

基于混沌因子及相空间重构后的神经网络 短期电价预测的研究

罗欣, 周渝慧, 郭宏榆

(北京交通大学电气工程学院, 北京 100044)

摘要: 影响电价因素众多,但在现实中不可能获得所有信息的资料,在这种信息不完全的情况下,为了更好地提高电价预测精度,通过分析电价和负荷时间序列的混沌性,用C-C方法分别重构其相空间,揭示出其本身蕴涵的规律,并采用数据挖掘技术中的相似搜索技术,挖掘出与预测日变化规律最相似的时间序列作为样本,利用BP神经网络这一具有高度自学习自适应能力的网络,拟合电价序列的重构函数。利用美国PJM电力市场的实际数据进行了实例预测,结果显示有良好的预测精度,并比传统BP网络能更好地预测休息日电价。

关键词: 电价预测; 混沌理论; BP神经网络; 数据挖掘技术; C-C法

Short-term price forecasting based on chaotic property and phase space reconstructed neural networks

LUO Xin, ZHOU Yu-hui, GUO Hong-yu

(School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: There are many factors which can affect the electricity prices, but in fact, we can not get all of the information of these factors. In this state, to improve the predictive precision, this paper analyzes the chaotic property of the price and load time series and reconstructs the attractors using C-C theory. Besides, the similarity search technique in data mining is adopted to find the most similar time series as training date. Then BP neural network which has high ability to self learning and self adapting is used to find the reconstructive function. The actual data of American PJM power market is forecasted based on the theory above. The results have good predictive precision, especially in rest days.

Key words: price forecasting; chaotic property; BP neural network; data mining; C-C theory

中图分类号: TM73; F123.9

文献标识码: A

文章编号: 1003-4897(2008)01-0048-04

0 引言

在电力市场中,电价是整个电力市场的杠杆和支点^[1],在整个电力市场中发挥着核心的作用,各个参与者之间发生的交易和关系都是以电价为基础的,市场参与者只有较为准确地预测到电力价格,才能在激烈的市场竞争中处于有利地位,谋取更多的利益。因此,近年来电价预测已成为电力市场中亟待研究和解决的热点课题之一。

电力市场中,影响电价的因素众多,包括用户负荷需求、系统的可用发电容量、发电商的报价模式、电网运行状况等因素密切相关。如果试图将电价的成因分析透彻,通过确定的预测模型,显然是不现实的。同时在现实市场当中,很多因素比如竞争对手的可用发电容量、机组运行状况等都属于商

业机密,更是不可能得到的。

在这种信息不完全的情况下,本文引入了混沌理论,通过分析负荷序列和电价序列本身的混沌性,直接根据数据序列本身所提取出来的客观规律来进行预测。在准确地重构了两个时间序列的相空间后,结合数据挖掘技术中的相似搜索技术,追踪相空间中临近点的变化趋势,并以BP神经网络为载体,根据美国PJM电力市场提供的电价数据进行了日前短期电价预测研究,并验证了该方法的有效性和可行性。

1 电价与负荷的混沌特性分析

1.1 相空间重构及Takens理论^[2]

相空间重构的目的是在高维相空间中恢复混沌吸引子。混沌吸引子意味着由一个混沌系统产生

的轨迹经过一定的时期变化后, 最终会变成一种有规律的运动, 这种特定的轨迹就是混沌吸引子。Takens证明了可以找到一个合适的嵌入维, 即如果延迟坐标的维数 $m \geq 2d + 1$, d 是动力系统的维数, 在这个嵌入维空间里就可以把混沌吸引子恢复出来。

