

Wavelet Gradient Boosting Regression Method Study in Short-Term Load Forecasting

Hong Wang, Gendai Gu

College of Mathematics and Science, North China Electric Power University, Baoding Hebei
Email: wanghong1566216@163.com, ggd999@ncepu.edu.cn

Received: Jul. 30th, 2015; accepted: Aug. 17th, 2015; published: Aug. 20th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The authors proposed gradient boosting regression method based on wavelet transform considering the influence of weather factors and the characteristics of the load and meteorological data. The load and meteorological data were decomposed into several subsequences in different band by wavelet transform respectively. Forecasting the load subsequence by building different gradient boosting regression model, lastly, the final forecasting result is attained via adding all child-load-serials forecasting results. It has been showed by load data of a city in north China that the method achieved good prediction accuracy.

Keywords

Electrical Power System, Short-Term Load Forecasting, Wavelet Transform, Gradient Boosting Regression

电力系统短期负荷的WGBR预测方法研究

王 虹, 谷根代

华北电力大学数理学院, 河北 保定

Email: wanghong1566216@163.com, ggd999@ncepu.edu.cn

收稿日期: 2015年7月30日; 录用日期: 2015年8月17日; 发布日期: 2015年8月20日

文章引用: 王虹, 谷根代. 电力系统短期负荷的 WGBR 预测方法研究[J]. 智能电网, 2015, 5(4): 189-196.
<http://dx.doi.org/10.12677/sg.2015.54023>

摘要

考虑到气象因素对电力负荷的影响，并结合负荷和气象数据的特点，本文提出了基于小波变换的梯度 Boosting 回归(Gradient Boosting Regression based on Wavelet transform, WGBR)预测方法。该方法将负荷和气象数据分别分解为不同频段的小波子序列，对各子序列分别利用 GBR 算法建模预测，最后综合得到负荷序列预测结果。经华北地区某市的负荷数据验证得本文方法取得了较好的预测精度。

关键词

电力系统，短期负荷预测，小波变换，梯度 Boosting 回归

1. 引言

短期负荷预测一般是预测未来 1 小时至 1 周的负荷，对电力系统的运行和生产有着重要的意义。目前，用于短期负荷预测的方法很多，总体分为传统和现代智能两类方法。传统的短期负荷预测方法有回归分析法、时间序列法等。实践表明，传统方法在应用中存在局限性，如负荷预测的输入和输出间呈非线性关系，就很难找到理想的预测模型[1]。因此，现代智能短期负荷预测方法应运而生，如神经网络法、小波分析法、支持向量机、灰色系统预测法等。

电力负荷具有不稳定性、非线性，且表现为以天、周、年为周期发生波动，大周期中套有小周期这些特殊的周期性。小波变换能将其呈周期性变化的子负荷序列分离出来，这样更有利于具有周期性的短期负荷预测精度的提高。小波分析在电力系统短期负荷预测中已得到广泛应用。

GBR 算法是 Friedman 提出的将 Boosting 算法扩展以解决回归问题的预测算法，在多个领域有着广泛应用，如生物研究、搜索排名和机器学习等。它对数据具有很强的鲁棒性，能够很好地处理不干净的、含噪声的数据，支持不同的损失函数，并对非线性数据有很强的预测能力。国内外研究人员将 GBR 算法应用于太阳能领域[2]的预测、汽车保险损失成本预测[3]、碳钢的土壤腐蚀规律研究[4]等。但在电力系统负荷预测领域应用还很少。

本文将小波变换与梯度 Boosting 回归相结合进行短期负荷预测。利用小波变换将负荷序列及气象序列(温度、湿度、气压、风速)分别分解到不同的频段上，选取输入特征向量对每个负荷分量建立基于相应气象因素影响的 WGBR 模型进行预测，最后综合得到负荷序列的最终预测结果。实例证明应用本文方法对电力系统短期负荷进行预测取得了较高的预测精度。

2. GBR 算法

Boosting 算法是基于统计学习理论的一种重要的机器学习方法，近年来在模式识别，回归分析，特征提取等方面得到了广泛的应用[5]，其中由 Freund 和 Schapire 提出的 AdaBoost (Adaptive Boost) 算法分类在理论分析和实际应用中取得了巨大的成功[6]。因此一些研究者试图将分类方法引入到回归问题中。Friedman 将梯度下降方法应用于回归算法中，提出了 Gradient Boosting 算法[7]，该算法利用沿着模型的损失函数的负梯度求解损失函数的最小值，迭代得到最佳模型。

