.82% 和 2.29%,明显好于其他方法。

为进一步比较,将以上 3 种修正方法处理后的样本和未经处理的样本用于某母线负荷预测系统,并以改进灰色模型预测法^[18]为例,预测 2012 年 9 月 10 日到 9 月 14 日的负荷曲线。依据国网公司下达的《电网母线负荷预测工作考核管理办法》中的相关

考核指标,母线负荷在 t 时刻的预测误差定义为:

$$E_t=(|y'(t)-y(t)|/y_B)\times 100\% \tag{17}$$

其中, $y'(t)$ 为实际值; $y(t)$ 为预测值; y_B 为负荷基准值,对应 110 kV 电压等级的 $y_B=114$ MW。日母线负荷预测准确率为:

$$A=\left(1-\sqrt{\frac{1}{96}\sum_{i=1}^{96}E_i^2}\right)\times 100\% \tag{18}$$

如表 5 所示,用本文方法处理后的样本进行预测,平均预测准确率较其他 2 种方法处理后的平均准确率分别提高了 0.23 % 和 0.13 %,较未处理时平均准确率提高了 0.83 %。

表 5 母线负荷预测准确率对比
Table 5 Comparison of bus load forecasting accuracy

日期	未处理的 准确率/%	处理后的准确率/%		
		方法 1	方法 2	方法 3
9 月 10 日	96.05	96.53	96.65	96.81
9 月 11 日	97.16	97.52	97.61	97.76
9 月 12 日	94.98	95.86	96.13	95.98
9 月 13 日	95.89	96.62	96.56	96.83
9 月 14 日	96.32	96.87	96.94	97.20
平均预测准确率/%	96.08	96.68	96.78	96.91

6 结论

a. 考虑母线负荷的纵向不确定性分布规律和横向连续性,采用异常数据复杂不确定性检测方法,其结果正确率高,可满足工程应用需要。

b. 基于历史修正云和当前修正云的母线负荷异常数据综合云修正模型,同时考虑了不同时间粒度的变化规律,所得修正结果更符合实际。

c. 本文方法在某地区电网母线负荷预测系统中的应用表明,采用本文方法后,母线负荷预测的准确率得到了明显提高,能更好地满足实际需要。

结合实际系统中母线负荷影响因素和变化规律,进一步考虑气象因素、计划检修、负荷计划性转移等的影响,是值得研究的课题。

参考文献:

[1] 李光珍,刘文颖,云会周,等. 母线负荷预测中样本数据预处理的新方法[J]. 电网技术,2010,34(2):149-154.
LI Guangzhen,LIU Wenying,YUN Huizhou,et al. A new date preprocessing method for bus load forecasting[J]. Power System Technology,2010,34(2):149-154.

[2] 康仁. 基于小波分析的母线负荷预测不良数据检测[J]. 中国电力,2011,44(8):5-8.
KANG Ren. Bad data identification based on wavelet in bus load forecasting[J]. Electric Power,2011,44(8):5-8.

[3] 康重庆,夏清,刘梅. 电力系统负荷预测[M]. 北京:中国电力出版社,2007:169-262.

[4] 莫维仁,张伯明,孙宏斌,等. 扩展短期负荷预测方法的应用[J]. 电网技术,2003,27(5):6-9.
MO Weiren,ZHANG Boming,SUN Hongbin,et al. Application of extended short-term load forecasting[J]. Power System Technology,

2003,27(5):6-9.

[5] 叶锋,何桦,顾全,等. EMS 中负荷预测不良数据的辨识与修正[J]. 电力系统自动化,2006,30(15):85-88.
YE Feng,HE Hua,GU Quan,et al. Bad data identification and correction for load forecasting in energy management system[J]. Automation of Electric Power Systems,2006,30(15):85-88.

[6] 陈建华,戴铁潮,张宁,等. 确定性合同分解中异常负荷数据的识别与修正[J]. 电力系统自动化,2009,33(6):21-25.
CHEN Jianhua,DAI Tiechao,ZHANG Ning,et al. Load outlier identification and correction for deterministic contract decomposition[J]. Automation of Electric Power Systems,2009,33(6):21-25.

