

文章编号:1671-6833(2010)03-0029-04

半参数回归分析法在电力负荷预测中的应用

杨丽徙¹, 张永锋¹, 许向伟², 冯越³

(1. 郑州大学 电气工程学院, 河南 郑州 450001; 2. 遂平县电业公司, 河南 遂平 463100; 3. 驻马店供电公司, 河南 驻马店 463400)

摘要: 为了融合参数模型和非参数模型的优点, 对半参数回归模型在电力负荷预测中的应用进行了研究. 在原始数据预处理的基础上, 运用相关性分析理论确定了半参数模型中的参数与非参数变量, 构建了半参数回归模型; 利用两阶段最小二乘法 and 权重系数法对回归系数和回归参数进行了估计. 算例结果表明, 半参数回归分析法具有较高的负荷预测精度和较广泛的适用范围.

关键词: 半参数回归模型; 两阶段最小二乘法; 权重系数法; 参数; 非参数

中图分类号: TM715 **文献标识码:** A

0 引言

负荷预测是电力系统运行管理与建设发展的基础工作, 也是长期以来的热点问题.

目前国内外研究的方法主要包括参数模型与非参数模型两大类. 前者包括基于状态空间理论、回归分析、神经网络、灰色理论、支持向量机和最优组合预测方法等^[1-4], 基本上都属于参数统计法的范畴, 因变量对自变量有较强的依赖关系, 当假设函数模型成立时, 预测精度较高, 当假设函数不成立时, 预测模型的拟合情况和预测精度都是不理想的. 后者降低了自变量对因变量的限制, 有较大的适应性, 但也可能会失去历史资料所提供的信息, 降低模型的解释能力.

半参数回归模型理论是20世纪80年代发展起来的统计模型^[5], 在交通、医疗、能源和人口等方面都有较广泛的应用^[6-9]. 它兼顾了参数和非参数回归方法型的优点, 集中了主要部分(参数部分)的信息, 又不忽略干扰项(非参数部分)的作用, 有着更强的适应性, 因此, 将半参数回归模型应用于负荷预测, 可以提高预测精度, 扩大预测模型的适用范围.

1 半参数回归模型

假设: y_j 是因变量, x_j 是参数部分自变量, z_j 是

非参数部分自变量, β 是回归系数, 即待求参数, g 是未知函数, ε_j 是随机误差, ε_j 相互独立且服从标准正态分布. 则有

$$y_j = x_j\beta + g(z_j) + \varepsilon_j, (j = 1, 2, \dots, n) \quad (1)$$

式中: $x_j\beta$ 反映了负荷预测可知的部分规律, 把握因变量 y 变化的趋势走向, 为参数部分; $g(z_j) + \varepsilon_j$ 反映了负荷预测的不确定因素的影响, 为非参数部分.

2 两阶段最小二乘估计

半参数模型中参数和函数估计的常用方法有: 两阶段最小二乘估计^[3]、偏残差估计^[5]、 P -范数估计^[10]、最小熵估计^[11]等. 两阶段最小二乘估计法的步骤如下.

2.1 模型的标准化处理

假设 $E(g(z_j))$ 、 $E(g(z_j))^2 < \infty$ 存在有限, 令 $a = E(g(z_j))$, 则:

$$e_j = g(z_j) - a + \varepsilon_j, (j = 1, 2, \dots, n) \quad (2)$$

e_1, e_2, \dots, e_n 相互独立且服从标准正态分布.

将(2)代入模型(1)中, 可将(1)转化为标准的线性回归模型:

$$y_j = x_j\beta + a + e_j, (j = 1, 2, \dots, n) \quad (3)$$

2.2 拟合权重的生成

该拟合权重反映了样本数据对于预测值的影响程度, 样本数据离所预测负荷的时刻越近, 权重

收稿日期: 2010-01-21; 修订日期: 2010-03-12

基金项目: 河南省教育厅自然科学基金资助项目(2009A470008)

作者简介: 杨丽徙(1956-), 女, 江苏盱眙人, 郑州大学教授, 博士, 研究方向为电力系统规划与可靠性, E-mail: yanglixixi@zzu.edu.cn.

