

# 中长期负荷预测的异常数据辨识与缺失数据处理

毛李帆<sup>1</sup>, 姚建刚<sup>1</sup>, 金永顺<sup>2</sup>, 李文杰<sup>2</sup>, 关石磊<sup>2</sup>, 陈芳<sup>2</sup>

- (1. 湖南大学 电气与信息工程学院, 湖南省 长沙市 410082;
2. 湖南湖大华龙电气与信息技术有限公司, 湖南省 长沙市 410082)

## Abnormal Data Identification and Missing Data Filling in Medium-and Long-Term Load Forecasting

MAO Li-fan<sup>1</sup>, YAO Jian-gang<sup>1</sup>, JIN Yong-shun<sup>2</sup>, LI Wen-jie<sup>2</sup>, GUAN Shi-lei<sup>2</sup>, CHEN Fang<sup>2</sup>

- (1. College of Electrical and Information Engineering, Hunan University, Changsha 410082, Hunan Province, China;
2. Hunan HDHL Electrical & Information Technology Co., Ltd., Changsha 410082, Hunan Province, China)

**ABSTRACT:** Historical load data is the basis of medium- and long-term load forecasting, thus the abnormal historical data and historical data missing seriously affect the accuracy and effectiveness of load forecasting. To remedy the insufficiency of traditional methods for abnormal data identification and data filling, a method for missing data filling, which is based on both  $T^2$  ellipse map to identify abnormal data and least square support vector machine (LSSVM), is proposed. The principal component of historical data is extracted by partial least square (PLS) to compute the accumulative contribution rate (ACR) of historical data to principal component and draw  $T^2$  ellipse, thereby the abnormal historical data that possesses too high contribution rate can be identified; the variation trend of historical data is fitted by LSSVM, thus the missing data can be filled. Results of calculation example show that the  $T^2$  ellipse map can effectively identify the abnormal samples in historical data and LSSVM possesses good data filling performance, therefore the proposed method is practicable.

**KEY WORDS:** abnormal data; data missing; accumulative contribution rate (ACR);  $T^2$  ellipse map; least square support vector machine (LSSVM); load forecast

**摘要:** 负荷历史数据是进行中长期负荷预测的基础。历史数据异常及缺失将严重影响负荷预测模型的精度及有效性。针对传统异常数据辨识方法和缺失数据填补方法的不足,提出了基于  $T^2$  椭圆图的异常数据识别和基于最小二乘支持向量机(least square support vector machine, LSSVM)的缺失数据填补方法。采用偏最小二乘法(partial least square, PLS)提取历史数据主成份,计算各历史样本对主成份的累积贡献率(accumulative contribution rate, ACR),并绘制  $T^2$  椭圆,从而识别出历史样本贡献率过大的异常数据。用最小二乘支持向量机拟合历史数据变化趋势,从而实现缺失数据的填补。算例结果表明:  $T^2$  椭圆图能有效识别历史数据中的异常样本;最小二乘支持向量机具有良好的数据填补特性,具有较强的实用价值。

**关键词:** 数据异常; 数据缺失; 累积贡献率;  $T^2$  椭圆; 最小二乘支持向量机; 负荷预测

## 0 引言

中长期负荷预测是根据负荷、经济、社会等历史数据,探索电力负荷历史数据变化规律对未来的影响,挖掘电力负荷与其它相关因素的内在联系,从而构建负荷预测模型,实现对若干年后负荷发展水平的预测。预测精度直接关系到电网规划的质量和年度检修计划、运行方式的制定。

历史数据是预测工作的基础,其准确与否、完整与否,对建立的负荷预测模型产生重要影响。然而,在中长期负荷预测的历史数据收集过程中常常出现数据异常和数据缺失的情况<sup>[1-4]</sup>。

数据异常主要分2类:一类是由于经济政策或其它原因产生的异常负荷,这类数据体现了负荷变化的历史特性,必须保留;另一类是由于统计口径不一致或手误错看等人为因素形成的,这种样本应尽可能纠正。而数据缺失主要是由于数据管理人员的变动更迭以及数据保管部门的疏忽造成的。小样本是中长期负荷历史数据的一大特点,如果抛弃含有缺失数据的历史样本点,就会大大削减建立负荷预测模型的可用信息,降低模型的预测性能。因此,对于缺失数据应采取合理的填补措施补充完整。

