

# 基于选择性集成 ARMA 组合模型的 零售业销量预测

常炳国<sup>1</sup>, 臧虹颖<sup>1</sup>, 廖春雷<sup>2</sup>, 毛丹华<sup>2</sup>

(1. 湖南大学 信息科学与工程学院, 长沙 410082; 2. 快乐购股份有限公司, 长沙 410003)

**摘要:** 准确预测商品销量的走向对零售企业具有重要意义, 构建自回归移动平均模型 (ARMA 模型, Auto-Regressive and Moving Average Model) 对零售商品时序销量数据进行预测分析; 传统 ARMA 模型无法准确描述商品销量中同时存在的非平稳非线性特征; 论文分别采用支持向量回归 (SVR, Support Vector Regression) 方法和极限学习机 (ELM, Extreme Learning Machine) 方法, 对时序模型中非线性误差进行预测并进行误差补偿, 提高了商品销量的预测精度; 提出了遗传优化的选择性集成定阶方法, 用以简化 ARMA 模型的复杂定阶过程, 降低了对数据平稳性程度要求; 论文收集了某电商平台商品销量数据, 对 ARMA、选择性集成 ARMA、ARMA-SVR、ARMA-ELM 四种预测模型的性能进行了对比分析, 结果表明, 选择性集成 ARMA 模型预测精度在平稳和非平稳时序数据下分别提高 23.58% 和 41.28%; 组合模型相比仅采用线性平稳时序模型的预测结果更符合实际, 其中, ARMA-SVR 模型在小样本、非平稳时序下预测精度比 ARMA-ELM 模型高出约三分之一。

**关键词:** 零售销量预测; 非平稳时序; 误差补偿; 自回归移动平均模型; 遗传优化算法

## Retail Sales Combination Forecasting Model Based on Selective Ensembled ARMA

Chang Bingguo<sup>1</sup>, Zang Hongying<sup>1</sup>, Liao Chunlei<sup>2</sup>, Mao Danhua<sup>2</sup>

(1. College of Information Science and Engineering, Hunan University, Changsha 410082, China;

2. Happy Go Share Co., Ltd., Changsha 410003, China)

**Abstract:** It is important for retail enterprises to forecast the sales volume as accurate as possible. By building auto-regressive moving average model (ARMA) can forecast and analyze the retail sales volume. The traditional ARMA model can not describe the unstable and nonlinear characteristics existing in the sales volume. The Support Vector Regression (SVR) method and Extreme Learning Machine (ELM) method are used to predict and compensate the nonlinear errors of ARMA time series model to enhance the accuracy of prediction. And the paper proposed a novel selective ensemble method to determine the parameter of ARMA. This method based on a genetic optimization algorithm and simplified the process of ARMA. We gathered some data to analyze and compare the performance of ARMA, selective ensembled ARMA, ARMA-SVR and ARMA-ELM. The experimental results show that the accuracy of selective ensembled ARMA model are increased by 23.58% and 41.28% in both stable and unstable series. The combination forecasting model is more in line with the actual data than time series forecasting model. Aslo, ARMA-SVR model's accuracy of prediction is about one-third higher than ARMA-ELM.

**Keywords:** retail sales forecast; unstable time series; error compensation; autoregressive moving average model (ARMA); genetic optimization algorithm

## 0 引言

全球化经济时代企业竞争日益强烈, 需要利用有价值的商业数据在竞争中取得优势地位。预测商品的销量有利于提高零售企业自身的库存周转率和运行效率<sup>[1]</sup>。为此, 国内外学者们提出了多种销量预测模型并结合具体实例进行了研究<sup>[2]</sup>。商品的销量随着时间不断变化, 将销量看成是时序数据, 可以通过时序挖掘相关方法对其进行预测研究。常用的时序分析模型有灰色模型<sup>[3]</sup>、马尔可夫模型<sup>[4]</sup>、自回归条件异方差模型<sup>[5]</sup>、贝叶斯向量自回归模型等<sup>[6]</sup>, 其中 ARMA 模型应用最广泛<sup>[7]</sup>。ARMA 模型通过抽取时序数据内部的规律用于时序的周期预

