

基于主成分分析的短期负荷预测模型*

吕佳

(重庆师范大学 数学与计算机科学学院, 重庆 400047)

摘要 :电力负荷预测的准确性直接影响到电力系统的安全性和经济性,但在应用神经网络进行短期负荷预测精度造成了显著的负面影响。针对这一问题,本文采用多元统计分析中的主成分分析,根据各主成分贡献率对输入空间进行约简,提取线性无关的输入变量,以此达到压缩变量维数的目的,然后利用考虑模型输入变量相互关系的递推合成BP网络进行预测,使之更符合电力短期负荷预测的特点,提高模型的预测精度。仿真实验的结果表明,该简化模型用于短期负荷预测建模速度快、预测精度高,是一种行之有效的办法。

关键词 :短期负荷预测;主成分分析;递推合成BP网络

中图分类号:TP18;TM715

文献标识码:A

文章编号:1672-6693(2007)03-0033-04

Short-term Load Forecasting Model Based on Principal Component Analysis

LÜ Jia

(College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047, China)

Abstract :The accuracy of short-term load forecasting in power system has directly influence on its safety and economy. When neural network forecasting model is used to perform short-term load forecasting, its input dimension number is commonly too great as well as its input variances have heavy self-correlation to have a negative effect on network training efficiency and forecasting accuracy of neural network. Focusing on solving this problem, a principal component analysis in multiple method of data processing is developed. The original input space is reduced by the contribution rate of each principal component to eliminate the multicollinearity of input variables, thus reducing the variable dimension. Then next the recurrent composite BP neural network in use of the correlations between input variances is used to forecast. All these improvement can make forecasting model in accordance with real condition, and improve the accuracy of load forecasting. The results of simulation experiments form a real power system to indicate that the simplified model is effectual. With this method the high modelling speed and high forecasting accuracy can be obtained.

Key words :short-term load forecasting; principal component analysis; recurrent composite BP neural network

电力系统短期负荷预测是电力部门重要的工作之一,是电力系统优化调度(如机组最优组合、经济调度、最优潮流计算等)的基础工具,又是电力企业日常经营管理工作的组成部分。准确的短期负荷预测是电网调度机构制定发供电计划和做好电网供需平衡的关键。神经网络由于具有高度的非线性运算能力和很强的自学习、自组织能力,常常被用作短期负荷预测建模的重要方法。其中BP网络是神经网络中较为常用的一种,随着应用的广泛,BP网络存在的问题也日益显现出来。如其存在的学习收

敛速度慢,不可避免地存在局部极小等问题影响了BP网络的进一步应用。故本文采用递推合成BP网络作为短期负荷预测的建模方法,该网络是在常规BP网络的基础上通过对网络结构、学习算法等加以改进而形成的。由于实际的电力负荷要受到经济、政策、气象及社会等因素的影响,而这些因素之间往往存在着较高的相关性。这样,当应用神经网络建模时,在由各个负荷分量组成的时间序列中各个数据点之间就相应地具有一定的相关性和差异性。在进行短期负荷预测时,若对所有分量采用同一个模

* 收稿日期 2006-02-16 修回日期 2006-07-04

作者简介:吕佳(1978-),女,四川达州人,讲师,硕士,研究方向为人工智能、数据挖掘。

型进行预测,则只顾及到其共性而忽略了其间的差异,若对所有分量单独建立预测模型,则所有分量处于同等重要的地位,当要预测的点数较多时,建模的工作量很大,无法采用一些复杂模型^[1-3]。目前针对这一问题已有很多研究,其基本思想大多是根据一定的方法找出相关性较大的一些相关量来代替所有相关量的空间,舍弃相关性较小的相关量。这样就在损失一些相关量信息的基础上,一定程度上克服了输入空间的自相关性,并降低了网络的维数^[4,5]。因此,本文采用多元统计学中的主成分分析法(Principal Component Analysis, PCA, 也称为主元分析法)先对负荷数据进行适当的处理以将其共同的特征和差异分离开,然后再对这些分离得到的分量应用递推合成 BP 网络建模进行预测。

1 主成分分析的原理

主成分分析是研究多个数值变量间相关性的一种多元统计方法^[6-7]。在基本保持原变量信息不变的前提下,能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。由于各主成分之间是相互独立的,所以由各个主成分组成的输入空间不存在自相关性。主成分分析法求解步骤如下。