GBR 本质上是一种利用 T 个基函数的加法展开式对目标函数进行逼近的方法。表达式[3]如下所示：

$$f(\mathbf{X}) = \sum_{t=1}^T f_t(\mathbf{X}) = \sum_{t=1}^T \beta_t h(\mathbf{X}; \boldsymbol{\alpha}_t) \quad (1)$$

其中 $h(\mathbf{X}; \boldsymbol{\alpha})$ 为一个简单的参数函数, 称为基函数, 本文中它是决策树。 $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots\}$ 为它的参数, 表示节点分枝变量、分枝规则、根节点的数目等[8]。 β 为展开式系数, T 为基函数的个数, $\mathbf{X} = (x_1, x_2, \dots, x_n)$ 为输入向量, y 为输出。

与其他的 Boosting 算法类似, 为了得到最终的模型, GBR 建立了分步向前可加模型进行迭代, 即

$$f_t(\mathbf{X}) = f_{t-1}(\mathbf{X}) + \beta_t h(\mathbf{X}; \boldsymbol{\alpha}_t) \quad (2)$$

也就是说, 我们要找到使所有样本的损失函数达到最小值时的 $\boldsymbol{\alpha}, \beta$ 即

$$\min_{\{\beta_t, \boldsymbol{\alpha}_t\}_{t=1}^T} \sum_{i=1}^n L\left(y^{(i)}, \sum_{t=1}^T \beta_t h(\mathbf{X}; \boldsymbol{\alpha}_t)\right) \quad (3)$$

其中 $(\mathbf{X}^{(i)}, y^{(i)})$ 为第 i 个样本, n 为样本数, 本文选定的损失函数都为平方损失函数, 即

$$L(y, f(\mathbf{X})) = \frac{1}{2}(y - f(\mathbf{X}))^2 \quad (4)$$

算法伪代码步骤[3] [9]如下:

给定初始模型

$$f_0(\mathbf{X}) = \sum_{i=1}^n y_i / n \quad (5)$$

2) for $t = 1$ to T do

3) 计算负梯度

$$g_i = -\left[\frac{\partial L\left(y^{(i)}, f\left(\mathbf{X}^{(i)}\right)\right)}{\partial f\left(\mathbf{X}^{(i)}\right)} \right]_{f(\mathbf{X})=f_{t-1}(\mathbf{X})} \quad i = \{1, 2, \dots, n\} \quad (6)$$

4) 通过输入向量 \mathbf{X} 及负梯度 g 拟合一个回归树模型 $h(\mathbf{X}; \boldsymbol{\alpha}_t)$, 即可得到 $\boldsymbol{\alpha}_t$

5) 通过最小化

$$L\left(y, f_{t-1}(\mathbf{X}) + \beta_t h(\mathbf{X}; \boldsymbol{\alpha}_t)\right) \quad (7)$$

求得 β_t

6) 更新模型

$$f_t(\mathbf{X}) = f_{t-1}(\mathbf{X}) + \tau \cdot \beta_t h(\mathbf{X}; \boldsymbol{\alpha}_t) \quad (8)$$

其中 τ 为学习率

7) end for

3. WGBR 预测模型

3.1. 预测模型

本文数据来自华北地区某市 2004 年 7 月 5 日至 8 月 22 日整点时刻的负荷数据及相应的气象数据。将 7 月 5 日至 8 月 15 日负荷及气象序列作为训练数据, 来预测未来一周(8 月 16 日至 22 日)各时的电力负荷值。应用 WGBR 方法的电力系统短期负荷预测的具体过程图如图 1 所示。

流程图第一步为利用小波进行数据处理: 本文将原始负荷序列 L 及气象序列(温度(T), 湿度(H), 气压(P), 风速(W))都选用具有时频紧支撑和高正则性的 Daubechies 小波分别进行分解。根据实验经验, 采用 Daubechies5 作为母小波, 对负荷序列及气象序列用小波变换和 Mallat 算法进行 3 层分解。 Lc_3 和 Ld_j ($j = 1, 2, 3$) 分别称为负荷序列在分辨率 2^{-j} 下的近似序列和细节序列。各层细节序列和近似序列是原