测和趋势分析, 现已广泛应用于市场销售额预测、消费模式变迁等长期追踪资料的研究<sup>[8]</sup>。传统的 ARMA 模型是线性回归模型, 只能捕捉平稳时序中的线性模式, 但是商品销量的变化是受多种因素影响的, 在其变化过程中同时存在非平稳非线性变化模式, 因此, 单一建立 ARMA 模型无法准确描述商品销量的变化趋势。为了克服传统时序线性模型预测精度低的问题, 多种非线性回归预测方法被提出, 其中表现较好的有 SVR 方法<sup>[9]</sup>以及 ELM 方法<sup>[10]</sup>。从理论上说, 构建非线性模型能以无限精度拟合原始序列, 但是需要大量的训练样本, 并且容易陷入过拟合。因此, 将传统的线性模型与智能非线性方法整合预测商品销量是有必要的, 已有许多研究表明, 通过整合不同的模型能显著提升预测性能。Yan 和 Ma 认为, ARIMA (Auto-Regressive Integrated Moving Average) 与径向基函数网络 (RBFN) 的异构组合模型性能要优于 ARIMA 和 RBFN 自身, 并在月降雨量预测中得到验证<sup>[11]</sup>。Babu 和 Reddy<sup>[12]</sup>使用 ARIMA 滤波器探索了自然种类的波动情况, 并将

收稿日期: 2017-08-15; 修回日期: 2017-10-23。

基金项目: 湖南省重点研发计划资助(2016GK2050)。

作者简介: 常炳国(1963-), 男, 陕西榆林人, 副教授, 硕士研究生导师, 主要从事信息集成方法与应用研究。

神经网络 (ANN) 模型进行了适当的应用。Bang-zhuzhu<sup>[13]</sup> 提出了一种新颖的异构 ARIMA 和最小二乘支持向量机 (LLSVM) 方法, 用于预测碳价格的波动情况; Wang 等人<sup>[14]</sup> 提出一种异构神经网络和支持向量机的模型用于地表 PM<sub>10</sub> 和 SO<sub>2</sub> 日浓度预测。

针对传统的时序模型无法捕捉商品销量中的非平稳和非线性特征, 论文提出一种将时序线性回归模型与智能非线性回归方法组合的预测模型。首先利用 ARMA 模型识别出商品销量数据中的线性成分, 再对其残差中的非线性成分分别采用 SVR 和 ELM 方法进行预测, 两部分组合后得到最终预测结果。由于 ARMA 模型存在平稳性假定且定阶过程复杂、随机性强, 论文提出一种选择性集成多个差异 ARMA 模型的定阶方法。经过某电商平台真实销售数据验证, 论文提出的选择性集成 ARMA 定阶方法定阶更准确, 改善了 ARMA 模型在平稳及非平稳时序下的预测效果。组合方法得到的销量预测结果相比仅采用线性平稳时序模型更符合实际, 稳定性更好, 因而验证了提出的组合模型具有一定的参考价值。

### 1 销量影响因素及组合预测过程

通过分析消费者的购买行为, 发现诸多商品之间的销量变化是相互影响的, 不同商品之间存在着竞争和合作关系<sup>[15]</sup>。某一商品的销量受其同类型商品销量的影响、关联商品销量的影响以及商家营销推广活动的影响。从内在影响因素映到外在表现, 销量变化呈现出一定的发展模式和潜在规律。通过挖掘历史销量数据中潜在的规则和模式, 建立数学模型对其进行描述, 可以达到预测销量未来走势的目的。

根据销量数据的时序特征, 论文采用 ARMA 模型对其中的线性因素、季节性周期性规律、随机因素进行预测。对除此之外的可能存在的突发因素及非线性变化部分, 仅基于统计时序模型来拟合和预测会产生较大的非线性误差, 故将影响商品销量变化过程看成是平稳随机过程以及非平稳非线性噪音的叠加。先通过 ARMA 模型识别出销量数据中的线性成分, 然后采用不同的非线性回归预测方法对残差数据作出预测并将预测结果补偿给 ARMA 模型。其误差补偿示意图见图 1。