1) 对原始数据 $X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$ 进行标

准化处理,共有 n 个样本,每个样本为 p 维,其中 x_{ij} 为第 i 个样本的第 j 维取值。用 x_1, x_2, \dots, x_p 表示 X 的各列向量,则有 $x_j^* = \frac{x_j - E(x_j)}{\sqrt{Var(x_j)}}$ $j = 1, 2, \dots, p$, 式中的 $E(x_j)$ 和 $Var(x_j)$ 表示 x_j 的均值和方差,计算公式为

$$E(x_j) = \frac{1}{n} \sum_{i=1}^n x_{ij}, Var(x_j) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - E(x_j))^2$$

2) 计算样本各维间相关系数矩阵 $R = (r_{ij})_{p \times p}$, 其中

$$r_{ij} = \frac{Cov(x_i, x_j)}{\sqrt{Var(x_i)Var(x_j)}} = \frac{1}{n} \sum_{k=1}^n x_{ki}^* x_{kj}^* \quad (i, j = 1, 2, \dots, p),$$

式中的 $Cov(x_i, x_j)$ 表示数据矩阵中第 i 列与第 j 列之间的协方差。

3) 求正交矩阵 P , 使得 $P^T R P = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ 式中 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 是 R 的 p 个特征值, diag 表示对角阵, $Y = (X^*) P$ 。令 $\beta_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$, 有

$\beta_1 + \dots + \beta_m = \alpha_m$, 设定 $0 \leq \alpha_0 \leq 1$, 选取前 m 个 β 使得 $\alpha_m > \alpha_0$, 则可称 y_1, y_2, \dots, y_m 为显著性水平为 α_m 的 m 个主成分, 每个主成分集中了随机变量 X 的各个分量不同的共同特征, 这样维数将由 p 个减少为 m 个, 而将其余 $y_{m+1}, y_{m+2}, \dots, y_p$ 看成是随机扰动。



图1 主成分分析示意图

在二维情形下(图1), 二维随机变量 (x_1, x_2) 的一些样本点散落在平面 $x_1 o x_2$ 上, 通过主成分法变换后, 得到一新的坐标系 $y_1 o y_2$, 在新的坐标系下, 这些样本点 y_2 的坐标几乎都是零, 而 y_1 坐标的变化(方差)很大, 因此, 可以把 y_2 视为随机扰动, 而 y_1 是主成分。

2 递推合成 BP 网络学习算法

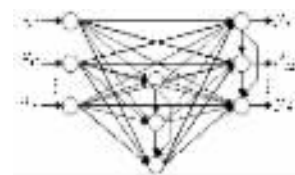


图2 递推合成 BP 网络结构图

递推合成 BP 网络的输入层至输出层之间具有直接连接权, 且在隐层内部及输出层内部各节点中引入了横向递推连接权^[8], 网络结构如图 2 所示。直接连接权和横向递推连接权的引入使网络各不同层节点及同一层节点之间相互关联, 这就充分考虑了时间序列值及其各影响因素之间的相互关系; 同时直接连接权还降低了隐层节点输出在网络输出层中输入的比重, 使因隐层节点的饱和而产生的误差得以减小。递推合成 BP 网络学习算法包括前向计算和误差的反向传播过程。在网络的前向计算中, 输出层及隐层各节点的输出值是由该节点前面各节点(包括同一层和不同层)的输出值推算得到; 而在误差的反向传播中, 采用有序偏导数公式对其误差进行修正。其具体过程如下。

2.1 前向计算

先将网络中输入层、隐层和输出层各层所有节点以自然数顺序依次标号, 各层节点数分别为 N, M, L 。对于一定的输入样本 $(x_{q1}, x_{q2}, \dots, x_{qN})$, 其前向计算过程如下。

1) 当 i 为隐层节点时, 其输出 $o_{qi} = 1 / (1 + \exp(-(\sum_{j=1}^{i-1} w_{ij} o_{qj} + b_i)))$;

