

基于 SPSS 的短期负荷特性分析及其预测研究

徐东升¹, 杨巍², 魏哲³, 刘国兴²

(1. 开滦热电公司, 河北 唐山 063103; 2. 东北电力大学电气工程学院, 吉林 吉林 132012;
3. 华北电力大学电气工程学院, 北京 102206)

摘要: 为了提高一个地区短期负荷预测的精度, 对该地区的负荷进行特性分析是十分必要的。而在分析的过程中则需要处理和分析大量的数据, SPSS 正是一款在数据处理和分析方面有着强大功能的易用软件。将 SPSS 软件的应用引入到电力系统中, 主要从主成分分析、聚类分析和多元回归分析 3 个方面, 阐述 SPSS 在电力系统负荷特性分析中的应用; 从时间 ARIMA 模型分析中说明 SPSS 在预测方面的应用。通过一些示例说明 SPSS 的具体应用, 从而显示它的重要功能: 可以更清楚地了解负荷数据间的内在联系、变化规律和影响短期负荷特性, 对提高负荷预测的水平具有重要的意义。

关键词: SPSS; 短期负荷特性; 负荷预测; 主成分分析; 聚类分析; 多元回归分析; ARIMA

Application of SPSS in characteristic of short load and its forecasting

XU Dong-sheng¹, YANG Wei², WEI Zhe³, LIU Guo-xing²

(1. Kailuan Thermal Power Co. Ltd, Tangshan 063103, China; 2. School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China; 3. School of Electrical Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: In order to improve the short load forecasting accuracy in the region, it is very necessary to analyze the characteristic of the load in the same area. A large amount of data is needed to process and analyze in the analysis of the characteristic of the load. SPSS is such a easy-used software which is very powerful in data processing and analysis. In this paper, SPSS software is applied in power system, and the application of SPSS software in power system load characteristic analysis is introduced from the principal component analysis, cluster analysis and multiple regression analysis. And the application of SPSS in load forecasting is introduced from the analysis of ARIMA time model. A number of examples are used to illustrate the specific application of SPSS, from which its important function is shown: it can more clearly explain the internal relations and the law of change of the load data, affect the short load characteristic and play a very important part in promoting the level of the load forecasting.

Key words: SPSS; short load characteristics; load forecasting; principal component analysis; cluster analysis; multiple regression analysis; ARIMA

中图分类号: TM714 文献标识码: A 文章编号: 1674-3415(2009)21-0147-05

0 引言

由于各地区的负荷特性的不同, 目前还没有一种预测方法适用于所有地区。因此, 对负荷进行特性分析, 找出数据间的内在联系、变化规律和影响负荷预测的特性, 就显得十分重要了。SPSS(Statistical Product and Service Solutions)统计分析软件以其清晰、直观、易学易用、涵盖面广等特点和具有强大的图形功能特色, 通过统计图可以直观、清晰地描述用于负荷预测的数据间的内在联系和变化规律^[1]。因此, 应用 SPSS 分析软件, 从负荷特性分析和预测两方面进行研究。

1 基于 SPSS 的负荷特性分析

1.1 主成分分析

主成分分析是在基本保持原变量信息不变的前提下, 通过原变量的少数几个线性组合来代替原变量并揭示原变量之间关系的一种分析方法^[2]。为了能用少量的气象因子来表征一天的天气状况, 这里将气象因子作为一个分析集合, 进行气象因子的主成分分析。以哈尔滨地区 1999 年春季的 5 个相关气象因素数据作为训练数据, 利用 SPSS 的因子分析对相关因素作主成分分析, 分析结果如表 1 和表 2 所示。

由表 1 分析, 变量间存在较强的相关关系, 即

日最高温度与日最低温度相关系数达到了 0.945, 这说明主成分分析是必要的。

表1 相关矩阵

Tab.1 Correlation matrix

	日最高温度	日最低温度	日平均温度	降水量	相对湿度
日最高温度	1.000	0.945	0.124	0.333	-0.267
日最低温度	0.945	1.000	-0.015	0.330	-0.114
日平均温度	0.124	-0.015	1.000	-0.253	-0.622
降水量	0.333	0.330	-0.253	1.000	0.098
相对湿度	-0.267	-0.114	-0.622	0.098	1.000