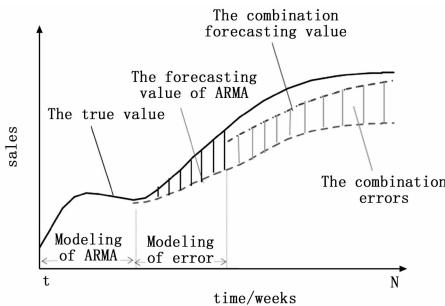


图 1 商品时序预测误差补偿示意图

如图 1 所示, 本文先根据历史销量数据建立 ARMA 模型。假设观测数据为一维, 当前时间为  $t$ , 总采样时间为  $N$ , 则建模后依次得到  $U_{t+1}, U_{t+2}, \dots, U_N$  个初步预测值。将初步预测值与真实值相减得到 ARMA 模型的非线性残差。通过滑动窗口法获得前  $d$  维销量数据作为 SVR 和 ELM 方法的输入, 得到  $t+1, t+2, \dots, N$  时刻的误差预测值  $E_{t+1}, E_{t+2}, \dots, E_N$ , 最后将初步预测值和误差预测值叠加得到商品销量的最终预测值。具体流程见图 2。

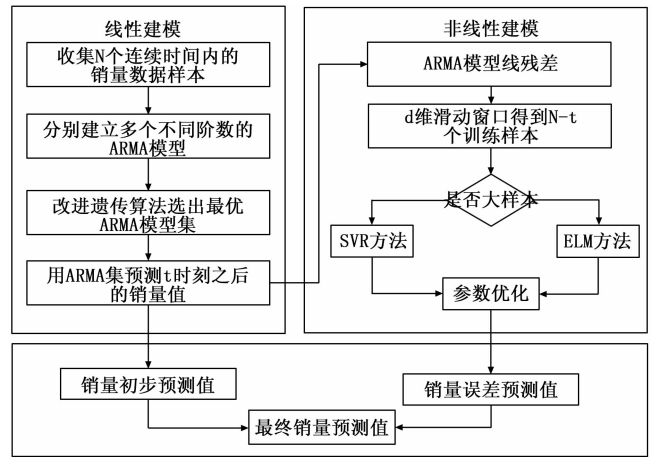


图 2 组合方法建模流程

## 2 非线性误差补偿模型

### 2.1 ARMA 模型

ARMA 模型是迄今最常用的拟合平稳时间序列的线性回归模型, 它由自回归模型 AR ( $p$ ) 和移动平均模型 MA ( $q$ ) 组合而成, 分别反映系统对历史数据和随机噪音的依赖关系。ARMA ( $p, q$ ) 的数学模型一般表示为:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\epsilon_t \\ E(\epsilon_t) = 0, \text{Var}(\epsilon_t) = \sigma_\epsilon^2, E(\epsilon_t \epsilon_s) = 0, s \neq t, \\ E(x_t \epsilon_s) = 0, \forall s < t \end{cases} \quad (1)$$

其中:  $\nabla^d x_t = (1 - B)^d$  是差分运算, 用于把数据处理成平稳时序, 序列  $\{\epsilon_t\}$  是白噪声序列,  $B$  为滞后算子,  $B^n x_t = x_{t-n}$ 。ARMA 模型经过多个参数的试算, 以及反复模型评定, 最终得到最优阶数  $p, q$ 。

### 2.2 选择性集成 ARMA 模型

ARMA 模型应用的关键是确定模型的回自阶数  $p$  和移动平均阶数  $q$ , 而往往在实际应用时无法准确定阶, 造成模型的准确率下降。本文提出了一个基于模型编码的选择性集成 ARMA 模型定阶的新方法, 该方法不直接对 ARMA 模型的阶数求解, 而是通过集成多个 ARMA 模型作为基学习器, 对所有基学习器二进制编码, 然后利用遗传优化算法<sup>[16]</sup> 选择出最优学习器组合。选择性集成方法通过选择部分精度较高、多样性大的个体学习器, 能获得比采用全部学习器更好的性能, 周志华等人<sup>[17]</sup> 最先从理论上证明了选择性集成的有效性。

选择性集成 ARMA 模型过程如下:

1) 设定  $p, q$  初始解空间: 阶数  $p, q$  的解空间表示最优阶数可能存在的范围, 假设  $p$  的解空间为  $(0, n)$ ,  $q$  的解空间是  $(0, m)$ , 则一共有  $(n+1) * (m+1)$  个 ARMA ( $p, q$ ) 模型参与集成。

2) 集成编码方法: 采用二进制模型权重编码方法, 即对每个模型的权重进行二进制编码, 权重为 1 表示选中该 ARMA 模型参与集成, 权重为 0 表示未选择该 ARMA 模型。

3) 模型预剪枝: 如果  $n$  与  $m$  的值过大, 遗传编码长度过长, 变异的多样性会受到抑制不利于产生最优解。因此, 在用遗传算法寻优之前, 需要预先剔除模型无法收敛的  $p, q$  组合, 以提高算法获取最优解的能力。

