

分类号：\_Q29\_\_\_\_\_

学 号：313070104005

学校代码：10638\_\_\_\_\_

密 级：\_\_\_\_\_

# 硕士学位论文

## SVR 模型及其用于经济数据预测的研究

论文作者 指导

教师 学科专业 袁秀芳

申请学科门类 郑伯川教授

所在学院 论文 应用数学

提交日期 论文 理学硕士

答辩日期 数学与信息学院

二〇一六年四月

二〇一六年五月

---

西华师范大学学位评定委员会

四川·南充

二〇一六年五月

**SVR Model and Its Predict for Economic  
Data**

**A Dissertation**

**Submitted to the Graduate Faculty**

**In Partial Fulfillment of the Requirement**

**For the Degree of Master of Natural Science**

**By**

**Yuan Xiufang**

**Supervised by**

**Professor Zheng Bochuan**

**Major in**

**Applied Mathematics**

**In**

**Department of Mathematics and Information**

**China West Normal University**

**Nanchong, Sichuan Province, China**

May, 2016

## 目录

摘要 .....	I
Abstract .....	II
第 1 章绪论 .....	1
1.1 课题的背景、目的和意义 .....	1
1.2 支持向量机国内外研究综述.....	1
1.3 支持向量回归机理论 .....	2
1.4 论文的主要成果和内容 .....	5
第 2 章基于 SVR 的多变量电力消费预测 .....	6
2.1 引言 .....	6
2.2 样本数据选取与来源 .....	6
2.3 数据归一化处理 .....	6
2.4 电力预测实验步骤 .....	7
2.5 电力预测实验结果 .....	7
第 3 章基于 SVR 模型的房价预测 .....	9
3.1 引言 .....	9
3.2 数据的选取 .....	9
3.3 房价预测实验结果 .....	10
第 4 章 基于组合模型和多尺度核函数模型的股票价格预测 .....	12
4.1 引言 .....	12
4.2 小波分析 .....	13
4.3 多尺度核函数 .....	17
4.4 ARIMA-SVR 模型的实验 .....	18
4.5 多尺度核函数模型的实验 .....	21
第 5 章总结与展望.....	23

5.1 本文总结 .....	23
5.2 今后展望 .....	23
参考文献 .....	24
致谢 .....	i
关于学位论文使用授权的声明 .....	ii
关于学位论文原仓 11 性的声明 .....	iii
在学期间的科研情况 .....	iv

## 摘要

支持向量机(Support Vector Machines, SVM)是在统计学理论上发展起来的一种新的、非常有效的机器学习方法。它最早被 Vapnik 教授在 20 世纪 90 年代提出,近年不管是理论研究、算法实现还是应用方面都取得了突飞猛进的发展。特别是在应用方面,SVM 被广泛应用,解决了许多实际应用问题。SVM 分为支持向量回归机(Support Vector Regression, SVR)和支持向量分类机(Support Vector Classification, SVC),SVR 具有很好的预测能力,且其预测模型推广能力很强,即使在较长区间内预测仍然具有较高的精度。因此,被广泛用于各种实际问题的预测。但在经济数据预测方面,SVR 的应用还比较鲜见,因此本文将主要研究应用 SVR 对电力消费、商品房价格和股票价格这 3 种经济数据进行回归预测。本文主要在以下方面进行研究和探讨:

(1) 对电力消费建立基于 SVR 模型的多变量和双变量预测模型并比较之,得出前者比后者的预测更加准确。对商品房价格建立基于 SVR 模型和自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA),得出 SVR 模型优于 ARIMA 的结论。

(2) 针对股票市场高度非线性特点,单一预测模型很难描述出股票价格趋势的整体特征,提出了一种 ARIMA-SVR 的组合预测模型。首先采用二进正交小波的 Mallat 算法对股票收盘价数据进行小波分解和重构,得到了低频信息和高频信息;然后对高频信息建立 ARIMA 模型,以及对低频信息建立 SVR 模型,应用 ARIMA 模型和 SVR 模型进行训练和预测;最后对分析结果进行整合,得出组合模型的预测结果。实验表明,ARIMA-SVR 模型优于单一 SVR 模型。

(3) 将小波分析和核函数的知识结合起来,构造多尺度小波核作为 SVR 的核函数。首先通过 MoHet 小波和多尺度核函数构造出多尺度小波(MWK)的核函数表达式;然后采用核函数,利用 SVR 模型对股票收盘价进行了预测。实验表明,多尺度核函数模型优于单一 SVR 模型。

