

基于因子和趋势分析反馈的多元回归负荷预测

胡怡霜¹, 夏翔², 丁一¹, 方建亮²

(1. 浙江大学 电气工程学院, 杭州 310007; 2. 国网浙江省电力有限公司, 杭州 310007)

A multivariate regression load forecasting algorithm based on factor and tendency analysis feedback

HU Yi-shuang¹, XIA Xiang², DING Yi¹, FANG Jian-liang²

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310007, China;

2. State Grid Zhejiang Electric Power Company, Hangzhou 310007, China)

摘要:提出一种基于因子和趋势分析反馈的考虑气象因素的多元回归预测模型,降低预测算法精度验证的计算负担。首先通过因子分析对气象和负荷指标进行相关性计算;其次,通过趋势分析量化历史负荷数据在时间尺度上的变化;然后,通过多元非线性回归模型,将气象指标作为自变量,负荷数据作为因变量进行函数拟合;最后,通过拟合出的历史预测数据与气象指标进行因子分析和趋势分析,并与第一步得到的数据进行动态相对误差计算,将动态误差反馈给多元回归预测模型。以某地区的实际负荷为例进行验证,结果表明该方法得到的预测精度高,且计算时间短。

关键词:趋势分析反馈;多元非线性回归;气象因素;负荷预测

Abstract: A weather concerning multivariate regression model combining with factor analysis feedback and tendency analysis feedback is proposed to reduce the counting burden in verifying accuracy of forecasting algorithm. Firstly, the relationship between weather and load indexes are evaluated by factor analysis; Secondly, a tendency analysis about time-varying load indexes is demonstrated; Thirdly, the forecasting equations are presented by multivariate nonlinear regression analysis; Lastly, in forecasting verification process, factor analysis and tendency analysis are applied in the forecasting data, and the error between the feedback of forecasting data and history data are treated as the accuracy verification principle. Numerical examples are given, presenting the accuracy and effectiveness of the proposed method.

Key words: tendency analysis feedback; multivariate nonlinear regression; weather factors; load forecast

中图分类号:F407.61;TM715 文献标志码:B

多年来,许多学者对负荷预测这一课题进行了深入研究,并提出了许多方法。文献[1]研究了神经网络算法在负荷预测中的应用,文献[2]将神经网络与线性回归结合,分析气象在负荷预测中的影响,文献[3]将支持向量机用于短期负荷预测,文献[4]在负荷预测中考虑温度修正效应,完善负荷预测模型,文献[5]采用小波分析考虑负荷预测,文献[6]基于配用电数据进行短期负荷预测。

这些文献在预测精度的验证上都面临着计算效率低的问题,尤其是在考虑气象指标的大数据负荷预测方面。传统的预测算法都是通过将历史数据代入负荷预测模型,得到历史预测数据,将其与实际的数据进行对比,从而验证精度,然而该过程计算时间长,且步骤重复。多元非线性回归预测是通过直接数据拟合得到非线性函数,但拟合精度限

定值是固定值,即预测精度是一个静态数值,且与实际负荷数据并没有对比分析,而拟合得到的最终负荷预测函数往往难以修改,计算效率和计算精度较低。为了解决这些问题,本文提出了一种精度反馈预测的算法。

该算法主要由三大基本算法组成:因子分析算法、趋势分析算法和多元非线性回归算法。首先,利用因子分析和趋势分析算法对历史数据进行分析;其次,利用多元非线性回归模型进行负荷预测,将气象指标作为自变量,负荷数据作为因变量进行预测函数拟合;然后,利用因子分析和趋势分析的数据反馈,进行精度验证,对拟合出的历史预测数据与气象指标进行因子分析和趋势分析,并与第一步得到的历史数据进行相对误差计算,并将该误差反馈给多元非线性回归预测模型。若预测相对误差小于规定值,则拟合结束,若该相对误差大于误差规定值,则修正多元非线性回归的系数和次数,使最终得到的负荷预测函数具有极高的准确性。该方法通过数据的自身反馈,解决了传统算法由于重复预测导致的计算负担问题,具有很高的计算效率,同时回归算法的精度是一个随着数据和预测函