4) 遗传算法寻优过程: 遗传算法解的构成为:

$$W(t) = \{\omega_1(t), \omega_2(t), \dots, \omega_D(t)\} \quad (2)$$

表示初始种群中某个染色体在  $t$  次迭代后的基因序列,  $D$  表示参与集成的 ARMA 个数,  $\omega_j$  表示每个 ARMA 模型权重, 初始种群权重的设置如下:

$$\omega_j(0) = \begin{cases} 1, \text{rand}(0,1) \geq 0.5 \\ 0, \text{rand}(0,1) < 0.5 \end{cases} \quad (3)$$

其中:  $\text{rand}()$  是随机函数。为了选出泛化能力强的 ARMA 模型组合, 适应度函数选择均方根误差函数的倒数。均方根误差越小, 适应度越大, 被遗传的概率也就越大, 这与遗传算法的规律相符。故对于一个特定的 ARMA 集合  $i$ , 适应度函数可描述为:

$$\text{fitness}_i = \frac{1}{\sum_{j=1}^D \omega_j C_j} = \frac{1}{\sum_{j=1}^D \omega_j \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i^j - y_i)^2}} \quad (4)$$

式中,  $m$  为验证样本个数;  $\hat{y}_i^j$  为第  $i$  个样本在第  $j$  个模型下的预测值;  $y_i$  为第  $i$  个样本的真实值。

为了产生验证集  $L_{\text{val}}$  用于集成选择, 采用滑动窗口法对时序数据进行标准预处理, 每次滑动一个窗口以产生一个验证样本。ARMA 模型的输入输出映射函数为:

$$f: (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}) \rightarrow \hat{x}_{i+d\tau} \quad (i = 1, 2, \dots, d) \quad (5)$$

其中:  $\tau$  为延迟时间, 论文设为 1;  $d$  是滑动窗口的维数。

### 2.3 组合模型预测过程

组合模型 ARMA-SVR、ARMA-ELM 预测过程如下。

步骤一: 数据准备与 ARMA 建模。将商品的销量数据以周为周期分布构建时间序列, 设定阶数  $p, q$  解空间大小, 根据解空间构建多个不同阶数的 ARMA 模型进行集成, 然后利用遗传优化算法选出最优 ARMA 模型集。用该 ARMA 模型集对商品未来销量进行预测, 真实值和预测值在  $t$  时刻的残差为:

$$e_t = y_t - \hat{y}_t \quad (9)$$

步骤二: 参数寻优与残差预测。利用  $L_{\text{val}}$  集合分别训练 SVR 和 ELM。SVR 模型中的惩罚参数  $C$  和不敏感损失函数通过萤火虫算法寻优得到, ELM 的激活函数选择 sigmoid 函数。用训练好的非线性模型对残差  $e_t$  进行预测, 得到预测值  $\hat{e}_t$ 。

步骤三: 合并 ARMA 模型的初步预测值  $\hat{y}_t$  和残差预测值  $\hat{e}_t$ , 最终预测结果为:

$$y_t^* = \hat{y}_t + \hat{e}_t \quad (10)$$

## 3 实例预测分析

收集某电商平台 2011 年 1 月 1 日到 2012 年 12 月 28 日期间某品牌手机销量数据、饼干销量数据作为实例研究对象。通过分析商品的货道设计等实际情况, 商品的备货通常不宜过于频繁, 因此本文旨在预测未来一周内的商品总销量。将数据以周为单位进行预处理后, 得到商品销量曲线图, 如图 3 所示。

### 3.1 选择性集成模型预测结果

利用 Eviews 工具对两个序列进行单位根检验以确定序列是否平稳, 检验结果见表 1。通过检验结果可知, 手机销量序列数据具有很好的平稳性, 饼干销量序列数据非平稳。

将非平稳序列进行差分运算直至平稳后, 再根据 AIC 定阶准则对序列反复试算, 最后得出手机序列最优模型为 ARMA(1, 1), 饼干序列最优模型为 ARMA(0, 1) (此处统称