表2 主成分统计信息

Tab.2 Tootal variance explained

主成分	初始特征值			提取平方载荷总和		
	特征根	方差贡献率%	累计贡献率%	特征根	方差贡献率%	累计贡献率%
1	2.189	43.788	43.788	2.189	43.788	43.788
2	1.713	34.250	78.038	1.713	34.250	78.038
3	1.102	14.031	92.069	1.102	14.031	92.069
4	0.357	7.131	99.201			
5	0.040	0.799	100.000			

表 2 是主成分的统计信息, 包括特征根由大到小的次序排列, 各主成分的贡献率累积贡献率: 第一主成分的特征根为 2.189, 它解释了负荷变化的 43.788%; 第二主成分的特征根为 1.713, 它解释了负荷变化的 34.250%; 第三主成分的特征根为 1.102, 它解释了负荷变化的 14.031%。前 3 个主成分累计贡献率达到了 92.069%, 这说明前 3 个主成分提供了原始数据的足够信息, 因此提取 3 个主成分。

由第 1 主成分, 第 2 主成分和第 3 主成分与原始变量关系, 可以写出主成分与原变量的线性组合, 根据表 3 所示的主成分矩阵, 主成分可用下列线性组合表示:

$$Z_1 = 0.971T_{\max} + 0.932T_{\min} + 0.182F_1 + 0.462F_2 - 0.364F_3 \quad (1)$$

$$Z_2 = 0.037T_{\max} + 0.184T_{\min} - 0.878F_1 + 0.524F_2 + 0.795F_3 \quad (2)$$

$$Z_3 = -0.189T_{\max} - 0.281T_{\min} + 0.119F_1 + 0.710F_2 - 0.263F_3 \quad (3)$$

表 3 主成分矩阵

Tab.3 Principal component matrix

	主成分		
	1	2	3
日最高温度	0.971	0.037	-0.189
日最低温度	0.932	0.184	-0.281
日平均温度	0.182	-0.878	0.119
降水量	0.462	0.524	0.710
相对湿度	-0.364	0.795	-0.263

其中: Z_i 为第 i 个主成分, T_{\max} 代表日最高温度, T_{\min} 代表日最低温度, T_{av} 代表日平均温度, P 表示降水量, H 表示相对湿度。

通过主成分分析降低了数据维数, 去除了数据间的相关性, 减少了噪音数据的影响。另外 SPSS 还会将计算出的每条记录的三个主成分作为新变量自动存储到原始数据文件中, 有利于下一步的负荷预测。

1.2 多元线性回归分析

SPSS 中的多元线性回归分析, 通过统计方法分析多个变量之间依存变化关系及线性相关特性。这里利用 SPSS 的多元线性回归中的逐步法, 分析负荷与气象因子的依存关系^[3,4]。使用逐步法时要事先确定选入自变量的标准, 开始时, 方程中只含常数项, 按自变量对因变量的贡献率大小由大到小依次选入方程。每引入一个自变量, 都会对已在方程中的变量进行 F 检验, 对符合剔除标准的变量逐一剔除。F 检验是似然比检验, 反映整体对因变量的影响。

为了反映气象因子对一天总负荷的影响作用, 采用某地区 1999 年春季的电力负荷进行日负荷 L 与气象因子(日最高温度 T_{\max} , 日最低温度 T_{\min} , 日平均温度 T_{av} , 降水量 P , 湿度 H)的逐步线性回归, 结果见表 4 和表 5。

表 4 给出了拟合未标准化和标准化之后的回归系数值; 表 5 反映的是多元线性回归拟合模型中没有进入模型变量的检验情况。通过表 6 最终得到方程:

$$L = 1066.324 - 6.838T_{\max} + 1.469P \quad (4)$$

表4 回归系数值
Tab.4 Coefficients for regression

模型		非标准化系数		标准化系数	T检验	显著性
		双向方差	标准差	β 值	双向方差	标准差
1	(常数项)	1 066.324	6.764		157.655	0.000
	日最高温度	-6.838	0.469	-0.838	-14.590	0.000
	降水量	1.469	0.217	0.763	9.330	0.000

表5 排除模型变量的检验情况
Tab.5 Excluded variables

模型		β 值	t检验	显著性	偏相关系数	共线性统计
1	日最低温度	0.043	0.245	0.807	0.026	0.907
	日平均温度	0.020	0.348	0.729	0.037	0.985
	相对湿度	-0.072	-1.177	0.242	-0.124	0.889