始序列在相邻的不同频率段上的成分。

由于用 Mallat 算法把序列分解, 分解后的序列长度减半, 这样各小波系数的序列长度必然不同, 故第二步将负荷分解后的序列进行重构增加序列的长度。即对 LC_3 和 LD_j ($j=1,2,3$) 分别进行重构[10], 得到 LC_3 和 LD_j ($j=1,2,3$)。它们和原始序列 L 的长度一样, 且有

$$L = LC_3 + LD_3 + LD_2 + LD_1 \quad (9)$$

同理可得到气象序列各自的重构序列: 温度重构序列 TC_3 , TD_3 , TD_2 , TD_1 , 湿度重构序列 HC_3 , HD_3 , HD_2 , HD_1 , 气压重构序列 PC_3 , PD_3 , PD_2 , PD_1 , 风速重构序列 WC_3 , WD_3 , WD_2 , WD_1 。

第三步为特征提取, 此步骤为预测序列构造特征以便 GBR 建模预测。通过对实时气象序列与负荷序列之间的相关性分析, 选取实时温度、实时湿度、实时气压、实时风速作为 GBR 模型的气象因素输入量。负荷序列本身又具有很强的自相关性及周期性。故选取的样本输入量如表 1 所示。标号(0), (-N), (0)-M, (-1)-M 分别表示预测时刻, 预测时刻前 N 个时刻, 预测前 M 天预测点的同一时刻, 预测前 M 天预测点的上一时刻。即表 1 所示的特征向量输入由预测时刻前 11 个时刻的负荷值, 前 11 天每天同一时刻的负荷值, 前 9 天预测时刻的上一时刻的负荷值, 预测时刻前 5 个时刻的温度值, 预测时刻前 5 个时刻的湿度值, 预测时刻前 2 个时刻的气压值及预测时刻前 2 个时刻的风速值组成。

第四步利用 GBR 预测重构的负荷序列 LC_3 , LD_3 , LD_2 , LD_1 , 对应的特征见第三步。我们的历史数据实际上包括 49 天的负荷数据及相应的气象数据, 我们把前 42 天的负荷及气象数据作为 GBR 模型的训练数据集, 后 7 天的负荷及气象数据作为测试预测数据集。利用训练集建立 GBR 模型, 将此模型应用到测试集进行预测, 得到各个预测负荷序列, 最后综合得到负荷序列的最终预测结果。

