

区间组合模型在中长期负荷预测中的应用

徐志向

(四川省电力工业调整试验所,四川成都 610072)

摘要: 由于年负荷的发展包含稳定增长趋势成分和随机成分,因此针对单一中长期负荷预测模型通常难以达到理想的预测精度,而组合模型可以对不同的预测模型进行优化组合和信息的综合利用,提出建立基于支持向量机的变权组合预测模型,实现了组合模型的结构风险最小化代替传统的经验风险最小化。同时针对以往预测模型仅仅实现点预测的不足,建立基于区间参数估计理论的中长期负荷区间组合预测模型,可以解决以往负荷模型预测的无精度范围的问题,对传统的中长期负荷预测模型进行了拓展。最后将该模型应用于实际负荷预测中,验证了该方法的有效性和可靠性。

关键词: 负荷预测;支持向量机;结构风险最小化;区间组合预测模型;区间值

Abstract: Because the annual load includes some steady increasing trend components and some stochastic components, it is difficult for a single medium and long term load forecasting model to achieve the desired forecasting accuracy, while the combination model can realize the optimal combination for the different forecasting models and can synthetically utilize the forecasting information of every model, so the time series variable weight combinational forecasting model based on support vector machine is proposed which can realize structural risk minimization of the combination forecasting model instead of the traditional experience risk minimization. At the same time, aiming at the shortcomings of the former load forecasting model which usually just realizes the point forecasting, the interval combination forecasting model of medium and long term load is established based on interval parameter estimation theory, which can solve the problem of the former forecasting model without the range of precision and develop the traditional forecasting model of medium and long term load. Finally, this model is utilized to the practical load forecasting, and the validity and reliability are verified.

Key words: load forecasting; support vector machine; structural risk minimization; interval combination forecasting model; interval value

中图分类号: TM715 文献标志码: A 文章编号: 1003-6954(2013)01-0064-06

0 引言

电力系统中长期的负荷预测是电力系统运行、规划等工作的重要基础,对保证电网运行的经济性尤为重要。准确的负荷预测有助于提高系统的安全性与稳定性,从而提高经济效益和社会效益。传统的预测方法有趋势外推法、时间序列法、回归分析法和灰色模型法等^[1-4]。由于年电力负荷由确定性和随机性成分组成,确定性成分主要由国民经济发展、长期气候变化、行政调控等因素决定,随机成分由偶然因素导致,具有不确定性,因此这些方法显得不够有效。近年来,基于人工智能理论如神经网络、支持向量机等方法得到了成功应用^[5-6]。但单个预测模型进行预测仍存在一些缺陷,如信息源的单一性、对

模型参数敏感等,而组合预测模型可以综合各模型的优势,实现更准确的负荷预测。

组合预测就是将不同的预测方法进行适当的组合,综合利用各种方法所提供的有用信息,从而尽可能地提高预测精度^[7]。大量的实验表明,组合预测往往优于单一预测模型^[8-10]。文献[11]综合多个准则对组合模型预测效果进行评价。文献[12]考虑了各预测方法各时点上的预测精度状态,通过加权马尔可夫链定性地推测出预测年份上各单项预测方法的预测精度状态,确定其在预测年份的权系数进而预测。文献[13]借鉴层次分析法的思想,采用方差-协方差优选组合预测方法和灰色关联分析分别确定各单一预测模型在各评价指标下的相对权重,最终确定组合预测模型中的组合权重。