根据Takens定理, 对于时间序列 $x_1, x_2, \dots, x_{n-1}, x_n, \dots$, 如果能选择合适的嵌入维数 m 和时间延迟 τ 重构相空间

$$X_i = (x_i, x_{i-\tau} \dots x_{i-(m-1)\tau}) \quad (1)$$

在拓扑等价的意义下就能恢复吸引子的动力学特性。

1.2 嵌入维数 m 和时间延迟 τ 的选择

嵌入维数 m 和时间延迟 τ 的正确选择具有十分重要的作用。延迟时间太小, 坐标的相关性太强, 难以保证相空间中每个相点都能提供新的信息, 但是若太大, 又难以保证轨迹的连续性。同样, 虽然高维相空间意味着提供更多的信息量, 但会增加计算的复杂度, 甚至会破坏真正的结构关系。目前学术上一种认为 m 和 τ 的选择互不相关, 常见的方法有自相关函数法与互相关函数法。另一种近来兴起的观点认为 m 和 τ 的选择是互相依赖的, 保持 $(m-1)\tau$ 为定值, 常用方法为C-C法^[3]。本文采用近年来公认的C-C法分别对电价和负荷时间序列重构相空间。

C-C方法基本原理是利用嵌入时间序列的关联积分函数同时估计出时间延迟 τ 和时间窗口

$\tau_w = (m-1)\tau$ 。关联积分定义如下:

$$C(m, N, r, t) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(r - d_{ij}), r > 0 \quad (2)$$

其中: N 为数据组的大小, $M = N - (m-1)\tau$, 对于本文中离散序列 $\tau = t$, $d_{ij} = \|X_i - X_j\|$,

$\theta(x) = 0$, 若 $x < 0$; $\theta(x) = 1$, 若 $x \geq 0$ 。

对于时间序列 $\{x_i\}$, 将其分成 t 个不相交的时间序列, 定义每个子序列的 $S(m, N, r, t)$ 为:

$$S(m, N, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, \frac{N}{t}, r, t) - C_s^m(1, \frac{N}{t}, r, t)] \quad (3)$$

根据BDS统计得到的 N 和 m , r 的恰当估计, 通过以下公式:

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t) \quad (4)$$

$$\Delta \bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 [\max\{S(m, r_j, t)\} - \min\{S(m, r_j, t)\}] \quad (5)$$

$$S_{cor}(t) = \Delta \bar{S}(t) + |\bar{S}(t)| \quad (6)$$

选取 $\bar{S}(t)$ 的第一个过零点或者 $\Delta \bar{S}(t)$ 的第一个极小值作为时间延迟 $\tau = t$, 同时选取 $S_{cor}(t)$ 的最小值作为时间窗口 $\tau_w = (m-1)\tau$, 并最终确定 m 的取值。

根据美国PJM电力市场提供的2005年全年的电价和负荷数据用C-C法重构相空间, 通过matlab仿真得到结果。图1, 图2分别为电价时间序列的以 $\Delta \bar{S}(t)$ 和 $S_{cor}(t)$ 为因变量的相空间重构仿真图。

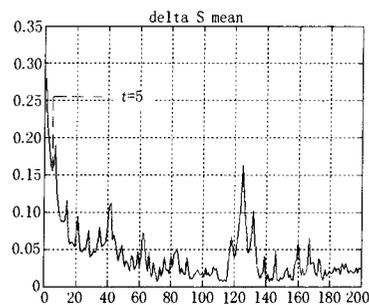


图1 $\Delta \bar{S}(t)$ 确定时间延迟 τ

Fig.1 Delay time τ by $\Delta \bar{S}(t)$

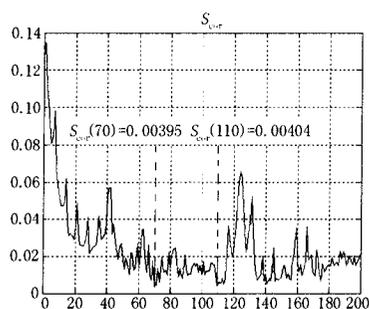


图2 $S_{cor}(t)$ 确定时间窗口 τ_w

Fig.2 The time window τ_w by $S_{cor}(t)$

由上图可得到电价时间序列的时间延迟 $\tau = 5$, 嵌入维数 $m = 15$ 。同理可得到负荷时间序列的时间延迟 $\tau = 5$, 嵌入维数 $m = 16$ 。