3.2. GBR 的实现环境及参数说明

在利用 Mallat 算法对数据进行小波分解后, 在 Spyder 环境下使用 Python 的 scikit-learn 包来实现 GBR 算法[2]。

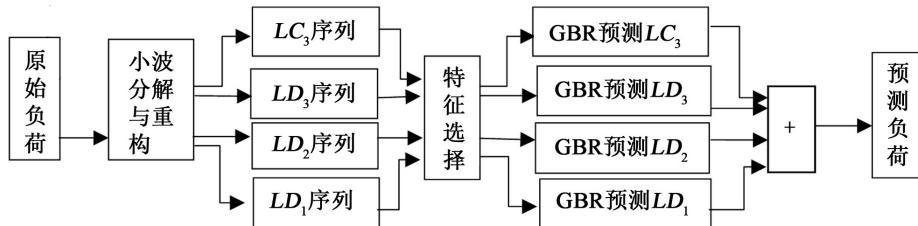


Figure 1. Short-term load forecasting model

图 1. 短期电力负荷预测流程图

Table 1. List of input features for different series

表 1. 各序列预测值选取的输入量

序列	特征输入	预测值
LC_3	$LC_3(-1) \sim LC_3(-11)$, $LC_3(0) - 1 \sim LC_3(0) - 11$, $LC_3(-1) - 1 \sim LC_3(-1) - 9$ $TC_3(-1) \sim TC_3(-5)$, $HC_3(-1) \sim HC_3(-5)$, $PC_3(-1) \sim PC_3(-2)$, $WC_3(-1) \sim WC_3(-2)$	$LC_3(0)$
LD_3	$LD_3(-1) \sim LD_3(-11)$, $LD_3(0) - 1 \sim LD_3(0) - 11$, $LD_3(-1) - 1 \sim LD_3(-1) - 9$ $TD_3(-1) \sim TD_3(-5)$, $HD_3(-1) \sim HD_3(-5)$, $PD_3(-1) \sim PD_3(-2)$, $WD_3(-1) \sim WD_3(-2)$	$LD_3(0)$
LD_2	$LD_2(-1) \sim LD_2(-11)$, $LD_2(0) - 1 \sim LD_2(0) - 11$, $LD_2(-1) - 1 \sim LD_2(-1) - 9$, $TD_2(-1) \sim TD_2(-5)$, $HD_2(-1) \sim HD_2(-5)$, $PD_2(-1) \sim PD_2(-2)$, $WD_2(-1) \sim WD_2(-2)$	$LD_2(0)$
LD_1	$LD_1(-1) \sim LD_1(-11)$, $LD_1(0) - 1 \sim LD_1(0) - 11$, $LD_1(-1) - 1 \sim LD_1(-1) - 9$ $TD_1(-1) \sim TD_1(-5)$, $HD_1(-1) \sim HD_1(-5)$, $PD_1(-1) \sim PD_1(-2)$, $WD_1(-1) \sim WD_1(-2)$	$LD_1(0)$

需要设置的参数有: 基函数的个数(“n_estimators”), 每棵树的深度(“max_depth”), 学习率(“learning_rate”), 损失函数(“loss”)等。树的最大深深度控制特征相互作用的程度。基函数越多, 树越深, 得到的方差就越大, 模型复杂度就越高。学习率是最优正则化过程中的参数, 最优正则化可以简化模型, 使预测结果更稳定。学习率是位于(0,1]的超参数, 它是衡量梯度下降过程的步长。经验表明, 小的学习率有利于得到更好的测试误差。推荐设置学习率为一个常数(如学习率≤0.1) [11]。scikit-learn 包通过及早停止树的增长来避免决策树学习中的过度拟合问题, 比如发现树的深度过高, 叶节点过多, 或者节点内数据过少。本文利用网格搜索法[12]解决参数设定问题, 再进行适当的调整, 找到最佳参数值。网格搜索法的优点是它能同时搜索多个参数值, 最终获取的是使训练误差达到最小时的最优参数组合。同时计算过程中各组参数相互解耦, 便于并行计算, 提高运行效率。损失函数用来衡量当前模型的拟合程度, 不同的损失函数将得到的不同的模型及结果。各预测序列所选则的参数值如表 2 所示。

4. 实例结果分析

为验证本文方法的有效性, 笔者将本文方法与单独使用 GBR 预测方法进行比较, 如图 2, 图 3, 图 4, 表 3, 表 4 所示。选取的误差如下:

绝对百分误差:

$$\varepsilon_{APE} = \left| \frac{y - \hat{y}}{y} \right| \times 100\% \quad (10)$$

平均绝对百分误差:

$$\varepsilon_{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (11)$$

式中: n 为预测时刻数; y_t 为 t 时刻的实际负荷值; \hat{y}_t 为 t 时刻的预测负荷。

图 2 是一周的 WGBR 预测值、GBR 预测值及真实值的比较, 可以看到 WGBR 方法的预测值更接近真实值, 预测效果比 GBR 的预测效果更好。图 3 是两种方法各时预测值的绝对百分误差的对比, 明显看出 WGBR 预测误差较小。表 3 是一周每天的平均绝对百分误差统计, 可以看出 WGBR 误差基本小于

Table 2. Parameter values of WGBR and GBR models

表 2. WGBR 及 GBR 模型参数选择值

预测序列	基函数的个数	树的最大深度	学习率
LC_3	600	4	0.009
LD_3	200	4	0.1
LD_2	200	4	0.1
LD_1	500	2	0.05
GBR 预测负荷	500	4	0.008

Table 3. Statistics of average relative error for a week

表 3. 一周的平均绝对百分误差统计

日期(日/月)	16/8	17/8	18/8	19/8	20/8	21/8	22/8
星期	一	二	三	四	五	六	日
WGBR %	1.59	1.59	1.41	1.37	1.51	1.75	1.55
GBR	3.58	2.61	2.92	2.43	1.83	3.19	2.38

Table 4. Load forecasting result on the Nineteenth day by the WGBR and the GBR methods
表 4. 19 日各时 WGBR 和 GBR 方法对负荷预测的结果