收稿日期:2018-08-06;修回日期:2018-08-20

基金项目:浙江省电力有限公司项目“电力现货市场下的公司售电量变化趋势及中长期差价合约管理策略研究”

作者简介:胡怡霜(1995),女,浙江海宁人,博士研究生,研究方向为电力市场;夏翔(1974),男,河南永城人,高级工程师,研究方向为电网信息化研究和管理;丁一(1978),男,北京人,教授,研究方向为电力市场。

数的变化而不断变化的动态数值,能有效提高预测精度。

1 算法分析

1.1 因子分析反馈算法

因子分析是多元统计学的一个分支。因子分析法研究的是以最少的信息丢失为基础,将众多的具有较高相关性的观测变量浓缩为少数几个假想变量^[7]。

进行因子分析的基本步骤为:

(1) 算变量的相关矩阵。因子分析的前提条件是原有的观测变量之间具有较强的相关性,在进行因子分析时,首先要计算观测变量之间的相关系数矩阵。

(2) 定因子变量。在样本数据的基础上,提取出少数几个可以反映大部分原有观测变量信息的因子,达到削减变量个数的目的。

(3) 算因子值。将因子变量表示为原始变量的线性组合,得到因子得分模型。

为了量化多个指标之间的影响关系,选择如下的相关系数计算函数

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中: n 为所需要比较指标的个数; x_i 和 y_i 分别为所需要比较的指标; $\bar{x}$ 、 $\bar{y}$ 分别为所有指标数据的平均值; r 为2个指标之间的相关性。

因子分析数据反馈得到的相关性越小,表明2个指标之间的关系越弱,即改变某一指标的数据,另一指标并不会太大的数据变化;相关性越大,表明2个指标之间的关系越强,这2个指标应作为重点指标加以考虑。同时,各个指标之间的相关性的大小应反映在多元非线性回归预测函数的系数上,也就是说,A指标与B指标具有强相关性,则A指标在B指标的预测函数上应具有较大的系数。反馈误差为

$$e = \frac{\text{各个气象指标相关性的两两比值之和}}{\text{各个气象指标预测函数系数绝对值的两两比值之和}} \quad (2)$$

1.2 趋势分析反馈

利用趋势分析反馈可以统计全年的日最高负荷和日最低负荷的分布情况^[8]。趋势分析反馈是指若历史预测数据和历史实际数据的峰值/谷值所对应的日期范围一致,则认为在该峰值/谷值处两者相对误差为0,该负荷参数函数曲线具有良好的预测精确性;若峰值/谷值所对应的日期范围不一致,则认为在该峰值/谷值处两者具有相对误差,该负荷参

数函数曲线不具有良好的预测精确性,需要修正预测函数。相对误差为

$$e = \frac{\text{两组数据值峰值天数差} + \text{两组数据值谷值天数差}}{\text{两组数据值峰谷值天数总和}} \quad (3)$$

1.3 多元非线性回归分析

多元非线性回归分析是确定2种或2种以上变量间相互依赖的定量关系的一种统计分析方法,常用于预测。如果回归分析中包括2个或2个以上的自变量,且自变量与因变量之间存在非线性关系,则称为多元非线性回归分析^[9]。

在实际问题中,不同变量的测量单位往往是不一样的。为了消除变量的量纲效应,使每个变量都具有同等的表现力,数据分析中常用消量纲的方法^[10]。多元非线性回归分析的模型为

$$y = \beta_0 + \beta_1 T_{\max} + \beta_2 T_{\min} + \beta_3 T_{\text{ave}} + \beta_4 H_r + \beta_5 R + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \quad (4)$$

式中: β_0 、 β_1 、 β_2 、 β_3 、 β_4 、 β_5 、 σ^2 均为与气象指标 $T_{\max}$ 、 $T_{\min}$ 、 T_{ave} 、 H_r 、 R 无关的未知参数,其中 β_0 、 β_1 、 β_2 、 β_3 、 β_4 、 β_5 称为回归系数。