通过回归方程可以发现：(1)日负荷除了和最高温度有着强相关外，其次和降水量的线性相关性也很明显；也就是说对日负荷影响最大的变量是最高温度，其次是降水量、最低温度和相对湿度对日负荷的影响最小。(2)最高温度与日负荷呈负相关，降水量与日负荷呈正相关。通过以上分析所得到的一些十分有益的结论，可为后面的负荷预测工作提供信息和数据支持，从而提高负荷预测的精度。

以下对线性回归方程拟合优度进行检定。通过观察图 2 所示的标准化残差直方图和图 3 所示的标准残差正态 P-P 图可以发现：由于残差具有正态分布的趋势^[5]，因此可以认为这里的回归模型是恰当的。

多元回归分析可以解释负荷与气象因子的依存关系，通过分析能够找出负荷与何种气象因素整体相关程度。对于相关程度小的因素可以排除不做分析，从而减小样本数据量冗余程度。

1.3 聚类分析

SPSS 的聚类分析，又称集群分析，是通过数值特征，来观察变量之间的亲疏关系，这种亲疏关系是用变量之间的距离来衡量的，距离定义之后，则把距离近的变量归为一类。

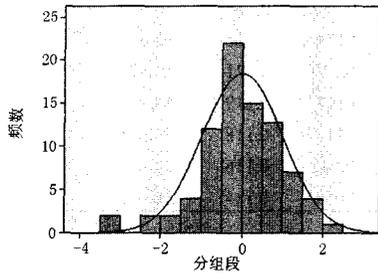


图1 标准化残差直方图

Fig.1 Standardized residuals

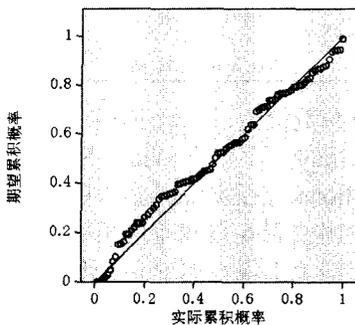


图2 标准残差正态P-P图

Fig.2 Standardized residuals normal P-P plot

*****HIERARCHICAL CLUSTER ANALYSIS*****

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine

CASE 0 5 10 15 20 25
Label Num

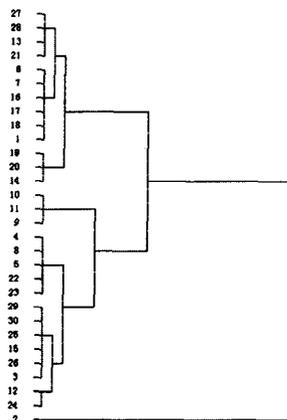


图3 聚类分析的树形图

Fig.3 Cluster analysis dendrogram

这里采用层次聚类法,并取哈尔滨地区的1999年4月所有工作日的有关数据为样本,把一天的日最高温度 T_{max} , 日最低温度 T_{min} , 降水量 P 和日负荷 L 看作一个 m 维向量 Y_i , 其中 $m=4, i=1,2,\dots,30$ (天数), 可得到聚类分析树形图, 也称龙骨图, 见图3。

对于最先归类的日期, 表明它们之间的亲密程度最大。由图可见9日, 10日与11日, 12日与24日等最为相似; 2日与其它各日相似性最差。通过以上的聚类分析, 可以找出历史相似日, 这对于电力系统的短期负荷预测十分有利。

2 基于 SPSS 的负荷预测

SPSS 提供了强大的时间序列分析的功能。由于电力系统负荷数据是以时间序列形式存储的, 所以很方便采用时间序列分析法进行负荷预测。时间序列分析法有趋势外推法、指数平滑法和 ARIMA (求和自回归滑动平均模型) 等^[6], 其中 ARIMA 被大量用于时间序列分析。

由于负荷序列存在季节性周期波动, 本文采用带有季节因子的 ARIMA 模型进行负荷预测。季节性参数的阶数是通过尝试和比较法确定, 季节性 ARIMA 模型的表述为 $(p,d,q) \times (P,D,Q)_s$ 。其中 p, d, q 分别为自回归、差分和滑动平均阶数, s 为季节参数, P, D, Q 分别表示以 s 为间距的自回归、差分和移动平均算子的阶数。