时刻 /h	实际负荷 /MW	WGBR		GBR	
		预测负荷 /MW	\mathcal{E}_{APE} /(%)	预测负荷 /MW	\mathcal{E}_{APE} /(%)
0	1121	1130.8	0.87	1093.4	2.46
1	1034.9	1048.1	1.28	1071.6	3.55
2	1022.6	1040.9	1.79	1017.3	0.52
3	1007.6	1005.5	0.21	1016.3	0.86
4	995.5	975.6	2.00	1007	1.16
5	1022.1	1019.5	0.25	1041.9	1.94
6	1134.2	1128.0	0.55	1108.1	2.30
7	1185.3	1195.9	0.89	1227	3.52
8	1299.9	1278.9	1.62	1291.6	0.64
9	1421.7	1399.2	1.58	1399.5	1.56
10	1442.8	1438.5	0.30	1442.5	0.02
11	1467.3	1471.8	0.31	1442.1	1.72
12	1335.4	1359.5	1.80	1359	1.77
13	1303	1286.1	1.29	1308	0.38
14	1316.9	1305.6	0.86	1355.3	2.92
15	1436.9	1410.6	1.83	1407	2.08
16	1502.4	1464.7	2.51	1442.9	3.96
17	1522.1	1511.7	0.68	1439.4	5.43
18	1516.5	1490.4	1.72	1427.2	5.89
19	1399.4	1435.4	2.57	1402.6	0.23
20	1596.7	1581.6	0.95	1459.2	8.61
21	1533.5	1532.1	0.09	1477.4	3.66
22	1383.2	1416.6	2.41	1368.5	1.06
23	1198.1	1254.2	4.68	1224	2.16