2) 当 i 为输出层节点时, 其输出 $o_{qi} = \sum_{j=1}^{i-1} w_{ij} o_{qj}$

+ b_i , 其中 w_{ij} 、 b_i 分别为相应节点 i 和 j 间的连接权及相应节点 i 的阈值, 下标“ q ”表示对应于第 q 个样本的相应值。

2.2 误差的反向传播计算

误差函数为 $E = \sum_{q=1}^Q \frac{1}{2} \sum_{i=N+M+1}^{N+M+L} (t_{qi} - o_{qi})^2$, 有序

偏导数的引入能反映变量间相关性, 考虑了过去时间序列值及其影响因素对误差的影响。其公式为

$\frac{\partial^+ E}{\partial y_i} = \frac{\partial E}{\partial y_i} + \sum_{j>i} (\frac{\partial^+ E}{\partial y_j} \times \frac{\partial y_j}{\partial y_i})$ (上标“+”表示有序偏导数)。网络的连接权值和阈值的有序偏导数修正公式如下。

1) 当 i 为输出层节点时, 当 $i = N + M + L$,

$$w_{ij}(m+1) = w_{ij}(m) + \Delta w_{ij},$$

$$\Delta w_{ij} = \eta \frac{\partial^+ E}{\partial w_{ij}} = -\eta \sum_{q=1}^Q \frac{\partial^+ E_q}{\partial net_{qi}} \times \frac{\partial net_{qi}}{\partial w_{ij}} =$$

$$-\eta (\frac{\partial E_q}{\partial net_{qi}} + \sum_{k=N+M+1}^{N+M+L} \frac{\partial^+ E_q}{\partial net_{qk}} \times \frac{\partial net_{qk}}{\partial net_{qi}}) \times \frac{\partial net_{qi}}{\partial w_{ij}} =$$

$$\eta \sum_{q=1}^Q \delta_{qi} o_{qj} = \eta \sum_{q=1}^Q (t_{qi} - o_{qi}) o_{qj},$$

$$b_i(m+1) = b_i(m) + \Delta b_i,$$

$$\Delta b_i = -\eta \frac{\partial^+ E}{\partial b_i} = -\eta \sum_{q=1}^Q \frac{\partial^+ E_q}{\partial net_{qi}} \times \frac{\partial net_{qi}}{\partial b_i} =$$

$$\eta \sum_{q=1}^Q \delta_{qi} = \eta \sum_{q=1}^Q (t_{qi} - o_{qi}),$$

当 $N + M + L - 1 \geq i \geq N + M + 1$ 时, 同理可计算

$w_{ij}(m+1)$ 与 $b_i(m+1)$ 的值, 其中

$$\delta_{qi} = (t_{qi} - o_{qi}) + \sum_{k=i+1}^{N+M+L} \delta_{qk} w_{ki}$$

2) 当 i 为隐层节点时, 过程同上, 其中

$$\delta_{qi} = o_{qi}(1 - o_{qi}) \sum_{k=N+1}^{N+M+L} \delta_{qk} w_{ki}$$

3 仿真试验及结果分析

以重庆市某供电局 2004 年 1 月至 2004 年 7 月的数据, 用主成分分析法对该样本数据进行计算, 即 $p = 96$ 、 $n = 210$, 设显著性水平为 α 的主成分个数为 m , 令 $Y = (X^*)P$, 则有 $X^* = YP^T$, Y 的前 m 列数据为主成分数据分量, 可对其进行重点建模预测, 而对后 $p - m$ 列数据, 其数值相对来说一般都很小, 其对预测精度的影响较小, 可进行简单建模预测, 也可看作随机扰动, 不必建模, 只须取近几日的数据作平均值即可。表 1 列出了 96 个主成分所对应的特征值和对应的显著性水平值。从表中可以看出, 前 9 个主成分的显著性水平就达到 99.04%, 如果取 α 等于 99%, 则原有的 96 维数据降维为 9 维。以第 60 天的日负荷数据为例, 其各分量曲线及对应的主成分曲线见图 3, 由图中可清楚地看出, 能量主要分布在前十几个主成分分量上。