以哈尔滨地区99年春季50天的日负荷数据作为样本数据, 后面4天的日负荷数据作为预测对比数据。首先是模型参数的识别, 负荷序列输出的 ACF (自相关) 和 PACF (偏自相关) 如图4和图5, 由图可见 ACF 呈拖尾衰减, PACF 一步截尾, 识别为 AR1 模型, 即 $p=1, d=0, q=0$ 的 ARIMA (1,0,0)。由于日负荷数据具有周周期性, 则 $S=7$, 通过 PACF 图也可以发现, PACF 并不呈通常的截尾或逐渐衰减, 而是在时滞7天处有突然的上升, 这种特征说明了日负荷序列具有周期性。本着由简单到复杂的尝试建模原则, 找到最佳模型结构为 $(1,0,0) \times (1,0,0)_7$ 。

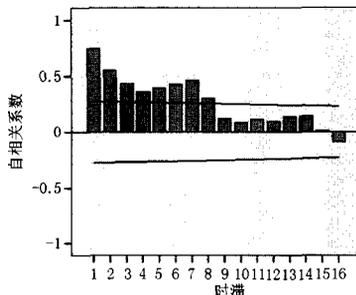


图4 自相关图
Fig.4 ACF

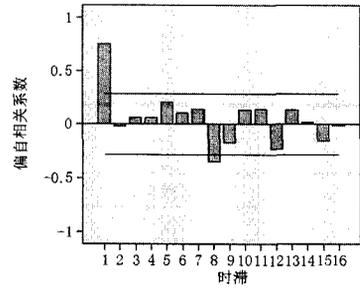


图5 偏自相关图

Fig.5 PACF

表6 残差自相关检验表

Tab.6 Residuals for autocorrelation test

时滞	自相关系数	标准差	Box-Ljung 统计		
			统计量	自由度	P值
1	-0.059	0.137	0.186	1	0.666
2	0.077	0.136	0.506	2	0.776
3	0.134	0.134	1.503	3	0.681
4	-0.184	0.133	3.418	4	0.490
5	0.131	0.132	4.415	5	0.491
6	0.068	0.130	4.686	6	0.585
7	-0.070	0.129	4.979	7	0.663
8	0.116	0.127	5.818	8	0.668
9	-0.268	0.126	10.377	9	0.321
10	-0.052	0.124	10.550	10	0.394
11	0.178	0.122	12.660	11	0.316
12	-0.130	0.121	13.822	12	0.312
13	0.117	0.119	14.781	13	0.321
14	0.115	0.118	15.731	14	0.330
15	-0.043	0.116	15.868	15	0.391
16	0.054	0.114	16.095	16	0.446

表7 参数估计表

Tab.7 Parameter estimates

		估计值	标准差	t检验	显著性
非季节性时滞	AR1模型	0.742	0.096	7.726	0.000
季节性时滞	季节性AR1	0.471	0.134	3.524	0.000
常数项		746.456	19.457	38.364	0.000

其次是参数检验, 以验证季节性模型的正确性。检验通过 Graphs 菜单下 Time Series 子菜单里的 Autocorrelations 过程实现, 选项中选择残差 (ERR) 为变量。输出的表6是残差自相关检验的结果, 如果在任何时滞上出现小的 P 值 (比如说小于 0.05), 则认为残差存在自相关性, 模型拟合不足, 需要改进。如本例, P 值都远大于 0.05, 可以认为残差为白噪声序列, 模型诊断得以通过。表7为参数的估计。表8为预测输出表。

表8 预测输出表
Tab.8 Outputs of forecasting

时间	实际值	点预测值	误差百分比 (PE)	95%置信区间 上限	95%置信区间 下限
4月13日	737.96	715.919 9	2.986 6	671.420 5	760.419 3
4月14日	740.72	759.047 6	-2.474 3	714.548 2	803.547 0
4月15日	755.74	724.946 6	4.074 6	680.447 2	769.446 0

从表 8 看出, 预测的效果总体上是令人满意的, 3 日内预测误差都小于 5%。

3 结论

论文应用 SPSS 研究了一些数理统计分析方法, 可以得到如下结论:

- 1) 通过主成分分析降低了数据维数, 去除了数据间的相关性, 减少了噪音数据的影响;
- 2) 通过多元线性回归分析, 能够找出负荷与何种气象因素整体相关程度;
- 3) 通过以聚类分析, 可以找出历史相似日;
- 4) 季节性 ARIMA 方法可以对天气因素不敏感的负荷序列进行很好的预测。

参考文献

- [1] 余建英, 何旭宏. 数据统计分析与 SPSS 应用 (第一版) [M]. 北京: 人民邮电出版社, 2003.
YU Jian-ying, HE Xu-hong. Analysis of Data Statistics and Application of SPSS (First Edition) [M]. Beijing: Posts & Telecom Press, 2003.
- [2] 王志征, 余岳峰, 姚国平. 主成份分析法在电力负荷预测中的应用 [J]. 电力需求侧管理, 2003, 24 (3): 21-24.