现得到 n 个独立观测数据 $(y_i, T_{i \max}, T_{i \min}, T_{i \text{ave}}, H_{ir}, R_i)$, 由模型得

$$y_i = \beta_0 + \beta_1 T_{i \max} + \beta_2 T_{i \min} + \beta_3 T_{i \text{ave}} + \beta_4 H_{ir} + \beta_5 R_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad i = 1, 2, \dots, n \quad (5)$$

其中

$$\begin{cases} X = \begin{bmatrix} 1 & T_{1 \max} & T_{1 \min} & T_{1 \text{ave}} & H_1 & R_1 \\ 2 & T_{2 \max} & T_{2 \min} & T_{2 \text{ave}} & H_2 & R_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ i & T_{i \max} & T_{i \min} & T_{i \text{ave}} & H_i & R_i \end{bmatrix} \\ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ \varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n]^T \\ \beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5]^T \end{cases} \quad (6)$$

通过相关数据建立回归模型,进行求解。

2 多元回归负荷预测算法

本文提出的精度反馈预测算法通过拟合出的历史预测数据与气象指标进行因子分析和趋势分析,并进行相对误差计算,将该误差动态反馈给多元非线性回归预测模型,通过动态精度来修正多元非线性回归的系数和次数,从而使最终得到的负荷预测函数具有极高的准确性。该方法充分考虑历史数据本身的特性趋势,即利用数据本身的特性来快速验证预测的准确性,大大减少原有算法的工作量,通过数据的自身反馈,解决了传统算法由于重复预测所带来的计算负担的问题。

基于因子和趋势分析反馈的考虑气象因素的多元回归负荷预测算法具体流程如下。

- (1) 设定负荷参数和气象因素;
- (2) 根据历史电力数据,采用趋势分析算法统计历史全年中各个负荷参数各自的分布情况,获得各个负荷参数各自以日为单位的变化曲线,每个负荷参数对应有一条变化曲线,并记录峰谷天数据;
- (3) 先将气象指标标准化,得到统一量纲后的气象数据,故在因子分析和多元非线性回归预测算法的模型中代入的气象数据均为标准化后的气象数据;
- (4) 采用因子分析,根据历史电力数据对负荷指标与气象指标的关系进行相关性分析,获得每个气象因素对每项负荷参数的相关性值,相关性越大,表明该气象因素对该负荷参数的影响越大;
- (5) 根据历史电力数据,采用多元非线性回归分析算法进行预测,获得负荷参数函数及其函数曲线;
- (6) 针对每项负荷参数,将第(5)步通过预测得到的负荷参数函数曲线与第(3)步得到的变化曲线进行趋势分析和相对误差计算。相对误差是指历史预测数据和历史实际数据之间峰值天和谷值天不同的天数和与所有峰值天和谷值天的天数和的比值。若相对误差小于8%,则进入下一步,否则需要改变多元非线性回归分析的预测函数次数,通过相对误差的动态修正,得到符合计算精度的负荷曲线;
- (7) 针对每项负荷参数由第(6)步保留得到的每个负荷参数函数,将负荷参数函数中各个气象因素对应的系数的绝对值分别与第(4)步得到各项气象因素的相关性值进行比较。反馈误差是指各个气象指标的相关性的两两比值之和与各个气象指标在预测函数上的系数的两两比值之和的比值,若该比值在90%~110%之间,则认为预测函数预测精度高,否则需要调整气象指标前的系数,重新预测;
- (8) 利用第(7)步获得的最终负荷参数函数对各自的负荷数据进行预测,获得预测结果。

3 算例分析

已知某城区从2012年1月1日至2015年1月10日的电力负荷数据(每15 min一个采样点)以及气象数据,如表1所示。气象数据统计结果见表2所示。为减少数据量,特选取5个气象指标和2个负荷指标代替负荷的整体变化趋势。气象指标是由气象数据中的5组数据得出的指标,是影响负荷变化的常见指标;负荷指标选择日最大负荷和日最小

负荷,通过峰谷值来确定负荷的变化范围。平台为MATLAB,硬件配置为双核处理器和4GB显存。

表1 日最高、最低负荷数据统计结果(局部) MW

日期	最高负荷	最低负荷
2015-01-01	4 082.142 304	2 543.075 584
2015-01-15	8 141.535 616	4 368.638 080
2015-01-25	7 699.712 512	4 163.356 576
2015-02-01	5 963.816 416	3 338.601 280
2015-02-15	3 216.371 872	1 801.964 320