## 1 异常数据的辨识

### 1.1 基本思想

每个地区都有逐年的经济社会发展统计年鉴,此类数据具有很高的可信度,因此可以考虑利用偏最小二乘法(partial least square, PLS)提取相关经济

发展指标和每年用电量水平的主成分进行分析。由于异常数据对回归模型具有明显的拉伸作用,对成分的贡献率将大大超过正常数据,所以可通过样本数据对成分的贡献率来辨识异常数据。

### 1.2 基于 $T^2$ 椭圆图的异常数据识别

由于篇幅限制,基于 PLS 的主成分提取方法不再赘述<sup>[5-7]</sup>。设从  $n$  组历史数据中共提取  $m$  个主成分,定义  $T_{hi}^2$  为第  $i$  个样本点对第  $h$  成分  $t_h$  的贡献率,则

$$T_{hi}^2 = \frac{t_{hi}^2}{(n-1)s_h^2} \quad (1)$$

式中  $s_h^2$  是主成分  $t_h$  的方差。由此,可计算出样本点  $i$  对成分  $t_1, \dots, t_m$  的累计贡献率  $T_i^2$ , 即

$$T_i^2 = \frac{1}{n-1} \sum_{h=1}^m \frac{t_{hi}^2}{s_h^2} \quad (2)$$

通常,  $T_i^2$  不宜过大,  $T_i^2$  过大会对预测模型造成明显的拉动。利用 Tracy 等人证明的统计量  $\lambda$ <sup>[8]</sup>, 即

$$\lambda = \frac{n^2(n-m)}{m(n^2-1)} T_i^2 \sim F_\alpha(m, n-m) \quad (3)$$

可以得到, 当

$$T_i^2 \geq \frac{m(n^2-1)}{n^2(n-m)} F_\alpha(m, n-m) \quad (4)$$

时, 可以认为样本点  $i$  对主成分的贡献过大。式(4)中  $\alpha$  显著水平。通常 2 个主成分即可概括数据的绝大部分信息, 所以为辨识方便, 不妨设定提取的主成分个数为 2。当  $m=2$  时, 判别条件为

$$\left( \frac{t_{1i}^2}{s_1^2} + \frac{t_{2i}^2}{s_2^2} \right)^2 \geq \frac{2(n^2-1)(n-1)}{n^2(n-2)} F_\alpha(2, n-2) \quad (5)$$

记

$$c = \frac{2(n^2-1)(n-1)}{n^2(n-2)} F_\alpha(2, n-2) \quad (6)$$

则

$$\frac{t_{1i}^2}{s_1^2} + \frac{t_{2i}^2}{s_2^2} = c \quad (7)$$

式(7)给出的椭圆方程即为  $T^2$  椭圆图<sup>[8]</sup>。在主成分平面图中绘制  $T^2$  椭圆, 如果所有样本点均落在  $T^2$  椭圆内远离椭圆界限, 则认为负荷历史数据的分布是均匀的, 没有异常数据。如果样本点落在椭圆之外或靠近椭圆边线, 则可认为该样本点是异常点。

对于通过  $T^2$  椭圆图检测出的异常负荷数据, 需要对其进行重新核实, 去掉人为因素造成的数据异常, 但对于数据自身的突变负荷数据, 则无需改变。

### 1.3 基于稳健回归分析的异常数据处理

以误差平方和最小为目标函数的经典最小二

乘回归是中长期负荷预测常用的预测工具。因方差为一不稳健统计量, 故最小二乘回归是一种不稳健的预测方法。如果历史数据存在突变负荷, 目标函数平方项会使预测模型明显向突变数据方向拉伸, 影响模型拟合效果, 甚至使整个预测模型发生畸变。