越大;反之,权重越小。

由样本数据 (z_j, y_j) 建立线性回归模型:

$$y_j = bz_j + \varepsilon_j \quad (4)$$

由最小二乘法求得 b 之后,求得逐点残差和残差平方:

$$\varepsilon_j = y_j - bz_j, 1 \leq j \leq n \quad (5)$$

$$h_j = \varepsilon_j \times \varepsilon_j, 1 \leq j \leq n \quad (6)$$

ε_j 和 h_j 值越大,回归模型拟合效果越差;反之,回归模型拟合效果越好。

借鉴方差分析中综合模型的权重系数的求法,可以得到拟合权重 W_j ,拟合误差越小,权重越大。

$$W_j = \frac{1}{h_j \times \sum_{i=1}^n \frac{1}{h_i}}, 1 \leq j \leq n \quad (7)$$

其中, n 是样本数据的个数。

拟合权重 W_j 为可测数据,应满足以下条件:

$$W_j \geq 0, \forall 1 \leq j \leq n \quad (8)$$

$$\sum W_j = 1, \forall n \in N \quad (9)$$

2.3 回归系数的两阶段估计

根据已知数据 (x_j, y_j) 对式(3)进行最小二乘回归分析,得到 β 和 α 的初次估计值 β' 和 α' 。则回归模型式(1)可以转化为

$$y_j - x_j\beta' - \alpha' = g(z_j) + \varepsilon_j, (j = 1, 2, \dots, n) \quad (10)$$

$$g(z_j) = g(z_j, \beta') = \sum_{i=1}^{j-1} W_i (y_i - x_i\beta' - \alpha'), \quad (j = 1, 2, \dots, n) \quad (11)$$

其中, W_i 是基于样本数据的拟合权重。

将式(11)代入式(1)中,得到:

$$y_j - g(z_j) = x_j\beta + \alpha + \varepsilon_j, 1 \leq j \leq n \quad (12)$$

由模型(12),利用最小二乘法得 β 和 α 的最终估计值 β'' 和 α'' ,则半参数回归预测模型为

$$y_j = \beta''x_{ij} + \alpha'' + \sum_{i=1}^{j-1} W_i (y_i - \beta'x_{ii} - \alpha') \quad (13)$$

欲进行 $j+1$ 时刻的负荷预测,则应将 j 时刻的预测值加入到原始数据中,重新计算非参数部分和 β'' 与 α'' ,然后利用式(13)对下一时刻的负荷进行预测,以此循环下去。

3 自变量的处理和选择

影响电力负荷变化的因素很多,主要分为两大类,一类是受国民经济发展、长期气候变化、行政调控政策等因素决定确定性成分;一类是由偶然因素导致的、不确定的,在时间上又表现为一

定的相依特性的随机成分。要准确预测负荷,就必须分别研究这两个不同性质的成分,因此在建立预测模型时必须考虑不确定性因素,同时兼顾确定性的规律,这样所建立的模型才具有较好的精度和较强的适应性。

3.1 数据预处理

(1)“异常数据”的存在给正常数据带来较大的干扰,影响电力负荷预测的精度,因此有必要对“异常数据”进行剔除和处理。

(2)如果原始数据中有数据缺失,要应用删除法、插补法等^[12]对数据进行缺失补遗,以减小由于数据缺失可能造成的估计偏差,保证数据的完整性。

3.2 自变量的选择

(1)自变量的聚类分析。对自变量之间的相关系数 r_{ij} 进行计算, r_{ij} 越大,变量之间的线性相关程度越高;利用一次形成聚类法对自变量进行聚类,当自变量之间的相关系数大于0.9时,把这些自变量归为一类。

$$r_{ij} = \frac{\text{COV}(X_i, X_j)}{\sqrt{D(X_i)} \sqrt{D(X_j)}} \quad (14)$$

式中: $\text{COV}(X_i, X_j)$ 为 X_i 与 X_j 之间的协方差; $D(X_i)$ 和 $D(X_j)$ 分别为 X_i 和 X_j 的方差。

(2)计算自变量与因变量之间的相关系数 r_{xy} ,计算公式同式(14)。

(3)选择模型中参数部分的自变量。在同一类自变量中选择相关系数最大且 $r_{xy} > 0.8$ 的变量为参数部分的自变量。

(4)选择模型中非参数部分的自变量。为了保持数据的完整性,将待选自变量中,除参数部分自变量以外的自变量中,相关系数 r_{xy} 介于(0.8 ~ 0.5)之间的自变量定义为非参数部分的自变量。

4 算例分析

以某年度某城市1980—2005年的统计数据为例^[13]。假设 y 表示该市社会总用电量, $x_1 \sim x_7$ 分别表示GDP、消费品零售总额、人均收入、固定资产投资、人均总可支配收入、人口数量和人均消费,原始数据略。

(1)数据的预处理。由已知数据可知各列数据没有异常数据,但在人均总可支配收入数据中,缺失1987和1997年的数据,可能影响电力负荷预测的准确性,所以采用插补法中的均值插补对缺失数据进行补遗,即

$$x_i = 0.5(x_{i-1} + x_{i+1}) \quad (15)$$

补遗后的数据分别是 1 827.39, 9 341.2.