表2 气象数据统计结果(局部)

日期	最高温度/℃	最低温度/℃	平均温度/℃	相对湿度(平均)	降雨量/mm
2012-01-01	19.5	12.1	15.8	63	0
2012-01-02	20.0	13.0	16.0	59	0
2012-01-03	18.7	14.2	15.8	72	0
2012-01-04	14.9	9.3	10.9	62	0
2012-01-05	9.2	5.1	6.9	78	2.9

2015年日最高负荷分布如图1所示。2015年日最低负荷分布如图2所示。2015年96点日负荷数据分布如图3所示。

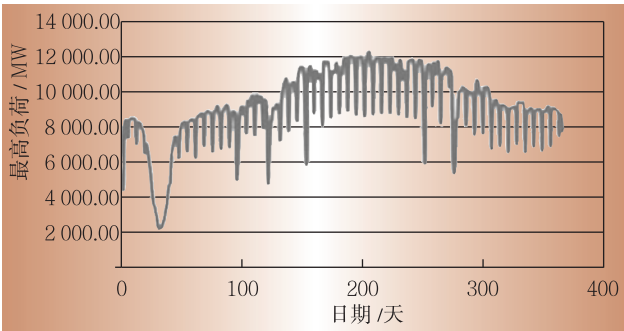


图1 2015年日最高负荷分布

图1和图2的横轴表示天数,即365天。图3的2个横轴,一个横轴表示天数,即365天,另一个横轴表示一天内的96个负荷采集点,每15 min采集一次。由图1、图2可知,最大和最小负荷都在7月和8月到了最高点,在3月和4月到了最低点,并且在一天的变化上,日负荷在中午达到峰值,在凌晨达到谷值,且每一天的负荷都呈现相同的趋势。提取2个负荷指标的在一年内的峰值天和谷值天,作为趋势算法精度反馈依据。

将各个气象指标标准化,并采用因子分析算法

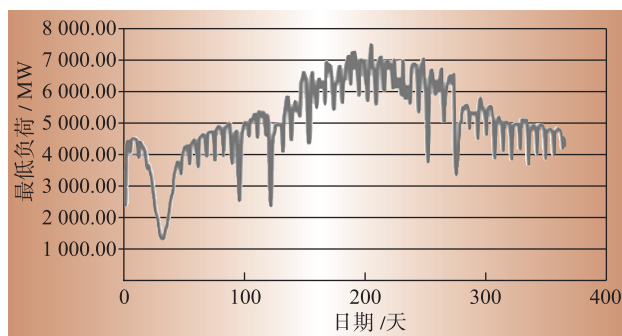


图2 2015年日最低负荷分布

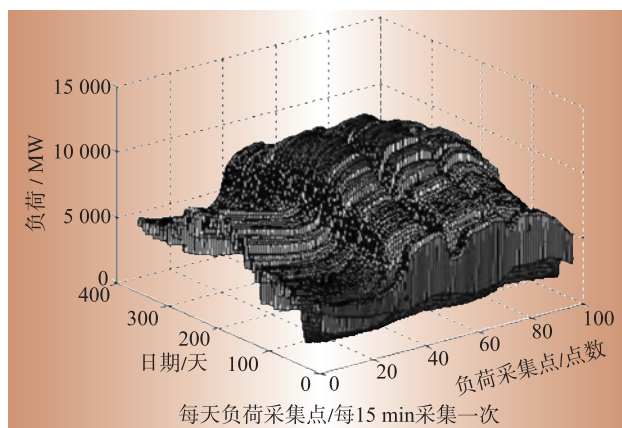


图3 2015年96点日负荷数据分布

中的相关性计算公式,可以得到各个气象指标和负荷指标之间的相关性值。式中: $L_{\max}$ 和 $L_{\min}$ 分别为日最高负荷和日最低负荷指标; $T_{\max}$ 、 $T_{\min}$ 、 T_{ave} 、 H_r 和 R 分别为日最高温度、日最低温度、日平均温度、日相对湿度和日降雨量指标。为有效进行后期的动态误差反馈,选取5个气象指标和2个负荷指标进行相关性计算。气象指标和负荷指标的相关性分布如表3所示。