稳健回归是稳健估计中的一种方法, 其主要思路是对异常值十分敏感的经典最小二乘回归中的目标函数进行修改, 从而降低突变负荷对预测模型的影响<sup>[9-10]</sup>。不同的目标函数定义了不同的稳健回归方法, 本文仅介绍一种基于最小化残差绝对值和的稳健回归预测模型。该稳健回归的目标函数为

$$\min S(\beta) = \sum_{i=1}^n |Y_i - X_i' \beta| = \sum_{i=1}^n |\varepsilon_i| \quad (8)$$

式中:  $Y_i$  和  $X_i$  分别为负荷值和相关影响因素;  $\beta$  为待求估计量;  $\varepsilon_i$  为各历史数据点的拟合误差。为求目标函数值, 引进 2 个变量  $\mu_i$  和  $\nu_i$ , 其中

$$\begin{cases} \mu_i = (|\varepsilon_i| + \varepsilon_i) / 2 \\ \nu_i = (|\varepsilon_i| - \varepsilon_i) / 2 \end{cases} \quad (9)$$

则

$$\begin{cases} |\varepsilon_i| = \mu_i + \nu_i \\ \varepsilon_i = \mu_i - \nu_i \end{cases} \quad (10)$$

将式(8)转化为如下线性规划模型进行求解:

$$\min S = \sum_{i=1}^n (\mu_i + \nu_i) \quad (11)$$

$$\text{s.t.} \begin{cases} (a' - a'') + (b' - b'') X_i - \mu_i - \nu_i = Y_i \\ \mu_i \geq 0 \\ \nu_i \geq 0 \\ a' \geq 0, a'' \geq 0, b' \geq 0, b'' \geq 0 \end{cases}$$

由式(11)可知, 基于最小化残差绝对值和的稳健回归预测模型采用一次函数作其目标值, 对异常负荷数据的敏感程度低。将该模型应用于存在异常数据的历史样本中, 能取得较好的预测效果。

## 2 缺失数据的处理

### 2.1 最小二乘支持向量机的原理

最小二乘支持向量机 (least square support vector machine, LSSVM) 是标准支持向量机的扩展, 用等式约束代替 SVM 的不等式约束, 将支持向量机 (support vector machine, SVM) 中的二次规划问题转化为线性方程组求解, 能够较好地解决小样本、非线性、高维数和局部极小等实际问题<sup>[11-17]</sup>。

假设训练样本为

$$D = \{(\mathbf{x}_i, y_i) | i = 1, 2, \dots, l\} \quad (12)$$

式中： $\mathbf{x}_i$  为输入量， $\mathbf{x}_i \in \mathbf{R}^l$ ； $y_i$  为输出量， $y_i \in \mathbf{R}$ 。

确定一个基于训练集的函数，即

$$f(\mathbf{x}) = \boldsymbol{\omega}\mathbf{x} + \mathbf{b} \quad (13)$$

式中： $\boldsymbol{\omega}$  为超平面的法向量； $\mathbf{b}$  为偏置，用它来逼近未知的回归函数。

对于非线性问题，可以通过非线性变换  $\varphi(\cdot)$  将输入向量映射到高维特征空间，转化成线性问题分析解决。按结构最小化原理，LSSVM 的目标函数为

$$\min \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 \quad (14)$$

s.t.  $y_i = \boldsymbol{\omega}\varphi(\mathbf{x}_i) + \mathbf{b} + e_i, i = 1, 2, \dots, l$

式中： $\gamma$  为正则化参数； $e_i$  为样本误差。

对 LSSVM，定义 Lagrange 函数为

$$L(\boldsymbol{\omega}, \mathbf{b}, \mathbf{e}, \mathbf{a}) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 - \sum_{i=1}^l a_i^2 |\boldsymbol{\omega}\varphi(\mathbf{x}_i) + \mathbf{b} + e_i - y_i| \quad (15)$$

根据优化条件

$$\begin{cases} \partial L / \partial \boldsymbol{\omega} = 0 \\ \partial L / \partial \mathbf{b} = 0 \\ \partial L / \partial e_i = 0 \\ \partial L / \partial a_i = 0 \end{cases}$$