但是目前所提出的组合预测方法以及预测

模型往往均是建立在实际值序列和预测值序列都为单点值基础上的。由于经济发展的复杂性,影响实际负荷变化的因素很多,同时对于未来负荷的规划,掌握负荷的精确变化是不可能的。而且负荷数据在一定范围内变化也是正常的。因此预测模型不但应给出未来的趋势变化,预测结果作为一区间值给出负荷的变化范围更为合理。区间数据的变化形态也越来越受到重视。区间计算的思想始于20世纪30年代,系统的理论体系则在1966年,国际著名计算数学专家R. E. Moore专著《区间分析》的出版,自此以后区间计算在数学、计算机、物理、工程等领域获得了广泛的应用^[14-16]。文献[17]对中长期负荷建立了模糊组合预测模型,并给出了一个区间范围,对区间组合预测仅进行了初步探讨。文献[18]在现有ARIMA电价预测模型的基础上,借助统计学的区间估计理论,给出了一种含置信区间的电价预测方法,但并没有对预测模型误差项进行正态性检验。

通过对中长期电力负荷数据及其预测模型进行深入分析,首先综合神经网络、灰色系统和最小二乘支持向量机各单一预测模型的优势,建立基于支持向量机的变权组合预测模型,实现了组合模型的结构风险最小化代替传统的经验风险最小化。同时采用基于统计推断学的区间参数估计理论,建立组合预测与区间估计理论相结合的电力负荷区间组合预测模型,给出未来电力负荷数据在一定置信度下的变化区间,从而实现了对中长期电力负荷发展趋势更加准确的预测。

1 组合预测模型的构建

首先讨论了灰色系统、神经网络和最小二乘支持向量机3种预测模型,在此基础上提出建立基于支持向量机的变权组合预测模型。

1.1 灰色预测模型

灰色系统理论认为所有随机过程都是在一定幅值范围、一定时区内变化的灰色量,称这个随机过程为灰色过程。灰色模型不需要任何原始序列的概率分布,可实现较少数据的建模^[4]。同时中长期电力负荷预测具有信息不完全和不确定的性质,符合灰色变量的特征,因此可采用GM(1,1)模型预测中长期电力负荷。

1.2 广义神经网络

人工神经网络技术在预测中得到了广泛的应用,其中应用较多的是BP神经网络及其改进网络,但BP神经网络有收敛速度慢、训练时间长、容易陷入局部极小点等缺点。广义神经网络(generalized regression neural network, GRNN)是由Donald F. Specht提出的,是RBF神经网络的一个分支,一种基于非线性回归理论的前馈式神经网络模型^[19]。该网络在逼近能力和学习速度上有较强的优势,预测效果也较好。此外该网络还可以处理不稳定的数据,所以经常用于对非线性函数的逼近。

GRNN神经网络的理论基础是非线性回归分析。设随机变量 x 和 y 的联合概率密度函数为 $f(x, y)$,已知 x 的观测值为 X ,则 y 相对于 X 的回归,即条件均值为

$$\bar{Y} = E[y|X] = \frac{\int_{-\infty}^{+\infty} f(X, y) dy}{\int_{-\infty}^{+\infty} f(X, y) dy} \quad (1)$$

Y 即为在输入为 X 的条件下, Y 的预测输出。如果GRNN神经网络的概率密度函数(隐含层的传递函数)采用高斯函数,则网络的输出为

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i \exp(-\frac{D_i^2}{2\sigma^2})}{\sum_{i=1}^n \exp(-\frac{D_i^2}{2\sigma^2})} \quad (2)$$

式中 $D_i^2 = [(X - X_i)^T (X - X_i)]$; $X_i, Y_i (i = 1, 2, \dots, n)$ 分别为第 i 个取样样本的输入和输出; σ 为光滑因子。

1.3 最小二乘支持向量机

SVM(support vector machine)是由Vapnik提出的一种统计学习方法,是在统计学习理论的VC维理论和结构风险最小原理的基础上发展起来的一种新的机器学习方法^[20]。具有理论完善、适应性强、全局优化、训练时间短和泛化能力强的特点。其基本思想为选择一个非线性变换 $\phi(\cdot)$ 将 n 维输入、1维输出样本向量 $[(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)] (x_i \in R^n, y_i \in R, i = 1, 2, \dots, l)$ l 为训练样本个数,从原空间映射到高维特征空间,并在该高维特征空间构造最优线性回归函数得

$$f(x) = w\phi(x) + b \quad (3)$$

最小二乘支持向量机^[21]在优化目标中选择的损失函数为误差 ξ_i 点的二范数。因此优化问题为

$$\min \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^l \xi_i^2 \quad (4)$$