表3 气象指标和负荷指标的相关性分布

指标	$T_{\max}$	$T_{\min}$	T_{ave}	H_r	R
$L_{\max}$	0.597	0.306	0.258	0.139	0.119
$L_{\min}$	0.592	0.387	0.166	0.13	0.153

从表3可以得到如下结论: $T_{\max}$ 和每一个负荷指标之间都具有最大的相关性, H_r 和每一个指标之间都具有最小的相关性,也就是说,如果 $T_{\max}$ 发生剧烈变换,那么负荷也会发生剧烈变化。通过表3可以得出一些指标规划上的结论,即重点关注 $T_{\max}$ 的变化趋势,从而将负荷大小维持在可接受的范围内。与此同时,各个指标之间的相关性的比值反映了负荷预测函数的指标系数的比值趋势,如果两者趋势满足精度要求,那么负荷预测具有较高的精确性,可用于精度验证。

通过标准化后的各个气象指标 $T_{\max}$ 、 $T_{\min}$ 、

T_{ave} 、 H_r 、 R 和负荷指标 $L_{\max}$ 、 $L_{\min}$ 间的相关性反馈以及负荷指标的趋势反馈,不断修正多元非线性回归模型负荷预测函数的自变量次数和系数,从而可以得到满足相对误差(历史预测数据和历史实际数据之间峰值天和谷值天不同的天数与所有峰值天和谷值天的天数和的比值)小于8%,以及反馈误差(各个气象指标的相关性的两两比值之和与各个气象指标在预测函数上的系数的两两比值和的比值)在90%~110%之间的负荷回归预测函数

$$L_{\max} = -0.1T_{\max}^4 + 5.8T_{\max}^3 + 147.5T_{\max}^2 + 2054.1T_{\max} - 5T_{\min}^3 + 9.8T_{\min}^2 + 1082.4T_{\min} - 2T_{\text{ave}}^3 + 114.1T_{\text{ave}}^2 - 969.9T_{\text{ave}} - 0.2H_r^2 + 418.1H_r - 402.7R \quad (7)$$

$$L_{\min} = 0 \times T_{\max}^5 - 0.1T_{\max}^4 + 2.6T_{\max}^3 + 56.8T_{\max}^2 + 571.6T_{\max} - 0 \times T_{\min}^4 - 1.2T_{\min}^3 + 39.3T_{\min}^2 - 438.6T_{\min} - 0.2T_{\text{ave}}^3 + 5.9T_{\text{ave}}^2 + 144T_{\text{ave}} - 0.2H_r^2 - 151H_r - 102.5R \quad (8)$$

由上述最高负荷和最低负荷预测函数可知,各个气象指标前的系数的绝对值比值与指标相关性的比值基本保持一致,并且预测函数经过趋势分析后得到的历史预测负荷曲线在峰谷值的时间与历史实际数据通过趋势分析得到的数据保持一致。

下面将从计算时间和精度2个方面,对基于因子分析和趋势分析反馈的多元非线性回归预测和传统的多元非线性回归分析算法^[12]进行比较,验证本文提出算法的预测性能。其中计算精度和计算时间为对比预测3次后的平均值,数据选取2014年12月1号至2014年12月30号的数据(负荷数据仅考虑日最高和日最低负荷)。计算效率和计算精度对比如表4所示。

表4 计算效率和计算精度对比

算法	负荷指标	计算精度/%	计算时间/s
反馈预测算法	$L_{\max}$	3	52
	$L_{\min}$	4	55
传统预测算法	$L_{\max}$	11	80
	$L_{\min}$	12	85

由表4可知,反馈预测算法的精度远远高于传统预测算法,并且计算时间大大缩短。

4 结论

本文提出的基于因子分析和趋势分析反馈的多元非线性回归预测算法,通过因子分析对气象和负荷指数进行相关性计算,再通过趋势分析量化历史负荷数据在单位时间尺度上的变化。之后通过多元非线性回归模型,得到预测函数。(下转第35页)