根据Takens理论, 必然存在重构函数 f , 使 $x_{i+1} = f(x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})$, 本文通过BP神经网络来拟合重构函数, 即将 $(x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})$ 作为样本, x_{i+1} 作为神经网络的输出来拟合该重构函数并进行预测。

2 BP神经网络简介

人工神经网络^[4] ANN (Artificial Neural Networks) 是由大量简单神经元广泛连接而成的, 用以模拟人脑行为的复杂网络系统。人工神经网络的模型有很多, 常用的有BP模型和Hopfield模型, 其中神经元是神经网络的基本单元, 其输入与输出之间可用函数来表示, 通常称之为激活函数或传递

函数, 每一神经元有一定的内部状态和阈值。

BP 神经网络是目前应用最为广泛的多级网络训练模型。该模型由输入层、隐藏层和输出层组成, 其中隐藏层可自由选择层数, 通常一层隐藏层, 即共三层的神经网络模型即可满足精度要求。其基本思路是误差反向传播的自动学习过程, 即首先假设函数的初始状态, 对样本输入作前向计算, 利用目标输出值与实际值之间的误差, 反向逐层地调节人工神经网络的连接强度和阈值, 逐步地减少输出误差, 直到达到允许的误差要求为止。其基本网络结构见图 3。

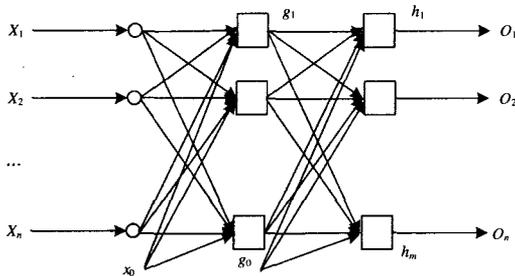


图 3 BP 神经网络网络结构

Fig.3 BP neural network

在用该模型进行预测前, 需要大量的样本进行训练, 在误差满足的条件下, 利用训练好的网络连接强度和阈值进行预测。因此, 训练样本的选择是至关重要的, 训练样本变化趋势越接近测试样本, 预测的精度越高, 反之预测精度很难提高。

传统的样本选择通常认为预测日前几日的变化趋势最接近预测日, 但这只适用于短期内变化趋势一定的时间序列, 但并不适用于电价和负荷序列, 例如在工作日和休息日, 两种序列的变化趋势会有很大的区别, 而如果在预测周六电价时仍采用前几日序列即工作日时间序列做样本, 预测精度很难精确。就此本文在下一节中提出了利用数据挖掘技术中的相似搜索技术来选择样本, 可在一定时间段内, 选择与测试样本趋势最接近的训练样本数据, 可有效地提高预测精度, 即使在休息日也能保持良好的预测精度。

3 电价预测模型的建立

3.1 多变量时间序列相空间重构

根据Takens的理论, 只要选取合理的嵌入维数和延迟时间, 单变量时间序列可以较好地重构相空间, 取得较理想的预测效果。然而, 在实际过程中, 所获得的时间序列长度有限, 且往往存在噪声, 由

单变量时间序列重构的相空间并不能十分准确地描述出动力系统状态变量的演化轨迹^[5]。

在这种情况下, 本文考虑到一方面历史负荷与电价关系最为密切, 另一方面在多数电力市场中, 历史电价和历史负荷数据是较容易获得的。因此, 采用多变量时间序列的相空间重构理论, 结合目前被广泛应用于预测学科的BP神经网络建立电价预测模型。