[7] 陈亮,文福拴,童述林. 基于密度估计的异常电力负荷数据辨识与修正[J]. 华南理工大学学报:自然科学版,2012,40(2):124-129.
CHEN Liang,WEN Fushuan,TONG Shulin. Identification and correction of abnormal electric load data based on density evaluation[J]. Journal of South China University of Technology: Natural Science Edition,2012,40(2):124-129.

[8] 张国江,邱家驹,李继红. 基于人工神经网络的电力负荷坏数据辨识与调整[J]. 中国电机工程学报,2001,21(8):104-113.
ZHANG Guojiang,QIU Jiaju,LI Jihong. Outlier identification and justification based on neural network[J]. Proceedings of the CSEE,2001,21(8):104-113.

[9] 孙谦,姚建刚,金敏,等. 基于特性矩阵分层分析的短期母线负荷预测坏数据处理策略[J]. 电工技术学报,2013,28(7):226-233.
SUN Qian,YAO Jiangang,JIN Min,et al. Bad data processing strategies for short-term bus load forecasting based on stratified analysis of characteristic matrix[J]. Transactions of China Electrotechnical Society,2013,28(7):226-233.

[10] 蒋雯倩,李欣然,钱军. 改进 FCM 算法及其在电力负荷坏数据处理的应用[J]. 电力系统及其自动化学报,2011,23(5):1-5.
JIANG Wenqian,LI Xinran,QIAN Jun. Application of improved FCM algorithm in outlier processing of power load[J]. Proceedings of the CSU-EPSA,2011,23(5):1-5.

[11] 孙谦,姚建刚,赵俊,等. 基于最优交集相似日选取的短期母线负荷综合预测[J]. 中国电机工程学报,2013,33(4):126-134.
SUN Qian,YAO Jiangang,ZHAO Jun,et al. Short-term bus load integrated forecasting based on selecting optimal intersection similar days[J]. Proceedings of the CSEE,2013,33(4):126-134.

[12] NIKHIL R P,CHARKRABORTY D. Mountain and subtractive clustering method:improvements and generalizations[J]. International Journal of Intelligent Systems,2000,15(4):329-341.

[13] 白杰华,宋向东. 总体正态性检测——W 检验法的改进[J]. 辽宁工程技术大学学报:自然科学版,2013,32(3):413-416.
BAI Jiehua,SONG Xiangdong. Improvement of winspection method in overall normal inspection[J]. Journal of Liaoning Technical University:Natural Science,2013,32(3):413-416.

[14] 李德毅,刘常显. 论正态云模型的普适性[J]. 中国工程科学,2004,6(8):28-34.
LI Deyi,LIU Changyu. Study on the university of the normal cloud model[J]. Engineering Science,2004,6(8):28-34.

[15] 杜鹄,宋自林,李德毅. 基于云模型的关联规则挖掘方法[J]. 解放军理工大学学报,2000,1(1):29-34.
DU Yi,SONG Zilin,LI Deyi. Mining association rules based on cloud model[J]. Journal of PLA University of Science and