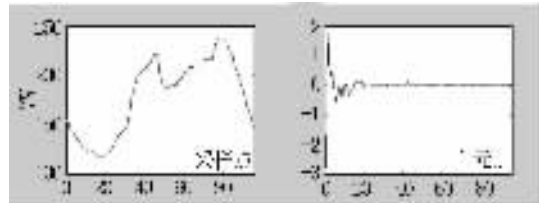


图 3 第 60 天负荷及主成分曲线

表 1 日负荷数据主成分计算结果

序号	特征值	显著水平/%	序号	特征值	显著水平/%
1	84.4489	87.6011	7	0.1962	98.7128
2	5.9247	93.7470	8	0.1616	98.8804
3	2.4212	96.2585	9	0.1540	99.0402
4	1.2581	97.5635	10	0.1342	99.1794
5	0.6095	98.1958
6	0.3022	98.5092	96	0.0002	100.0000

表 2 工作日 24 小时负荷预测值及误差比较

时点	实际负荷 /MW	方法 1		方法 2		时点	实际负荷 /MW	方法 1		方法 2	
		预测值 /MV	相对误差	预测值 /MV	相对误差			预测值 /MV	相对误差	预测值 /MV	相对误差
1	340.2	350	-2.8807	346.7	-1.9106	13	362.4	357	1.4901	368.3	-1.6280
2	311.1	295.3	5.0788	305.6	1.7679	14	384.9	374.6	2.6760	395.2	-2.6760
3	290.1	277.5	4.3433	297.4	-2.5164	15	390.1	383.2	1.7688	396.9	-1.7431
4	275.6	259.2	5.9507	281.9	-2.2859	16	381.7	375.1	1.7291	390.1	-2.2007
5	261.2	270.1	-3.4074	267.7	-2.4885	17	383.4	368.6	3.8602	381.6	0.4695
6	248.1	230	7.2954	239.6	3.4260	18	376.7	388.6	-3.1590	369.4	1.9379
7	222.7	217.3	2.4248	217.4	2.3799	19	364.2	383.3	-5.2444	362.5	0.4668
8	210.9	209.4	0.7112	215.4	-2.1337	20	368.7	345.4	6.3195	373.4	-1.2747
9	262	268.7	-2.5573	268.4	-2.4427	21	398.3	377.5	-5.2444	407.5	-2.3098
10	317.7	328.8	-3.4939	326.3	-2.7070	22	410.9	395.9	3.6505	429.2	-4.4536
11	339.3	347.6	-2.4462	354.2	-4.3914	23	414.4	397.9	3.9817	406.6	1.8822
12	363.9	365.2	-0.3572	370.3	-1.7587	24	393.7	377.7	4.0640	389.2	1.1430
平均相对误差		3.5056		2.1831		平均相对误差		3.5056		2.1831	

应用递推合成 BP 网络建立短期负荷预测模型,其预测误差见表 2,方法 1 为直接用 RBF 神经网络的预测值,方法 2 为本文提出的基于主成分分析的递推合成 BP 网络的预测值。从表中可以看出方法 2 的预测相对误差为 2.1831%,低于方法 1 的 3.5056%,具有较高的预测精度。

4 结束语

本文引入了多元统计学中的主成分分析法来选择短期负荷预测模型的输入特征,避免了各维特征变异程度差异信息的丢失,能准确反映原始数据所包含的全部信息,有效地减少建模的工作量,其短期负荷预测模型具有较高的预测精度。

参考文献:

- [1] 谢宏,陈志业,牛东晓.短期电力负荷预测的数据主成份分析[J].电网技术,2001,24(1):43-46.
[2] 程晔.常用预测方法及评价综述[J].四川师范大学学

报(自然科学版)2002,25(1):70-73.

- [3] 卫贵武,姚恒申,刘正龙.多维 AR(p)与指数平滑组合预测算法的研究及应用[J].西华师范大学学报(自然科学版)2004,25(3):238-241,255.
[4] 张晓星,周淙,任海军,等.基于增量约简算法确定电力负荷预测模型输入参数的研究[J].电力系统自动化,2005,29(13):40-44.
[5] 钟波,周家启,肖智.基于粗糙集与神经网络的电力负荷新型预测模型[J].系统工程理论与实践,2004,24(6):113-119.
[6] 常学降,陈敏,王明生.时间序列分析[M].北京:高等教育出版社,1993.
[7] JOHNSON R A, WICHERN D W.实用多元统计分析[M].陆旋译.北京:清华大学出版社,2001.
[8] 吕佳.基于递推合成 BP 网络的多变量时间序列预测模型[J].重庆师范学院学报(自然科学版)2004,21(1):30-32.

(责任编辑 欧红叶)