得到

$$\begin{bmatrix} 0 & \mathbf{E}^T \\ \mathbf{E} & \mathbf{Z}\mathbf{Z}^T + \gamma^{-1}\mathbf{E} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} \quad (16)$$

式中： $\mathbf{E} = [1, 1, \dots, 1]^T$ ； $\mathbf{Z} = [\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_n)]^T$ 。

解上述线性方程组求  $\mathbf{b}$  和  $\mathbf{a}$ ，得到的最小二乘支持向量回归机函数为

$$f(\mathbf{x}) = \sum_{i=1}^l a_i [\varphi(\mathbf{x})\varphi(\mathbf{x}_i)] + \mathbf{b} \quad (17)$$

式(17)用输入空间的一个核函数  $K(\mathbf{x}, \mathbf{x}_i)$  等效高维空间的内积形式，解决了高维计算问题，所以回归函数为

$$f(\mathbf{x}) = \sum_{i=1}^l a_i K(\mathbf{x}, \mathbf{x}_i) + \mathbf{b} \quad (18)$$

## 2.2 基于 LSSVM 的缺失数据的填充

历史数据少是中长期负荷预测的一大特点，所以每个历史负荷样本都对预测模型非常重要。当历史数据存在缺失时，如果不考虑这些缺失数据而建立负荷预测模型，就难以有效地模拟负荷发展变化规律，模型的预测精度也同样不能得到保证。鉴于此，对中长期负荷预测中缺失样本数据的合理填充具有重要意义。

通过归一化处理原始数据，除缺失数据外，将其余年份的社会经济发展指标做为模型输入，并将负荷值作为输出进行训练。训练完成后，根据缺失负荷年份的社会经济指标对其负荷值进行填补。

## 3 算例结果与分析

### 3.1 异常数据辨识算例

湖南省某县 1990—2005 年社会经济发展指标及全社会用电量见表 1。图中：不带“\*”的数据为历史拟合数据；带“\*”的数据为负荷预测模型检测数据。先采用 T<sup>2</sup> 椭圆图对其历史样本的异常数据进行辨识分析，再构建预测模型进行预测。

表 1 湖南省某县 1990—2005 年社会经济指标及全社会用电量

Tab. 1 Social economic factors and annual electric load of some Hunan county from 1990 to 2005

年份	社会经济指标				用电量/ 万 kW·h
	第一产业/ 万元	第二产业/ 万元	第三产业/ 万元	人均/ 元	
1990	32 221	14 733	17 229	948	12 211
1991	33 017	16 161	18 043	989	13 985
1992	35 414	17 972	20 660	1 086	16 024
1993	45 265	24 343	25 076	1 385	17 622
1994	62 507	31 783	33 364	1 860	20 407
1995	79 166	43 670	47 671	2 471	25 216
1996	91 329	56 248	58 407	2 985	21 377
1997	93 813	70 764	70 601	3 408	21 276
1998	96 279	79 775	79 289	3 627	21 778
1999	96 184	84 517	86 434	3 883	22 558
2000	98 461	87 816	95 667	4 098	21 979
2001	105 418	101 059	109 202	5 214	22 774
2002	107 900	114 134	124 721	5 705	29 607
2003	121 423*	131 384*	141 570*	6 458*	29 387*
2004	157 550*	157 363*	168 336*	7 904*	30 584*
2005	172 931*	191 988*	189 333*	9 244*	34 139*

根据表 1 的数据，采用偏最小二乘法对用电量数据提取主成分，结果见表 2，其中  $u_1$ 、 $u_2$  分别代表第 1 主成分和第 2 主成分。为叙述简便，将 1990—2002 年的历史数据依次编号为 1~13。

表 2 主成分  $u_1$ 、 $u_2$  的计算结果

Tab. 2 Results of components  $u_1$  and  $u_2$

历史数据序号	$u_1$	$u_2$
1	-1.788 5	-0.696 1
2	-1.406 9	-0.339 3
3	-0.968 2	0.042 2
4	-0.624 4	0.207 2
5	-0.025 2	0.519 2
6	1.009 4	1.191 3
7	0.183 5	0.066 3
8	0.161 8	-0.188 9
9	0.269 8	-0.233 3
10	0.437 6	-0.168 9
11	0.313 0	-0.412 5
12	0.484 0	-0.602 3
13	1.954 1	0.614 8