- Technology, 2000, 1(1): 29-34.
- [16] 杨薛明, 苑津莎, 王建峰. 基于云理论的配电网空间负荷预测方法研究[J]. 中国电机工程学报, 2006, 26(6): 30-36.
- YANG Xueming, YUAN Jinsha, WANG Jianfeng. A new spatial forecasting method for distribution network based on cloud theory[J]. Proceedings of the CSEE, 2006, 26(6): 30-36.
- [17] 孙晓璐, 肖先勇, 尹星露, 等. 基于模型有效度的地区电网母线负荷组合预测与系统实现[J]. 电力自动化设备, 2014, 34(12): 106-110.
- SUN Xiaolu, XIAO Xianyong, YIN Xinglu, et al. Combination forecasting based on model availability for district network bus loads and its system implementation[J]. Electric Power Automation Equipment, 2014, 34(12): 106-110.
- [18] 廖峰, 刘清良, 贺辉, 等. 基于改进灰色模型与综合气象因素的母线负荷预测[J]. 电网技术, 2011, 35(10): 183-188.
- LIAO Feng, LIU Qingliang, HE Hui, et al. Bus load forecasting based on improved grey model and meteorological elements[J]. Power System Technology, 2011, 35(10): 183-188.
- [19] EFSTATIOS P, THEOFANIS S. Short-term load forecasting: the similar shape functional time-series predictor[J]. IEEE Transactions on Power Systems, 2013, 28(4): 3818-3825.
- [20] 童述林, 文福拴, 陈亮. 电力负荷数据预处理的二维小波阈值去噪方法[J]. 电力系统自动化, 2012, 36(2): 101-104.
- TONG Shulin, WEN Fushuan, CHEN Liang. A two-dimension wavelet threshold de-noising method for electric load data pre-processing[J]. Automation of Electric Power Systems, 2012, 36(2): 101-104.
- [21] 张平, 潘学萍, 薛文超. 基于小波分解模糊灰色聚类 and BP 神经网络的短期负荷预测[J]. 电力自动化设备, 2012, 32(11): 121-125.
- ZHANG Ping, PAN Xueping, XUE Wenchao. Short-term load

forecasting based on wavelet decomposition, fuzzy gray correlation clustering and BP neural network[J]. Electric Power Automation Equipment, 2012, 32(11): 121-125.

- [22] 刘文霞, 徐晓波, 周樾. 基于支持向量机的纯电动公交车充/换电站日负荷预测[J]. 电力自动化设备, 2014, 34(11): 41-47.
- LIU Wenxia, XU Xiaobo, ZHOU Xi. Daily load forecasting based on SVM for electric bus charging station[J]. Electric Power Automation Equipment, 2014, 34(11): 41-47.
- [23] 马瑞, 姜飞, HUANG G M. 基于电力系统负荷变动速率的美国德州 ERCOT 短期负荷预测[J]. 电力自动化设备, 2012, 32(2): 81-84.
- MA Rui, JIANG Fei, HUANG G M. Short-term load forecasting based on load fluctuation rate[J]. Electric Power Automation Equipment, 2012, 32(2): 81-84.

作者简介:



尹星露

尹星露(1990—), 女, 山东莱州人, 硕士研究生, 研究方向为负荷预测、不确定性理论在电力系统中的应用(E-mail: yinxinglu@yeah.net);

肖先勇(1968—), 男, 四川宜宾人, 教授, 博士研究生导师, 博士, 从事不确定性理论在电力系统中的应用、电能质量、智能电网等的教学和研究工作(E-mail: xiaoxianyong@

163.com);

孙晓璐(1987—), 女, 山东鄄城人, 硕士研究生, 研究方向为负荷预测、不确定性理论在电力系统中的应用(E-mail: sunxiaolu2008@126.com)。

Complex uncertainty detection and synthesized cloud correction model for abnormal bus load data

YIN Xinglu, XIAO Xianyong, SUN Xiaolu

(College of Electrical Engineering and Information, Sichuan University, Chengdu 610065, China)

Abstract: The complex uncertainty of abnormal bus load data is a key issue of bus load forecasting, power grid operating mode determination and security check. The cluster analysis is applied to determine the similar set of daily load to be forecasted and a method of complex uncertainty detection for abnormal data is proposed based on the vertical distribution regularity and horizontal continuity of bus load. The mathematical characteristics of bus load data, such as expectation, entropy and hyper entropy, are studied and an abnormal data correction model based on the synthesized cloud is proposed. The 110 kV bus load data of a power grid are analyzed and forecasted by the proposed method and the results demonstrate its feasibility, correctness and effectiveness.

Key words: bus load; electric load forecasting; abnormal data; similar set; uncertainty analysis; mathematical characteristics; detection and correction; synthesized cloud model; clustering algorithms