令电价时间序列记为 $\{x_{1i}\}$, 负荷时间序列为 $\{x_{2i}\}$ 。采用延迟坐标法重构相空间

$$X_i = (x_{1i}, x_{1i-\tau_1}, \dots, x_{1i-(m_1-1)\tau_1},$$

$$x_{2i}, x_{2i-\tau_2}, \dots, x_{2i-(m_2-1)\tau_2}) \quad (7)$$

式中: τ_1 、 τ_2 分别表示电价和负荷时间序列的时间延迟, m_1 、 m_2 分别表示电价和时间负荷序列的嵌入维数。需要注意的是使用的负荷序列均为次日的预测负荷, 因此应用时需要比电价序列超前一个时间单位。

采用三层BP神经网络, 将重构后的电价和负荷序列作为输入单元, 由第2节知电价序列时间延迟 $\tau_1=5$, 嵌入维数 $m_1=15$; 负荷序列时间延迟 $\tau_2=5$, 嵌入维数 $m_2=16$ 。则输入层节点为 $15+16=31$, 隐藏层节点经实验计算为40, 输出节点为1。

3.2 样本数据归一化

本文BP神经网络采用s形函数, 即输出限制在 $(0, 1)$ 之间, 为此, 需要对原始数据进行处理, 分别将电价和负荷序列数据进行预处理。

设 X' 为归一化以后的数据, X 为未进行处理前的数据, X_{\max} 为输入的最大值, X_{\min} 为样本输入的最小值。

$$X' = \frac{(0.9 - 0.1)(X_0 - X_{\min})}{X_{\max} - X_{\min}} + 0.1 \quad (8)$$

经过上述公式处理后, 原始数据归一化到 $[0.1, 0.9]$ 之间。

在模型的输出预测结果时, 只要把输出代入式 (8) 中进行反变换求出原始值即可。

3.3 应用数据挖掘技术选择训练样本

数据挖掘就是从海量数据中提取隐含在其中的、人们事先未知但又有用的信息和知识, 并将其表示成最终能被人理解的模式的高级过程^[8]。通过分析时间序列的相似性从大量历史数据中挖掘出有效信息的技术称为相似搜索技术。本文就是通过该技术搜索与测试样本序列相似的序列作为训练样本来进行预测的。

搜索依据也就是相似度量, 是构造一个能判别

序列相似程度的函数Sim(), 对于任意的两时间序列X和Y, 都可以通过计算Sim(X,Y)来确定X与Y是否相似。通常取两序列的欧几里距离为相似度量, 即

$$d(X,Y) = \left(\sum_{i=1}^n |x_i - y_i|^2 \right)^{\frac{1}{2}} \quad (9)$$

并给定一个阈值, 当距离大于该阈值时认为两个序列不相似, 反之认为相似。这里根据与BP神经网络的结合, 本文提出将重构后的各序列的欧几里距离进行排序, 选择距离最小的前20组序列作为训练样本。

4 实例验证

选择美国PJM电力市场提供的2005年1月1日到11月31日电价和负荷数据分别重构相空间, 对12月5日到11日, 即周一到周日一周每天24小时的电价进行预测, 预测结果与传统BP算法比较如表1所示, 表中法1表示传统BP算法, 法2表示混沌理论与数据挖掘技术同时应用在BP算法中的混合预测算法。

表1 两种算法结果比较

Tab.1 The different results of two methods

预测日		<3%	3%~	5%~	>10%	平均 误差/(%)
			5%	10%		
5日	法1	7	1	14	2	9.56
周一	法2	14	6	4	0	4.12
6日	法1	4	3	11	6	10.51
周二	法2	14	7	3	0	3.08
7日	法1	8	7	8	1	7.13
周三	法2	16	5	3	0	3.04
8日	法1	4	6	9	5	9.25
周四	法2	15	7	2	0	2.48
9日	法1	7	12	5	0	5.29
周五	法2	13	7	4	0	4.44
10日	法1	6	1	5	12	12.14
周六	法2	13	8	3	0	4.23
11日	法1	2	1	4	17	24.72
周日	法2	13	6	5	0	4.32