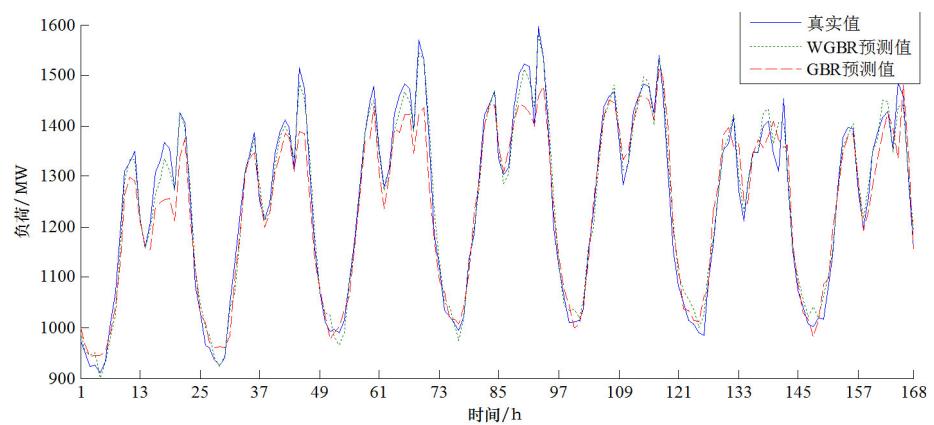


Figure 2. Actual load and forecasting load curve of a week
图 2. 一周的负荷真实曲线与预测曲线

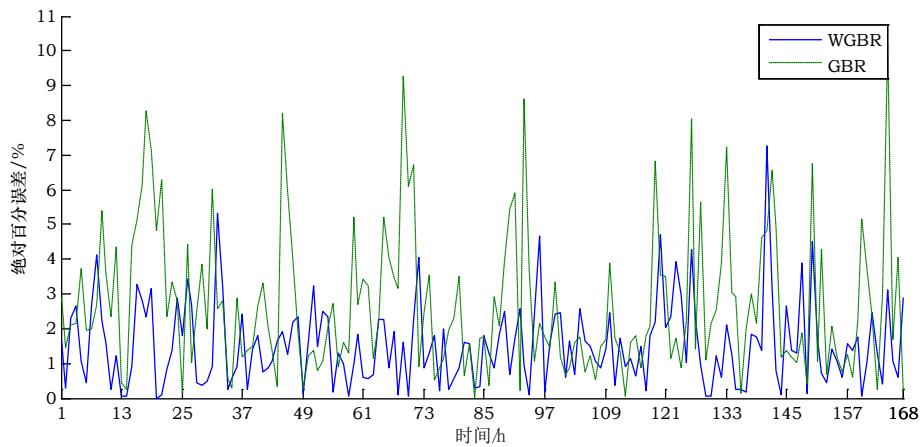


Figure 3. Comparison of forecasting load errors for a week
图 3. 一周的负荷预测误差比较

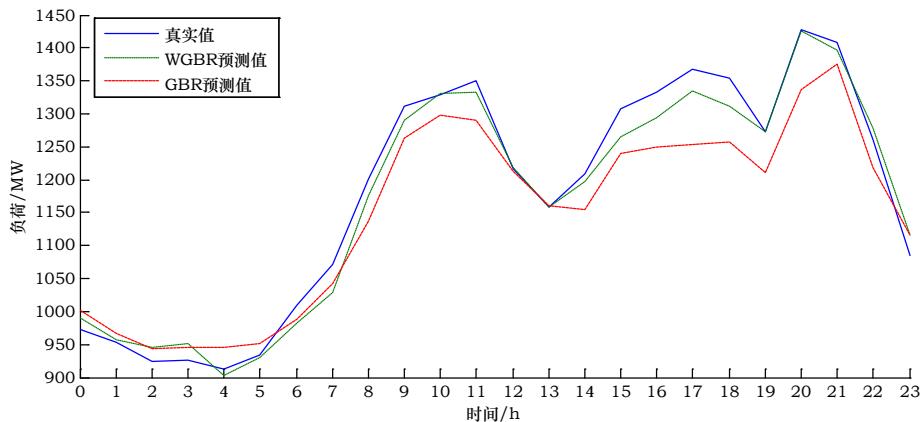


Figure 4. Actual load and forecasted load curve of a Monday
图 4. 周一各时的负荷真实曲线与预测曲线

1.5%，取得了较好的预测精度。表 4 是 8 月 19 日(周四)各时的预测值与真实负荷值。图 4 是周一各时的预测值与真实值的曲线图，能够较清晰地看出 WGBR 的预测效果。

5. 结论

本文提出了一种基于小波变换的梯度 Boosting 回归预测电力系统短期负荷的方法，对不同频段的小波分量采用不同参数模型的 WGBR 方法进行预测更具针对性。通过对比 GBR 方法的预测效果，结果表明，本文的预测得到了更好的预测精度，其误差基本小于 1.5%。小波分析强劲的数据处理能力及 GBR 算法高效的预测能力显而易见。

基金项目

本文工作由国家自然科学基金(11401208)项目资助完成。

参考文献 (References)

- [1] 葛少云, 贾鸥莎, 刘洪 (2012) 基于遗传灰色神经网络模型的实时电价条件下短期电力负荷预测. 电网技术, 1, 224-229.
- [2] Prakash, A. and Singh, S.K. (2014) Towards an efficient regression model for solar energy prediction. 2014 Innovative

Applications of Computational Intelligence on Power Energy and Controls with Their Impact on Humanity (CIPECH), 18-23. <http://dx.doi.org/10.1109/CIPECH.2014.7019040>

- [3] Guelman, L. (2012) Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, **39**, 3659-3667. <http://dx.doi.org/10.1016/j.eswa.2011.09.058>
- [4] 鲁庆, 穆志纯 (2014) 应用提升回归树研究碳钢的土壤腐蚀规律. *中南大学学报(自然科学版)*, **6**, 1879-1886.
- [5] 高云龙, 潘金艳, 吉国力, 高峰 (2011) 基于 Boosting 梯度下降理论的时间序列建模方法. *中国科学: 技术科学*, **7**, 929-943.
- [6] 乐斌 (2004) Boosting 算法研究及其在光谱分析中的应用. 硕士论文, 浙江大学, 杭州.
- [7] 李航 (2012) 统计学习方法. 清华大学出版社, 北京.
- [8] 吴喜之, 马景义, 吕晓玲, 闫洁 (2009) 数据挖掘前沿问题. 中国统计出版社, 北京.
- [9] Friedman, J.H. (2001) Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, **5**, 1189-1232. <http://dx.doi.org/10.1214/aos/1013203451>
- [10] 徐科, 徐金梧, 班晓娟 (2001) 基于小波分解的某些非平稳时间序列预测方法. *电子学报*, **29**, 566-569.
- [11] Hastie, T., Tibshirani, R. and Friedman, J. (2009) Elements of statistical learning: Data mining inference, and prediction. Springer-Verlag New York Inc., New York. <http://dx.doi.org/10.1007/978-0-387-84858-7>
- [12] 王兴玲, 李占斌 (2005) 基于网格搜索的支持向量机核函数参数的确定. *中国海洋大学学报(自然科学版)*, **35**, 859-862.