约束条件为

$$y_i [w^T \phi(x_i) + b] = 1 - \xi_i \quad (i = 1, \dots, l) \quad (5)$$

一般, 由于 w 可能为无限维, 直接求解式的优化问题极其困难, 因此将这一优化问题转化到其对偶空间中, 引入拉格朗日函数得

$$L = \frac{1}{2} w^T w + \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l a_i \{ y_i [w^T \phi(x_i) + b] - 1 + \xi_i \} \quad (6)$$

通过最小二乘法求出 a 和 b , 由此得到预测输出

$$y(x) = \sum_{i=1}^l a_i \phi^T(x) \phi(x) + b \quad (7)$$

1.4 预测模型的最优组合

由于目前各种组合预测方法都是基于经验风险最小化原则, 而支持向量机是基于结构风险最小化原则。同时支持向量机的拓扑结构仅是由支持向量决定, 并且无需确定具体的函数表达式, 较好地解决了小样本、非线性、高维数和局部极小等问题, 具有较强的泛化能力。文献 [22] 对应用支持向量机的组合预测过程和参数计算方法进行了详细的探讨, 这里不再赘述。该模型综合了线性模型、指数模型、乘幂模型、非等间隔灰色 GM(1, 1) 模型和非等间隔灰色 Verhulst 模型。虽然线性模型、指数模型和乘幂模型具有模型简单、计算量较小的特点, 但是实际的负荷数据不是单调线性变化的, 而是波动性较大的数据, 尤其对于非线性发展的中长期电力负荷。因此单调模型不能很好跟踪其变化趋势, 于是提出将广义神经网络、灰色模型与最小二乘支持向量机 3 种不同机理的非线性预测模型首先作为单一预测模型分别对训练集的负荷进行预测, 然后将 3 个模型的拟合结果作为 SVM 的输入, 实际的负荷数据作为输出, 这样就得到下一年的负荷组合预测模型。每次使用组合模型预测都需要对支持向量机进行训练后再预测。因此, 每次模型训练得到的组合权重都是基于结构风险最小化原则的, 隐含于训练模型中。

2 区间预测理论的应用

通过建立的基于支持向量机的变权组合预测模型可以实现对下一时刻的预测, 却无法给出精度范围。由于影响负荷预测对象的因素很多, 某些参数在一定范围内波动属正常现象。因此预测模型不但应给出负荷的未来变化趋势, 更重要的是应能够给

出负荷在某一置信度下的变化范围。

1937 年, Neyman 提出的置信区间 (confidence interval) 理论已被广泛应用, 区间估计和点估计是其重要的组成部分。置信区间估计方法包括枢轴量法、假设检验法、大样本法等。这里采用较常见的枢轴量法构造置信区间。

设 $y_n = [y_{n-k+1} \ y_{n-k+2} \ \dots \ y_{n-1} \ y_n]$ 为电力负荷第 $n+1$ 时刻以前的真实数据, \hat{y}_{n+1} 为电力负荷第 $n+1$ 时刻的预测输出。二者的关系可以表示为

$$\hat{y}_{n+1} = g(y_n, w) + \varepsilon \quad (8)$$

$g(y_n, w)$ 为组合预测模型, 随机变量 ε 为模型输出误差服从 $N(\mu, \sigma^2)$ 分布。可以通过 W 检验、 A^2 检验或 W^2 检验等方法对负荷预测模型的误差进行正态性验证。当确认预测模型的误差服从正态分布的假设, 可得如下推导^[23]。