注: 表中第2-5列数值单位为个数

从表1的结果可以分析如下:

(1) 混沌理论与数据挖掘技术应用到BP神经网络中后, 平均预测精度有了大幅度的提高, 由传统神经网络的10%左右, 提高到5%左右, 并且大大提高了预测精度在5%以内的预测点的个数。

(2) 休息日预测精度有了较大幅度的提高, 这是由于混沌理论的应用使预测不再局限于前几日的历史趋势, 而是深度挖掘时间序列内部客观规律, 比传统神经网络应用更加广泛。

(3) 传统神经网络预测速度较快, 应用混沌理论后由于其遍历性使速度有所下降, 有待进一步探

索。

12月11日, 即周日两种算法电价实际值与预测值曲线图如图4所示。

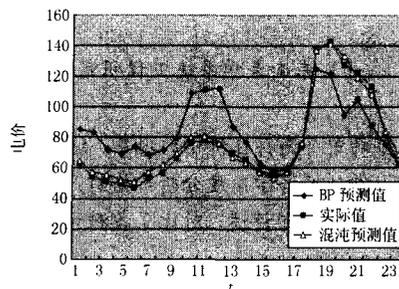


图4 两种方法与实际值比较

Fig.4 Two methods compared with the real price

5 结论

通过以上的的工作, 得到具体结论如下:

(1) 本文应用混沌理论于电价预测领域中, 分析电价与负荷时间序列的混沌性, 采用C-C方法分别重构两序列相空间, 将影响电价变化规律的重要因素——负荷的时间序列与电价的时间序列关联, 使预测模型更加丰满。

(2) 将数据挖掘技术应用到BP神经网络的样本选择中来, 采用其中的相似搜索技术选择与预测日最相似的样本进行训练, 算法结果证明了样本选择的有效性。

(3) 采用美国PJM电力市场的历史数据, 对一周的数据进行了实例预测, 并与传统BP神经网络进行了比较, 证明混沌理论与数据挖掘技术的应用不仅可以提高工作日的精度, 更能大幅度提高休息日的精度。

(4) 进一步地提高算法的收敛速度还有待探讨, 并可通过提高收敛速度进一步提高精度。

参考文献

- [1] 宋守信. 电力市场机制[M]. 北京: 中国电力出版社, 2002年.
SONG Shou-xin. Power Market[M]. Beijing: China Electric Power Press, 2002.
- [2] 吕金虎. 混沌时间序列分析及其应用[M]. 武汉: 武汉大学出版社, 2003.
Lü Jin-hu. Chaotic Times Series Analysis and Application[M]. Wuhan: Wuhan University Press, 2003.
- [3] Kim H S, Eykholt R, Salas J D. Nonlinear Dynamics, Delay Times, and Embedding Windows[J]. Physica D, 1999, 127: 48-60.

(下转第 72 页 continued on page 72)