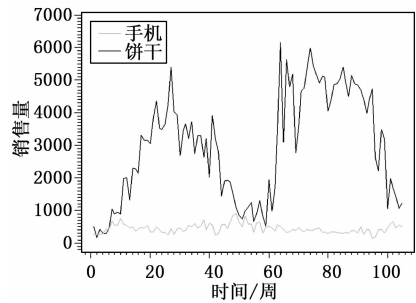


图 3 商品销量曲线图

表 1 序列单位根检验

商品	显著性水平	检验临界值	$t$ 统计量	概率值 $p$
手机		↓	-4.184779	0.0003
	1% level	-3.495677		
	5% level	-2.890037		
	10% level	-2.582041		
饼干		↓	-2.189676	0.2113
	1% level	-3.495021		
	5% level	-2.889753		
	10% level	-2.581890		

为 ARMA-1)。分别采用 ARMA-1、选择性集成 ARMA 模型 (简称为 ARMA-2) 对两组序列的后 45 周数据进行预测, 得到残差序列, 如图 4 所示。

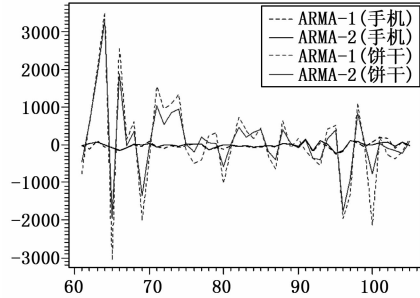


图 4 两类商品的残差序列

根据图 4 可知, ARMA-2 降低了 ARMA 模型的平稳性假定条件, 简化了模型的定阶过程。选择性集成方法通过集成了部分差异性大的 ARMA 模型作为组合学习器, 在平稳和非平稳序列上的预测结果都更接近真实值。

### 3.2 组合模型预测结果

ARMA-1 和 ARMA-2 在某些节点均产生了较大的非线性误差, 可能是受到一些突发事件 (如商家促销活动) 的影响, 因此对残差序列  $e_t$  建立非线性回归模型是必要的。论文分别用 SVR 方法、ELM 方法对残差序列  $e_t$  进行预测, 采用萤火虫算法对 SVR 的参数寻优, 最后得到交叉验证下的参数  $C = 13.6, \epsilon = 0.012$ 。将残差预测结果与 ARMA-2 预测值合并为序列  $y_t^*$ , 进一步得到 ARMA-SVR、ARMA-ELM 的预测值, 不同方法下的预测结果见图 5, 预测精度对比见表 2。

论文分别采用 MAE (Mean Absolute Error, 平均绝对误差)、RMSE (Root Mean Squares Error, 均方根误差) 和 MAPE (Mean Absolute Percentage Error, 平均绝对百分误差)

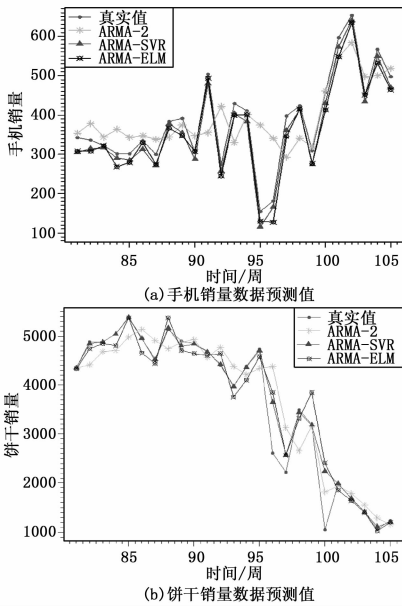


图 5 三类模型的销量预测值

表 2 模型预测效果比较

性能指标	模型	手机(平稳)	饼干(非平稳)
MAE	ARMA-1	87.28349222	537.3351136
	ARMA-2	62.53684838	354.5720451
	ARMA-SVR	22.45632044	121.2818484
	ARMA-ELM	25.62867939	259.1897546
RMSE	ARMA-1	106.6000241	758.6079273
	ARMA-2	83.17734401	517.6454307
	ARMA-SVR	24.50485841	323.8438132
	ARMA-ELM	30.11284418	424.7625384
MAPE / %	ARMA-1	27.9778790	22.9326688
	ARMA-2	21.3795996	17.4650316
	ARMA-SVR	6.6567527	7.3815995
	ARMA-ELM	7.5953922	11.4591856