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \quad (9)$$

$$\text{其中 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i;$$

由上式可得

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad (10)$$

因此统计量 T 服从 $t(n-1)$ 分布。由于 T 是 μ 的点估计 \bar{X} 和 μ 的函数, 又与参数 $\theta = (\mu, \sigma^2)^T$ 无关, 因此有

$$\begin{cases} p_\theta \left[\frac{\sqrt{n}(\bar{X} - \mu)}{S} \leq t_{1-\alpha} \right] = 1 - \alpha \\ p_\theta \left[-t_{1-\alpha} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{S} \right] = 1 - \alpha \end{cases} \quad (11)$$

由此可得 μ 的置信水平为 $1 - \alpha$ 时的双侧同等置信区间为 $[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \quad \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)]$ 在正态分布总体的 UMPU 检验下, 负荷预测模型误差

$$I_w(\varepsilon_{n+1}) = [\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \quad \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)] \quad (12)$$

是 μ 的置信水平为 $1 - \alpha$ 的 UMAU 同等置信预测区间。式 (12) 即为预测模型的误差区间, 结合式 (8) 的组合预测模型, 得出第 $n+1$ 时刻 y_{n+1} 模型的预测区间为

$$I_w(y_{n+1}) = g(y_{n+1}, \mu) + I_w(\varepsilon_{n+1}) \quad (13)$$

3 应用实例

为验证提出的区间组合预测方法,以某省1978—1998年21年年用电量数据^[24]的实际数据为例对该系统的年用电量进行预测,并对预测结果进行评价分析。运用1978—1993年的年用电量数据进行模型的构建,1994—1998年的年用电量数据进行模型的验证。并同时与各单项模型进行了对比。表1为1978—1998年的用电量。

表1 1978—1998年用电量

单位: GWh

年份	负荷值	年份	用电量
1978	80.13	1989	17 267
1979	90.78	1990	17 711
1980	93.95	1991	18 815
1981	94.25	1992	19 437
1982	100.29	1993	21 891
1983	108.15	1994	24 755
1984	116.37	1995	27 350
1985	124.92	1996	29 210
1986	129.83	1997	29 917
1987	151.12	1998	30 408
1988	159.92		

最小二乘支持向量机模型采用了RBF径向基函数作为核函数。

$$K(x, x_k) = -\frac{\|x - x_k\|^2}{2\gamma^2} \quad (14)$$

式中 γ 为核宽。核参数 γ 和惩罚参数 C 采用了交叉验证方法对参数进行寻优($\gamma = 1.61, C = 960$)。GRNN神经网络的光滑因子 $\sigma = 0.01$ 。作为组合模型的支持向量机函数svr采用Quadratic损失函数,通过参数寻优得到RBF核参数 $\gamma = 100, C = 10^4$,误差 $\varepsilon = 0.001$ 。

通过Matlab提供的正态分布检验函数lillietest对所建模型误差项均进行正态性检验, $\alpha = 0.01$,通过对组合预测模型误差进行假设检验,均得到 $h = 0$ 因此接受负荷组合预测模型误差项服从正态分布的假设。图1为含有前20个数据量的负荷组合预测模型误差的正态性检验图。从图中可以清楚地看出,所提出的区间组合预测模型的误差基本符合正态性

标准线,可认为该预测模型误差近似服从正态分布,因此可直接对模型误差应用区间估计理论。如果预测模型的误差不能满足正态性分布的要求,可通过Box-Cox变量变换法^[25]将模型数据转换为正态分布样本。

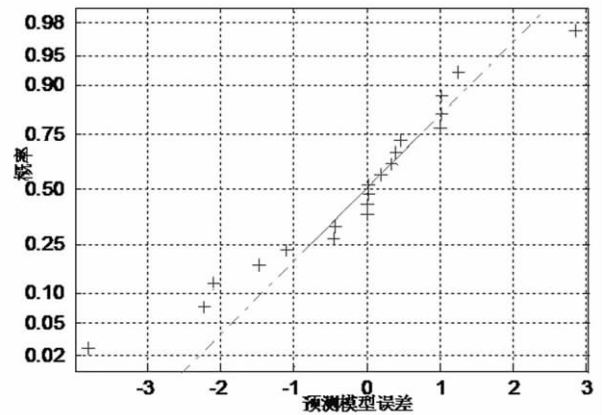


图1 负荷组合预测模型误差的正态性检验图

图2为各负荷预测模型MAPE的变化情况。从图可明显看出随着模型数据量的增加,LSSVM模型和GRNN模型误差基本保持不变,GM模型误差一直处于较高的位置,而所提出的组合预测模型对数