根据表1,按式(2)计算各历史样本对成分的累积贡献率(accumulative contribution rate, ACR),如表3所示。

表3 各历史样本对主成分的 ACR  
Tab.3 ACR of historical load data

$N$	1	2	3	4	5	6	7
$T_i$	0.410 5	0.199 1	0.078 6	0.045 2	0.080 1	0.506 4	0.004 1
$N$	8	9	10	11	12	13	—
$T_i$	0.012 8	0.022 2	0.024 4	0.058 7	0.127 3	0.430 5	—

由表3可知,样本1、6、13(1990年、1995年和2002年数据)的累积贡献率分别为0.4105、0.5064和0.4305,明显大于其余历史样本。

根据历史数据,取显著水平 $\alpha$ 为0.15,因此,可按式(7)计算出 $T^2$ 椭圆方程,即

$$\frac{u_1^2}{1} + \frac{u_2^2}{0.2828} = 5.004 \quad (19)$$

根据表2绘制 $u_1$ 和 $u_2$ 主成分图,并在图上绘制出式(19)的 $T^2$ 椭圆图,如图1所示。

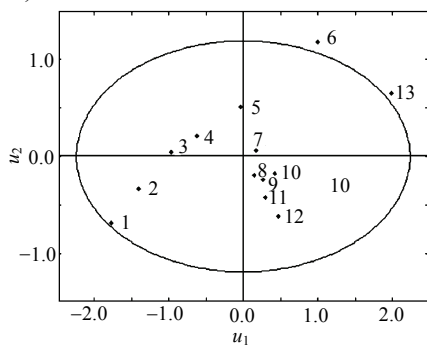


图1 原始历史数据的 $T^2$ 椭圆图

Fig. 1  $T^2$  ellipse map of original historical data

由图1可以发现,样本6和13(1995年和2002年的数据)是异常数据,而样本1(1990年的数据)也位于 $T^2$ 椭圆的边缘,需要重新核实。经核实发现,1995年和2002年的实际电量数据分别为21256和26907万kW·h;1990年的数据无误,属于突变负荷。文献[3]以负荷均值的 $\pm 20\%$ 作为判定依据。以此标准,在区间(16419, 24629)内的历史数据为有效数据,尽管也能识别出1990年、1995年、2002年的异常负荷值,但同时也将1991年、1992年正常的历史负荷误判为异常数据。

经过核实后重新计算 $T^2$ 椭圆方程,即

$$\frac{u_1^2}{1} + \frac{u_2^2}{0.1664} = 5.004 \quad (20)$$

重新绘制 $u_1$ 和 $u_2$ 主成分图、 $T^2$ 椭圆图,如图2所示。

对比图1和图2可以发现,核实数据后,样本5和13均出现在 $T^2$ 椭圆图内部,这说明已纠正人

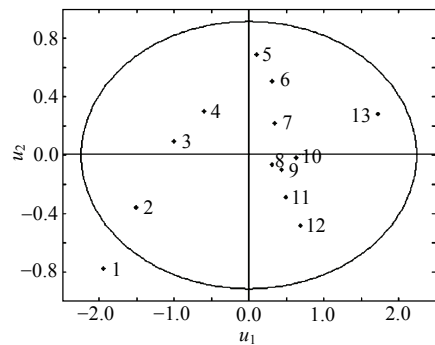


图2 核实后历史数据的 $T^2$ 椭圆图

Fig. 2  $T^2$  ellipse map of historical data after verifying

为的异常数据。而1990年的历史数据是突变负荷,所以仍然在椭圆外部。如果按负荷均值 $\pm 20\%$ 准则再次判断核实后的数据,则1990年、1991年和2002年将显示数据异常。