(2) 自变量的聚类分析与选择. 利用(14) 可求得自变量 $x_1 \sim x_7$ 之间的相关系数 r_{ij} , 如表 1 所示.

表 1 自变量之间的相关系数 r_{ij}

Tab. 1 The correlation coefficients r_{ij} of inter-variables

变量	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1.000	0.991	0.991	0.970	0.997	0.798	0.987
x_2	0.991	1.000	0.998	0.981	0.990	0.837	0.998
x_3	0.991	0.998	1.000	0.980	0.992	0.838	0.999
x_4	0.970	0.981	0.980	1.000	0.963	0.822	0.985
x_5	0.997	0.990	0.992	0.963	1.000	0.810	0.989
x_6	0.798	0.837	0.838	0.822	0.810	1.000	0.851
x_7	0.987	0.998	0.999	0.985	0.989	0.851	1.000

由上表可知, 除 x_6 之外的自变量之间的相关系数都大于 0.9; 因此, 把自变量 $x_1 \sim x_5$ 和 x_7 归为一类; 利用(14) 计算因变量 y 与类中各自变量之间的相关系数, 选择相关系数最大的 x_1 表征这类变量, 因 y 与 x_1 相关系数为 0.993 大于给定阈值, 故定义 x_1 为参数部分自变量; y 与 x_6 相关系数为 0.792 小于给定阈值, 选择 x_6 作为非参数部分的影响变量.

用 1980—2002 年数据进行半参数回归建模, 分别对 2003、2004 和 2005 年的负荷进行预测, 并通过样本数据进行检验模型.

(3) 模型参数的初步估计. 以 x_1 为参数变量, 进行线性回归分析, 得到线性回归系数为 0.082 2, 回归常数为 178, 把它们分别作为 β 和 α 的估计初值 $\beta' = 0.082 2$ 与 $\alpha' = 178$, 得半参数回归模型:

$$y_j = 0.082 2x_{1j} + 178 + g(x_{6j}), 1 \leq j \leq n$$

(4) 拟合权重的求取. 首先以 x_6 为自变量, 进行线性回归分析, 得到线性回归系数为 2.398 4, 回归常数为 -2 601, 得线性回归模型: $y = 2.398 4x_6 - 2 601$; 然后根据公式(5) 和(6) 求得逐点残差和残差平方, 最后根据公式(7) 求得预测 2003 年负荷时的拟合权重 W_{24} (24 是 2003 年年度序号), 即:

$$W_{24} = [0.075 6, 0.036 4, 0.009 8, 0.001 3, 0.000 1, 0.003 4, 0.001 4, 0.037 9, 0.057 4, 0.102 8, 0.097 9, 0.071 8, 0.042, 0.027 2, 0.009 5, 0.001 4, 0.000 4, 0.006 5, 0.021 9, 0.025 8, 0.070 5, 0.103 5, 0.182 9]^T$$

(5) 模型参数的最终估计. 利用已知样本数据和公式(11) 进行非参数部分的估计, 求得非参数部分为

$$g(x_{6j}) = \sum_{i=1}^j W_i(x_{6j})(y_i - 0.082 2x_{1i} - 178), 1 \leq j \leq n.$$

然后可以得到数据列 $(x_{1j}, y_j - g(x_{6j}))$, 基于此数据列和公式(12) 利用最小二乘法可以得到 β 和 α 的最终估计值 β'' 与 α'' , 即:

$$y_j = 0.082 6x_{1j} + 177.3 + \sum_{i=1}^{j-1} W_i(y_i - 0.082 2x_{1i} - 178)$$

其中: j 为 2003 年年度序号 24, i 为 1980 年年度序号 1.

(6) 模型检验. 应用线性回归分析法和半参数回归分析法对样本数据进行建模、预测和拟合分析, 在进行 $j+1$ 时刻的负荷预测, 将 j 时刻的预测值或样本数据加入到原始数据中, 重新计算各相关系数, 然后利用(13) 对下一时刻的负荷进行预测, 以此循环下去. 拟合曲线如图 1 所示.

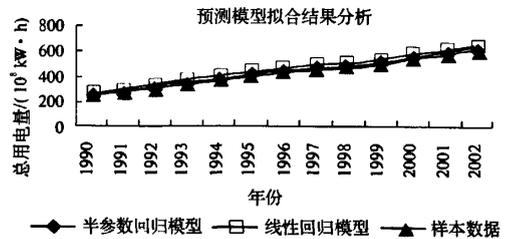


图 1 预测模型拟合分析图

Fig. 1 The fitting analysis diagram of forecast models

数据分析表明, 半参数回归模型的相对拟合误差范围为 0.035% ~ 3.443%, 平均相对误差为 1.256%; 线性回归模型的相对误差范围为 0.009% ~ 9.076%, 平均相对误差为 3.285%; 相比较而言, 两者的拟合相对误差大致趋势一致, 拟合值与样本数据大致趋势相同, 但半参数回归模型的拟合效果略好于线性回归分析模型.