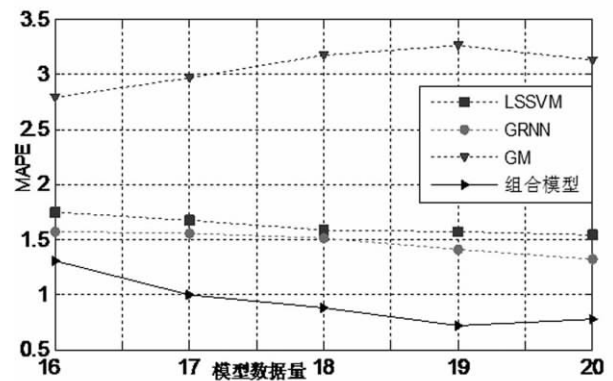


图2 负荷预测模型的MAPE图

据进行了比较好的拟合。主要是由于模型通过SVM进行非线性组合,充分挖掘了原始数据和单一模型的信息和实现了利用组合模型结构风险最小化原则代替传统的经验风险最小化。通过后验误差 ε 和平均绝对百分误差(MAPE)对各模型进行了评价。通过表2的各预测模型的预测结果对比可以看出组合预测方法的 ε 和MAPE都比单一预测方法小,预测可靠性更高。后验误差 ε 定义为

$$\varepsilon = \frac{\hat{y}_{t+i} - y_{t+i}}{y_{t+i}} \quad (15)$$

式中 y_{t+i} 为校验实际值; \hat{y}_{t+i} 为模型预测值。同时以

表 2 各预测模型的预测结果比较 (单位: 100 GkW·h)

年份	实际值	负荷预测值									
		LSSVM	GRNN	GM	组合模型	下限	上限				
1994	247.55	238.45	-0.038 2	216.08	-0.145 6	230.70	-0.073 0	243.53	-0.016 5	237.05	250.02
1995	273.5	278.14	0.016 7	244.19	-0.120 0	252.86	-0.081 6	270.96	-0.009 4	264.98	276.94
1996	292.1	302.794	0.035 3	270.43	-0.080 1	278.30	-0.049 6	292.94	0.002 9	287.74	298.15
1997	299.17	309.91	0.034 7	289.88	-0.032 0	303.53	0.014 4	304.15	0.016 4	299.45	308.85
1998	304.08	313.93	0.031 4	298.29	-0.019 4	324.75	0.063 6	301.23	-0.009 5	296.53	305.93
MAPE			3.12		7.95		5.65		1.09	1.90	1.36

平均绝对百分误差(MAPE)来衡量各模型的拟合精度及预测精度,其定义为

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_{t+i} - y_{t+i}}{y_{t+i}} \right| \cdot 100 \quad (16)$$

从图 3 各模型的负荷预测变化情况可看出,所提出的变权组合模型预测效果明显优于其他 3 种模型,能较好地预测负荷的发展。图 4 为负荷在 $1 - \alpha$ 置信水平下的预测区间($\alpha = 95\%$)。由图可清晰地看出,负荷的发展均落在了置信范围为 95% 的区间内。由于置信水平越大预测区间包含未来真实数据的概率越大,区间范围更可靠;反之,区间范围更精确。因此在实际中应当根据实际情况选取最优的置信水平。