三个性能指标来度量每个模型的预测误差。由预测结果可以看出, ARMA-2 的 MAPE 在平稳时序上比 ARMA-1 模型提高了 23.58%。在非平稳时序上提高了 41.28%。组合模型中 ARMA-SVR 的预测精度分别在 ARMA-2 的基础上提高了 68.86% (平稳) 和 45.18% (非平稳); ARMA-ELM 的预测精度分别提高了 64.47% (平稳) 和 14.90% (非平稳), 这也印证了 ARMA 模型不能识别销量数据中非线性影响因素的假设。对比两种组合方法, ARMA-SVR 在小样本、非平稳时序下预测准确度达到 92.62%, 对比 ARMA-ELM 精度更高、稳定性更好。

#### 4 结束语

短期销量预测对现代零售企业采购和库存具有很大战略指导价值。通过分析零售商品销售趋势, 企业能更好地把握消费者的未来消费走向, 有利于企业制定有效的市场营销策略。论文根据当前 ARMA 模型定阶准则和方法的不足, 提出了选择性集成的定阶方法, 将定阶过程自动化, 为线性平稳时序数据的分析提供了一种新思路。针对 ARMA 模型只能捕捉商品销

量中的线性及平稳因素的特点, 论文提出了将 ARMA 模型与智能非线性回归模型组合的预测方法, 引入非线性预测方法中表现较好的 SVR 方法和 ELM 方法对残差当中的非线性、非平稳因素进行预测并补偿。实验证明, 选择性集成 ARMA 方法定阶更准确, 有效改善 ARMA 模型在非平稳时序上的预测效果, 组合模型得到的销量预测结果相比仅采用线性平稳时序模型更符合实际, 稳定性更好。商品销量受外界因素的影响较大, 收集完整影响因素数据存在一定困难, 论文仅对销量历史数据进行了分析, 继续研究中需进一步整理较完整的销售数据, 不断完善预测模型, 以期达到预期的效果。

#### 参考文献:

- [1] 何伟, 徐福缘. 需求依赖库存且短缺量部分拖后的促销商品库存模型 [J]. 计算机应用, 2013, 33 (10): 2950-2953.
- [2] Kanko Y, Yada K. A Deep Learning Approach for the Prediction of Retail Store Sales [A]. IEEE 16th International Conference on Data Mining Workshops (ICDMW) [C]. Barcelona, 2016, 531-537.
- [3] 何斌, 蒙清. 灰色预测模型扩广方法研究 [J]. 系统工程理论与实践, 2002 (9): 137-140.
- [4] Schoof JT, Pryor SC. On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States [J]. Journal of applied meteorology and climatology, 2008, 47 (9): 2477-2786.
- [5] Engle R F. Autoregressive conditional heteroskedasticity with estimate of the variance of United Kingdom inflation [J]. Econometrica, 1982 (50): 987-1007.
- [6] 樊重俊. 贝叶斯向量自回归分析方法及其应用 [J]. 数理统计与管理, 2010, 29 (6): 1060-1066.
- [7] Box G E P, Jenkins G N, Reinsel G C 著. 时间序列分析预测与控制 [M]. 顾岚主译. 中国统计出版社, 1997: 51-58.
- [8] Esling P, Agon C. Time-series Data Mining [J]. ACM Computational Survey, 2012, 45 (1): 1-34.
- [9] Chang C C, Lin C J. Training v-Support Vector Regression: Theory and Algorithms [J]. Journal of Neural Computation, 2002, 14 (8): 1959-1977.
- [10] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70 (1-3): 489-501.
- [11] Qiao Yan. Application of integrated ARIMA and RBF network for groundwater level forecasting [J]. Forecasting Environmental Earth Sciences [J]. 2016, 75 (5): 1-13.
- [12] Babu CN, Reddy BE. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data [J]. Appl. Soft Comput, 2014, 23: 27-38.
- [13] Bang Zhuzhu, Wei Yiming. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology [J]. Omega, 2013, 41 (3): 517-524.
- [14] Wang P, Liu Y, Qinz ET. A novel hybrid forecasting model for PM<sub>10</sub> and SO<sub>2</sub> daily concentrations [J]. Total Environ Sci, 2015, 505, 1202-1212.
- [15] 何周舟. 基于时序数据的结构学习与模式预测联合优化算法研究 [D]. 杭州: 浙江大学, 2016.
- [16] Piotr Faliszewski, Jakub Sawicki, Robert Schaefer, et al. Multiwinner Voting in Genetic Algorithms [J]. IEEE Intelligent Systems, 2017, 32 (1): 40-48.
- [17] Zhou ZH, Wu jx, Tang W. Emsembling neural networks: Many could be better than all [J]. Artificial Intelligence, 2002, 137 (1-2): 239-263.