WANG Zhi-zheng, YU Yue-feng, YAO Guo-ping. The Application of Principal Component Analysis in Load Forecasting [J]. Power Demand Side Management, 2003, 24 (3): 21-24.

- [3] 王治华, 杨晓梅, 李扬. 气温与典型季节的电力负荷关系的研究 [J]. 电力自动化设备, 2002, 22(3): 16-18.
WANG Zhi-hua, YANG Xiao-mei, LI Yang. Study on Relationship between Temperature and Electrical Loads in Typical Seasons [J]. Electric Power Automation Equipment, 2002, 22 (3): 16-18.
- [4] Heinemann G T, Nordman D A, Plant E C. The Relationship Between Summer Weather and Summer Loads-a Regression Analysis [J]. IEEE Trans on PAS, 1996, 85(11):1144-1154.
- [5] 宇传华. SPSS 与统计分析 (第一版) [M]. 北京: 电子工业出版社, 2007.
YU Chuan-hua. SPSS and Statistical Analysis (First Edition) [M]. Beijing: Publishing House of Electronics Industry, 2007.
- [6] 于尔铿, 刘广一, 周京阳. 能量管理系统 (第二版) [M]. 北京: 科学出版社, 1998.
YU Er-keng, LIU Guang-yi, ZHOU Jing-yang. Energy Management System (Second Edition) [M]. Beijing: Science Press, 1998.

收稿日期: 2008-11-01; 修回日期: 2008-12-06

作者简介:

徐东升 (1981-), 男, 工程师, 从事电力系统运行和电力市场方面研究; E-mail: weiyang1984@126.com

杨巍 (1984-), 男, 硕士研究生, 从事电力系统规划、可靠性和电力市场方面研究;

魏哲 (1983-), 男, 硕士研究生, 从事电力系统运行、电力系统规划和电力市场方面研究。

(上接第 98 页 continued from page 98)

- [9] 段振国, 高曙, 杨以涵. 电网结构知识表示方法研究与管理系统的开发 [J]. 中国电力, 1999, 32(2): 34-37.
DUANG Zhen-guo, GAO Shu, YANG Yi-han. Study of Power Network Structure Knowledge Representation and Development of Knowledge Management System [J]. Electric Power, 1999, 32(2): 34-37.
- [10] 杨以涵, 唐国庆, 高曙. 专家系统及其在电力系统中的应用 [M]. 北京: 水利电力出版社, 1995.
YANG Yi-han, TANG Guo-qing, GAO Shu. Expert System and Its Application in Power System [M]. Beijing: Hydraulic and Electric Power Press, 1995.
- [11] 周明, 林静怀, 杨桂钟, 等. 新型智能电网调度操作票自动生成与管理系统的研究 [J]. 电力系统及其自动化, 2004, 6(11): 71-74.
ZHOU Ming, LIN Jing-huai, YANG Gui-zhong, et al. A New-type Intelligent Dispatching Operation Order

System [J]. Automation of Electric Power Systems, 2004, 28(11): 71-74.

- [12] 杨继涛, 胡明, 吴琼, 等. 电网调度操作票专家系统的设计与开发 [J]. 继电器, 2004, 8(15): 45-48.
YANG Ji-tao, HU Ming, WU Qiong, et al. Design and Development of Dispatching Sheet Expert System for Power Network [J]. Relay, 2004, 32(15): 45-47.

收稿日期: 2008-11-12; 修回日期: 2009-03-08

作者简介:

王艳 (1982-), 女, 硕士, 主要研究方向为电力系统调度自动化、电力系统分析与控制等; E-mail: ahwy8212@163.com

杨以涵 (1927-), 男, 教授, 博士生导师, 主要研究方向为电力系统运行、分析与控制, 电力系统调度自动化, 光学互感器等。