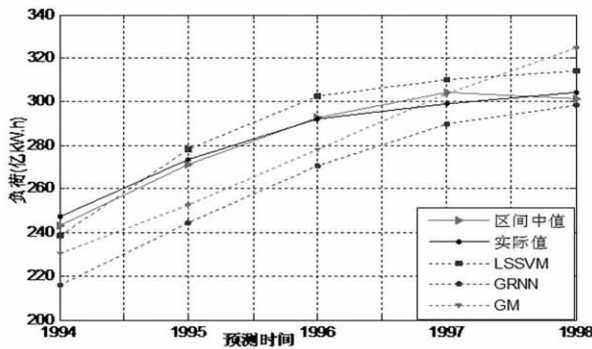


图 3 各模型的负荷预测值

4 结 论

针对单一电力负荷预测模型预测精度不高的情况下,所提出的基于支持向量机的时间序列变权组合预测模型,可以综合各单一预测模型所具有的优势,实现组合模型的结构风险最小化代替传统的经验风险最小化,增强了预测模型的稳定性。同时根据统计推断学的区间参数估计理论,将中长期负荷模型的点预测推广到了区间预测,以一定置信度给

出数据发展的区间预测范围,解决了以往负荷模型预测的无精度范围的问题,对传统负荷预测模型进行了拓展。同时通过与 GRNN、GM、LSSVM 3 种单一模型对实际负荷进行对比预测,实例充分表明所提出的负荷区间组合预测模型能有效提高电力系统负荷预测的精度,具有较好的工程应用价值。区间置信度的选取与负荷的实际发展情况、专家经验等很多因素相关,因此如何根据负荷发展的实际情况选取最优的置信度有待进一步研究。

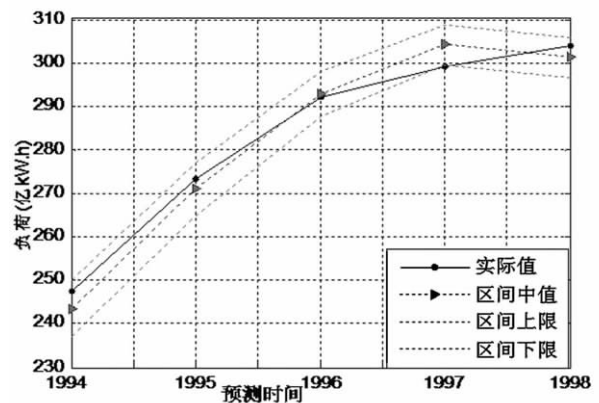


图 4 负荷在 $1 - \alpha$ 置信水平下的预测区间($\alpha = 95\%$)

参考文献

- [1] 牛东晓,曹树华,赵磊,等. 电力负荷预测技术[M]. 北京: 中国电力出版社, 1999.
- [2] 康重庆,夏清,刘梅. 电力系统负荷预测[M]. 北京: 中国电力出版社, 2007: 112 - 114.
- [3] 雷绍兰,孙才新,周涪,等. 电力短期负荷的多变量时间序列线性回归预测方法研究[J]. 中国电机工程学报, 2006, 26(2): 25 - 29.
- [4] 张友泉. 一种基于灰色系统理论的中长期电量预测模型[J]. 电网技术, 1999, 23(8): 47 - 50.
- [5] 李光珍,刘文颖. 基于 LSSVM 和马尔可夫链的母线负

荷短期预测[J]. 电力系统保护与控制, 2010, 38(11): 56-58.

[6] 牛东晓, 王建军, 李莉, 等. 基于粗糙集和决策树的自适应神经网络短期负荷预测方法[J]. 电力自动化设备, 2009, 29(10): 30-34.

[7] Bates J M, Granger C W J. Combination of Forecasts[J]. Journal of Operational Research Quarterly, 1969(20): 451-468.

[8] 方仍存. 电力系统负荷区间预测[D]. 武汉: 华中科技大学, 2008.

[9] 马永开, 唐小我. 两种组合预测优化模型的分析 and 比较[J]. 电子科技大学学报, 1998, 27(1): 99-103.