- Evaluation of Neural Network for Fault Diagnosis[A]. In: Proceedings of Second Symposium on Expert System Application to Power Systems (ESAP'89)[C]. Seattle: 1989.378-384.
- [4] Dy Liacco T E, Kraynak T J. Processing by Logic Programming of Circuit-Breaker and Protective-Relaying Information[J]. IEEE Trans on PAS, 1969, 88(2):171-175.
- [5] 文福拴,邱家驹,韩祯祥. 利用断路器信息诊断电力系统故障的高级遗传算法[J]. 电工技术学报, 1994, 11(2):58-64.
WEN Fu-shuan, HAN Zhen-xiang, et al. A Refined Genetic Algorithm for Fault Section Estimation Using Information from Circuit Breakers[J]. Journal of Electric Machines and Power Systems, 1994, 11(2):58-64.
- [6] Wen Fushuan, Han Zhenxiang. Fault Section Estimation in Power Systems Using a Genetic Algorithm[J]. Journal of Electric Power Systems Research, 1995,34(3): 165-172.
- [7] Wen Fushuan, Chang C S. A New Approach to Fault Diagnosis in Electrical Distribution Networks Using a Genetic Algorithm[J]. Journal of Artificial Intelligence in Engineering, 1998, 12(1): 69-80.
- [8] 文福拴,韩祯祥,等.基于遗传算法的电力系统故障诊断的解析模型与方法——第一部分:模型与方法[J].电力系统及其自动化学报,1998,10(3):1-18.
WEN Fu-shuan, HAN Zhen-xiang, et al. An Analytic Model and Genetic Algorithm Based Method for Fault Diagnosis in Power Systems(1)[J].Proceedings of the EPSA, 1998,10(3):1-18.
- [9] 文福拴 韩祯祥,等.基于遗传算法的电力系统故障诊断的解析模型与方法——第二部分:软件实现[J].电力系统及其自动化学报,1998, 10(3):1-18.
WEN Fu-shuan, HAN Zhen-xiang, et al. An Analytic Model and Genetic Algorithm Based Method for Fault Diagnosis in Power Systems(2)[J]. Proceedings of the EPSA, 1998, 10(3):1-18.
- [10] 韩祯祥,钱源平,文福拴. 基于模糊外展推理和 Tabu 搜索方法的电力系统故障诊断[J]. 清华大学学报(自然科学版),1999, 39(3):56-60.
HAN Zhen-xiang, QIAN Yuan-ping, WEN Fu-shuan. Tabu Search Approach to Fault Diagnosis in Power Systems Using Fuzzy Abductive Inference[J]. J Tsinghua Univ (Sci&Tech), 1999, 39(3):56-60.
- [11] Yogendra Kumar, Biswarup Das , Jaydev Sharma. Multi-objective Reactive Tabu Search for Service Restoration in Distribution System[J]. IEE Proc. Generation, Transmission & Distribution.
- [12] Glover F, Memillan C, Novick B. Interactive Decision Software and Computer Graphics for Architerctural and Space Planning[J]. Annals of Operations Research, 1985: 557-573.
- [13] Glover F, Kelly J, Laguna M. Genetic Algorithms and Tabu Search: Hybrids for Opimization[D]. Boulder University of Colorado, 1992.
- [14] Battiti. The Reactive Tabu Search[J]. ORSA Journal on Computing, 1994, 6(2):126-140.

收稿日期: 2007-09-07

作者简介:

李文清(1982-), 女, 硕士研究生, 主要研究方向为电力系统继电保护; E-mail:liwqstudent@163.com

刘前进(1967-), 男, 副教授, 主要研究方向为电力系统继电保护及电力系统智能控制;

姜伟(1982-), 男, 硕士研究生, 主要研究方向为电力系统故障诊断。

(上接第 51 页 continued from page 51)

- [4] 牛东晓,等. 电力负荷预测技术及其应用[M]. 北京: 电力工业出版社, 1998.
NIU Dong-xiao, et al. Study and Application on Load forecasting[M].Beijing: Electric Power Industry Press, 1998.
- [5] 杨洪明, 段献忠. 电价的混沌特性分析及其预测模型研究[J]. 电网技术, 2004, 28(3): 59-64.
YANG Hong-ming, DUAN Xian-zhong. Research on Chaotic Characteristics of Electricity Price and Its Forecasting Model[J].Power System Technology, 2004,28(3):59-64.

- [6] Han J, Kamber M. Data Mining Concepts and Techniques[M].Beijing: China Machine Press, 2001.

收稿日期: 2007-05-21; 修回日期: 2007-08-08

作者简介:

罗欣(1983-), 女, 硕士研究生, 从事电力市场运营和电价预测方面的研究; E-mail:05121630@bjtu.edu.cn

周渝慧(1955-), 女, 教授, 从事电力市场, 电价分析和电力节能等方面的研究;

郭宏榆(1983-), 男, 硕士研究生, 从事电力市场软件应用及嵌入式系统开发的研究。