$T^2$ 椭圆图辨识方法与传统均值 $\pm 20\%$ 准则在数据核实前后对异常值的识别效果见表4。

表4  $T^2$ 椭圆图与传统方法的辨识效果

Tab.4 Identification results of  $T^2$  ellipse and traditional method

异常数据核实前后	异常值个数		
	实际	$T^2$ 椭圆识别	$\pm 20\%$ 准则识别
核实前	3	3	5
核实后	1	1	3

由表4可知,由于各地区的负荷发展速度相差较大,用传统的 $\pm 20\%$ 准则来识别历史数据中的负荷值具有一定的局限性。

对于突变负荷,采取普通的线性回归法拟合负荷发展变化。突变负荷的影响会降低预测模型的预测精度。采用稳健回归法进行模型预测,由于该方法对突变负荷的敏感程度低,故能有效提高预测效果。

为体现异常值对整个负荷模型的影响,先以未纠正的历史数据建立传统的最小二乘回归预测模型,记为 $f_1(x)$ ,再以辨识核实后的历史样本构建稳健回归模型,记为 $f_2(x)$ ,分别对2003—2005年的全社会用电量进行预测。

传统最小二乘回归模型为

$$f_1(x) = 8329.87 + 0.1658x_1 - 0.4334x_2 + 0.2315x_3 + 3.6239x_4 \quad (21)$$

根据式(11)构建线性规划模型,得到的稳健回归预测模型为

$$f_2(x) = 10102 + 0.1125x_1 - 0.1212x_2 + 0.006x_3 + 2.7747x_4 \quad (22)$$

式(21)(22)中: $f_1(x)$ 和 $f_2(x)$ 代表全社会用电量; $x_1$ 、 $x_2$ 、 $x_3$ 和 $x_4$ 分别代表第一、第二、第三产业产值和

人均 GDP。

利用式(21)(22)分别对 2003—2005 年的全社会用电量进行预测, 结果见表 5。

表 5 基于最小二乘回归模型和与稳健回归模型得到的预测结果

**Tab. 5 Forecasting results based on least square regression model and robust regression model**

年份	实际值/万 kW·h	$f_1(x)$ 预测值/万 kW·h	$f_2(x)$ 预测值/万 kW·h
2003	29 387	27 697	26 607
2004	30 584	33 864	31 695
2005	34 139	31 124	33 073

以表 5 数据为基础, 现以均方根误差作为以上 2 种模型的评价指标, 则最小二乘回归模型的均方根误差为 2 750.92 万 kW·h; 稳健回归模型的均方根误差为 1 834.89 万 kW·h, 这说明通过对异常值进行辨识纠正, 并建立稳健回归负荷预测模型, 预测精度可显著提高。

### 3.2 缺失数据填补算例

湖南省某市 1996—2005 年社会经济发展指标及全社会用电量见表 6。表中假定 1999 年和 2003 年的全社会用电量为缺失值, 用“—”表示。

表 6 湖南省某市 1996—2005 年社会经济指标及全社会用电量

Tab. 6 Social economic factors and annual electric load of a city in Hunan province from 1996 to 2005

年份	第一产业/ 万元	第二产业/ 万元	第三产业/ 万元	用电量/ 万 kW·h
1996	784 300	693 500	657 200	282 138
1997	842 200	795 700	772 700	283 818
1998	904 300	913 000	908 500	304 062
1999	971 100	1 047 600	1 068 200	—
2000	1 042 800	1 201 900	1 256 200	344 625
2001	1 048 500	1 363 500	1 450 200	378 816
2002	1 020 000	1 335 100	1 626 000	415 165
2003	1 051 400	1 573 400	1 829 800	—
2004	1 360 200	1 947 500	2 104 600	544 120
2005	1 497 100	2 273 500	2 130 800	603 652

以第一、二、三产业作为 LSSVM 的输入, 以全社会用电量作为输出, 选取如下归一化公式:

$$x'_{ij} = \frac{x_{ij} - \max x_j}{\max x_j - \min x_j} \quad (23)$$

式中:  $x_{ij}$  代表历史数据;  $x'_{ij}$  代表归一化值;  $x_j$  代表  $x_{ij}$  所在的第  $j$  列,  $\max$  和  $\min$  分别代表取最大和最小值。