[10] Wang Jianzhou, Zhu Suling, Zhang Wenyu, et al. Combined Modeling for Electric Load Forecasting with Adaptive Particle Swarm Optimization [J]. Energy, 2010(35): 1671-1678.

[11] 周四清, 王坚强. 基于多准则优化的组合预测方法[J]. 系统工程与电子技术, 2009, 31(7): 1651-1654.

[12] 毛李帆, 姚建刚, 金永顺, 等. 中长期电力组合预测模型的理论研究[J]. 中国电机工程学报, 2010, 30(16): 53-58.

[13] 周淦, 任海军, 李健, 等. 层次结构下的中长期电力负荷变权组合预测方法[J]. 中国电机工程学报, 2010, 30(16): 47-52.

[14] Zhao Junhua, Dong Zhaoyang, Xu Zhao, et al. A Statistical Approach for Interval Forecasting of the Electricity Price[J]. IEEE Trans. on Power System, 2008, 23(2): 267-276.

[15] Khosravi A, Nahavandi S, Creighton D. Construction of Optimal Prediction Intervals for Load Forecasting Problems[J]. IEEE Trans. on Power System, 2010, 25(3): 1496-1503.

[16] 徐惠莉, 吴柏林, 江韶珊. 区间时间序列预测准确度探讨[J]. 数量经济技术经济研究, 2008, 25(1): 133-140.

[17] 游仕洪, 程浩忠, 谢宏, 等. 模糊组合预测在中长期负荷预测中的应用[J]. 电力系统及其自动化学报, 2004(3): 54-56.

[18] 曾鸣, 刘玮, 汪晓露. 含置信区间的改进 ARIMA 电价预测[J]. 电力系统保护与控制, 2009, 37(18): 25-31.

[19] Donald F, Specht. A General Regression Neural network [J]. IEEE Trans Neural Netw, 1991, 2(6): 568-576.

[20] Cortes C, Vapnik V. Support Vector Networks [J]. Machine Learning, 1995, 20(1): 273-297.

[21] Suykens J A K, Gestel V T. Least Squares Support Vector Machines [M]. World Scientific Publishing Company, 2003.

[22] 赵文清, 朱永利, 张小奇. 应用支持向量机的变压器故障组合预测[J]. 中国电机工程学报, 2008, 28(25): 14-19.

[23] 安伟光, 孙振明, 张辉. 预测区间技术在航天器数据处理中的理论与应用研究[J]. 宇航学报, 2006(27): 109-112.

[24] 陈攀, 李燕辉, 王文圣, 等. 电力负荷预测的灰色随机组合模型[J]. 四川水力发电, 2004, 23(3): 80-81.

[25] 李建蹦, 吴小俊. 基于 Box-Cox 变换的分类器性能改进[J]. 计算机工程, 2009, 35(23): 172-174.

作者简介:

徐志向(1980), 男, 电气工程专业工学硕士, 工程师, 主要从事电力系统工程技术相关工作及研究。

(收稿日期: 2012-07-25)

(上接第15页)

[17] 张靠社, 王媛, 胡德海. 基于 Crowbar 保护的双馈感应发电机组的低电压穿越研究[J]. 电网与清洁能源, 2011, 10(27): 66-71.

[18] 孔宪国, 刘宗歧. 基于主动式 Crowbar 的双馈风电机组 LVRT 性能优化分析[J]. 现代电力, 2012, 29(1): 77-81.

[19] 苏平, 张靠社. 基于主动式 IGBT 型 Crowbar 的双馈

风力发电系统 LVRT 仿真研究[J]. 电力系统保护与控制, 2010, 23(38): 164-171.

作者简介:

魏丽丽(1987), 女, 硕士研究生, 研究方向为洁净能源与并网技术;

晁勤(1959), 女, 教授, 博士生导师, 研究方向为并网型风力发电系统及其控制方面。

(收稿日期: 2012-07-09)