(7) 预测结果分析. 根据公式(12) 和两阶段最小二乘算法计算得到 2003—2005 年的 β'' 和 α'' 分别为 0.082 6、0.081 3、0.080 6 与 177.3、179.2、180.3, 利用公式(13) 对负荷进行预测, 得到预测结果和相对误差如表 2 所示.

表 2 预测结果和相对误差比较

Tab. 2 The comparison of Results and errors

年份	预测值 / (10 ⁴ kW·h)		原始数据 / (10 ⁴ kW·h)	相对误差 / %	
	半参数	回归分析		半参数	回归分析
2003	737.50	733.23	745.97	1.14	1.71
2004	836.12	839.89	821.44	-1.79	-2.25
2005	929.85	931.62	908.97	-2.30	-2.49

由表2可知,距离样本数据时间越近,预测误差越小,预测精度越高;半参数回归模型在参数部分的基础上进行局部非参数部分调整,达到了提高精度的目的,本算例证明了半参数回归模型的有效性.

5 结论

(1)半参数回归模型对样本数据采用数据及时更新方式,可以更好地提高负荷预测精度.

(2)为了提高负荷预测的精度,通过相关系数进行自变量的聚类分析,将其处理为参数和非参数部分,计及相关因素更加全面.

(3)为了进一步提高半参数模型方法的精度,还可以进一步考虑把半参数模型和其它预测方法综合使用,构建组合预测模型.

参考文献:

- [1] 李翔,陈昊. 基于时变参数模型的中长期负荷预测[J]. 电力需求侧管理, 2009,11(2):32-34.
- [2] 尤焕苓,丁德平,王春华,等. 应用回归分析和BP神经网络方法模拟北京地区电力负荷预测[J]. 气象科技, 2008,36(6):801-805.
- [3] CUI H R. A long-term electrical power load forecasting model based on grey feed-back modification [C]//Proceedings of the seventh International conference on Machine Learning and Cybernetics. Kunming: Yunan University, 2008:2198-2201.
- [4] 张亚军,刘志刚,霍柏超,等. 基于支持向量机的电力负荷组合预测模型[J]. 电力需求侧管理, 2007,9(2):14-17.
- [5] 柴根象,洪圣岩. 半参数回归模型[M]. 安徽:安徽教育出版社, 1995.
- [6] 王如义,王慈光,郭改政,等. 铁路货物周转量的半参数回归模型预测[J]. 西南交通大学学报, 2008,43(1):96-100.
- [7] 赵林峰,韩国艳,李勇,等. 基于半参数分析的电力需求预测算法[J]. 中国安全科学学报, 2006,16(8):17-21.
- [8] 尹逊汝. 基于纵向数据半参数模型的AIDS疗法评价及预测[J]. 泰山学院学报, 2008,30(6):28-32.
- [9] 姜爱平,张德生,武新乾,等. 预测我国人口总量的具有外生变量的半参数自回归模型[J]. 河南科技大学学报, 2007,28(1):97-100.
- [10] PAN X. P-norm Semi-parametric regression model [C]//Machine Learning and Cybernetics. Harbin: Harbin University of Technology, 2007:2461-2466.
- [11] WOLWYNSIK R, THIERRY L R, PRPNZATO L. Minimum entropy estimation in semi-parametric models [C]//IEEE International Conference. New York:IEEE, 2004:ii-1045-1048.
- [12] 金勇进. 缺失数据的插补调整[J]. 数理统计与管理, 2001,20(5):47-53.
- [13] 上海统计局. 2006上海统计年鉴[M]. 北京:中国统计局出版社, 2006.

Application of Semi-parametric Regression Analysis Method in Electric Power Load Forecasting

YANG Li-xi¹, ZHANG Yong-feng¹, XU Hong-wei², FENG Yue³

(1. School of Electric Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. Suiping Electric Power Supply Company, Suiping 463100, China; 3. Zhumadian Electric Power Supply Company, Zhumadian 463400, China)

Abstract: In order to integrate the advantages of parameter model and non-parameter model, a semi-parametric regression model was studied for power load forecasting. On the basis of pretreatment of the original data, the use of relevant theory determined the parameters of semi-parametric model and non-parametric variables and constructed the semi-parametric regression model. The regression coefficients and regression parameters were assessed with the two-stage least squares method and the weight coefficients. The example shows that this method is more precise and has broader adaptability.

Key words: semi-parametric regression model; two-stage least squares method; weight coefficients; parametric; non-parametric