径向基函数(radial basis function, RBF)因具有优秀的局部逼近特性应用得最为广泛, 所以选取径向基函数作为 LSSVM 的核函数。

归一化后选取除 1998 年和 2003 年以外的其余

8 年数据作为训练样本, 对 LSSVM 进行训练, 并用训练好的模型对 1999 年和 2003 年的用电量进行拟合填补, 拟合结果见表 7。

表 7 采用 LSSVM 得到的拟合结果  
Tab. 7 Fitting results by LSSVM

年份	真实值/万 kW·h	拟合值/万 kW·h	相对误差/%
1996	282 138	282 105	-0.011 4
1997	283 818	283 924	0.037 3
1998	304 062	303 983	-0.025 8
1999	—	327 462	—
2000	344 625	34 466	0.010 8
2001	378 816	378 819	0.000 9
2002	415 165	415 144	-0.004 9
2003	—	454 163	—
2004	544 120	544 123	0.000 7
2005	603 652	603 632	-0.003 2

由表 7 可知, LSSVM 对训练历史负荷具有良好的拟合效果, 并通过 LSSVM 对 1999 年和 2003 年全社会的用电量缺失进行填补, 填补的结果分别为 327 462 和 454 163 万 kW·h。上述 2 年的实际用电量分别为 307 772 和 481 841 万 kW·h, 填补的相对误差绝对值仅分别为 6.4% 和 5.7%, 这说明 LSSVM 具有良好的填补效果。

## 4 结论

1) 本文利用  $T^2$  椭圆图对中长期负荷预测中的异常历史数据进行辨识分析, 对含有突变负荷的数据样本建立稳健回归负荷预测模型, 并通过 LSSVM 对历史样本中的缺失数据进行了填补。算例结果表明: 根据  $T^2$  椭圆图对历史样本 ACR 进行比较, 能有效识别出负荷历史数据中的异常值; 采用稳健回归建立负荷预测模型能降低突变负荷影响, 提高预测精度; LSSVM 具有良好的泛化性能, 对缺失数据有良好的填补效果。

2) 本文的方法均以相关变量, 即社会经济指标可信为前提条件, 如何不依赖相关变量而单从负荷数据自身挖掘规律, 进行数据异常辨识和缺失数据填补, 是今后的研究重点。

## 参考文献

- [1] 牛东晓, 曹树华, 赵磊, 等. 电力负荷预测技术及其应用[M]. 北京: 中国电力出版社, 1998: 7-8.
- [2] 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: 12-13.
- [3] 韦钢, 王飞, 张永健, 等. 负荷预测中历史数据缺损处理[J]. 电力科学与工程, 2004(1): 16-19.  
Wei Gang, Wang Fei, Zhang Yongjian, et al. Methods of processing shortage and damage of historical data in load forecasting[J]. Electric Power Science and Engineering, 2004(1): 16-19(in Chinese).

- [4] 陈泽淮, 张尧, 武志刚. RBF神经网络在中长期负荷预测中的应用[J]. 电力系统及其自动化学报, 2006, 18(1): 15-19.  
Chen Zehuai, Zhang Yao, Wu Zhigang. Application of RBF neural network in medium and long-term load forecasting[J]. Proceedings of the CSU-EPSSA, 2006, 18(1): 15-19(in Chinese).
- [5] 王文圣, 丁晶, 赵玉龙, 等. 基于偏最小二乘回归的年用电量预测研究[J]. 中国电机工程学报, 2003, 23(10): 17-21.  
Wang Wensheng, Ding Jing, Zhao Yulong, et al. Study on the long term prediction of annual electricity consumption using partial least square regressive model[J]. Proceedings of the CSEE, 2003, 23(10): 17-21(in Chinese).
- [6] 毛李帆, 江岳春, 龙瑞华, 等. 基于偏最小二乘回归分析的中长期电力负荷预测[J]. 电网技术, 2008, 32(19): 71-77.  
Mao Lifan, Jiang Yuechun, Long Ruihua, et al. Medium and long term load forecasting based on partial least squares regression analysis [J]. Power System Technology, 2008, 32(19): 71-77(in Chinese).
- [7] 张伏生, 汪鸿, 韩梯, 等. 基于偏最小二乘回归分析的短期负荷预测[J]. 电网技术, 2003, 27(3): 36-40.  
Zhang Fusheng, Wang Hong, Han Ti, et al. Short-term load forecasting based on partial least-squares regression[J]. Power System Technology, 2003, 27(3): 36-40(in Chinese).
- [8] 王惠文. 偏最小二乘回归方法及应用[M]. 北京: 国防工业出版社, 1999: 184-185.
- [9] 童恒庆. 计量经济学[M]. 北京: 科学出版社, 2007: 566-567.
- [10] 杨桂元. 残差绝对值之和最小准则下回归方程的求法[J]. 运筹与管理, 2001, 10(1): 20-23.  
Yang Guiyuan. Solution to regression equations on the rule of minimum sum of absolute errors[J]. Operations Research and Management Science, 2001, 10(1): 20-23(in Chinese).
- [11] 杨延西, 刘丁. 基于小波变换和最小二乘支持向量机的短期电力负荷预测[J]. 电网技术, 2005, 29(13): 60-64.  
Yang Yanxi, Liu Ding. Short-term load forecasting based on wavelet transform and least square support vector machines[J]. Power System Technology, 2005, 29(13): 60-64(in Chinese).
- [12] 耿艳, 韩学山, 韩力. 基于最小二乘支持向量机的短期负荷预测[J]. 电网技术, 2008, 32(18): 72-76.  
Geng Yan, Han Xueshan, Han Li. Short-term load forecasting based on least squares squares support vector machines[J]. Power System Technology, 2008, 32(18): 72-76(in Chinese).
- [13] 李元诚, 方廷建, 于尔铿. 短期负荷预测的支持向量机方法研究[J]. 中国电机工程学报, 2003, 26(3): 55-59.  
Li Yuancheng, Fang Tingjian, Yu Erkeng. Study of support vector machines for short-term load forecasting[J]. Proceedings of the CSEE, 2003, 26(3): 55-59(in Chinese).
- [14] 谢宏, 魏江平, 刘鹤立. 短期负荷预测中支持向量机模型的参数选取和优化方法[J]. 中国电机工程学报, 2006, 26(22): 17-22.  
Xie Hong, Wei Jiangping, Liu Heli. Parameter selection and optimization method of SVM model for short-term load forecasting [J]. Proceedings of the CSEE, 2006, 26(22): 17-22(in Chinese).
- [15] 赵登福, 王蒙, 张讲社, 等. 基于支持向量机方法的短期负荷预测[J]. 中国电机工程学报, 2002, 22(4): 26-30.  
Zhao Dengfu, Wang Meng, Zhang Jianshe, et al. A support vector machines approach for short-term load forecasting[J]. Proceeding of the CSEE, 2002, 22(4): 26-30(in Chinese).
- [16] 张林, 刘先珊, 阴和俊. 基于时间序列的支持向量机在负荷预测中的应用[J]. 电网技术, 2004, 28(10): 38-41.  
Zhang Lin, Liu Xianshan, Yin Hejun. Application of support vector machines based on time sequence in power system load forecasting [J]. Power System Technology, 2004, 28(10): 38-41(in Chinese).
- [17] 李波, 顾冲时, 李智录, 等. 基于偏最小二乘回归和最小二乘支持向量机的大坝涌流监控模型[J]. 水利学报, 2008, 39(12): 1390-1394.  
Li Bo, Gu Chongshi, Li Zhilu, et al. Monitoring model for dam seepage based on partial least-squares regression and partial least square support vector machine[J]. Journal of Hydraulic Engineering, 2008, 39(12): 1390-1394(in Chinese).



毛李帆

收稿日期: 2010-03-03.

作者简介:

毛李帆(1982—), 男, 博士研究生, 主要从事电力系统自动化方面的研究工作, E-mail: sail\_1234@163.com;

姚建刚(1952—), 男, 教授, 博士生导师, 主要从事电力系统自动化和电力市场方面的教学与科研工作。

(责任编辑 杜宁)