关键词:支持向量机;支持向量回归机;电力消费;商品房价格;股票价格;

ARIMA-SVR 模型;小波分析;多尺度核函数

## Abstract

Support Vector Machine is a new and highly efficient machine learning method developed from the basis of statistical theory. It was firstly proposed by the professor Vapnik in the 1990s. In recent years, the researches on SVM had made rapid development, including theoretical research, algorithm implementation research and application research. Particularly in the field of application research, SVM is widely used to solve many of the practical application problems. SVM is divided into Support Vector Regression and Support Vector Classification. SVR has good performance on data prediction, and the generalization ability of SVR model is strong, even in long range, the accuracy of predicted values is still high. Thus, SVR is widely used to predict data for a lot of practical problems. However, in the economic research field, the researches that adopted SVR to predict economic data is still lack, this thesis will apply SVR to predict three economic data, including electricity consumption prediction, commercial house price prediction and stock price prediction. This thesis mainly studied and discussed in the following areas:

(1) For electricity consumption prediction, the multi-variable SVR model and two-variable SVR model are both established and compared. The experiments show that multi-variable SVR model has higher prediction accuracy than two-variable SVR model. This thesis also compares SVR with Autoregressive Integrated Moving Average Model, it is concluded that SVR has better performance than ARIMA when they are both applied to predict electricity consumption.

(2) Because of the highly nonlinear character of stock price, a single prediction model is difficult to describe the overall character of stock price. So this thesis proposes a combination prediction model of ARIMA-SVR. Firstly, mallat algorithm of binary orthogonal wavelet is used to wavelet decompose and reconstruct stock closing price, the low frequency and high frequency information has have been obtained. Then the ARIMA model of the high frequency information and the SVR mode of the low-frequency information are established to train and predict data, respectively. Finally, the prediction results of combination model are got by integrating the results of two models. The experimental results show that ARIMA-SVR model is better than the single SVR model.

(3) Multi-scale wavelet kernel is constructed as SVR<sub>kernel</sub> function by combining wavelet analysis knowledge with kernel function knowledge. Firstly, MWK kernel function expression is constructed through Morlet wavelet and

multi-scale kernel function; then adopting MWK kernel function, SVR model is used to predict the stock closing price. The experiments show that the SVR model adopting multi-scale kernel function has better performance than the single SVR model.

**Key Words:** Support vector machine; Support vector regression; Electricity consumption; Real estate prices; Stock price; ARIMA-SVR model; Wavelet analysis; Multi-scale kernel

## 第 1 章绪论

### 1.1 课题的背景、目的和意义

支持向量机(Support Vector Machines, SVM)<sup>[1-2]</sup>是由 Vapnik 于 1995 年提出, 包括支持向量回归机(Support Vector Regression, SVR)和支持向量分类机(Support Vector Classification, SVC),可以用来解决小样本、非线性、高维数、局部极小点等实际问题。尤其是在解决小样本和非线性问题方面的有效性使得它在预测领域中具有其他方法无法比拟的优势。它的最大特点是不仅改进了传统的经验风险最小化原则, 而且建立在结构风险最小化原则上, 因此具有很好的泛化能力。SVR 具有很好的预测能力, 且其预测模型推广能力很强, 即使在较长区间内预测仍然具有较高的精度, 因此被广泛用于各种实际问题的预测。本文拟应用 SVR 模型研究分析经济数据, 以电力消费、商品房价格和股票价格为例, 建立 SVR 模型进行预测, 给予经济方面的参考建议。

电力作为主要能源消费品种, 以其清洁高效的特质在现代经济社会发展进程中发挥着重要作用, 因此研究电力消费势在必行。不论是在工业应用中, 还是在经济运行中, 它都具有举足轻重的地位。

居者有其屋, 是中国人千百年来梦想。在市场经济条件下, 具有经营资格的房地产开发公司通过出让方式取得土地使用权后经营销售的住宅就是商品房。进入 21 世纪以后, 商品房价格不断上涨, 房价已经成为关乎国计民生的问题, 对商品房价格的研究也成为了学术界的一个热点。

股票市场在整个国家的经济体系中占有举足轻重的作用, 它在很大程度上反映了一个国家的经济实力, 是市场经济的重要组成部分, 是经济发展趋势的晴雨表(3)。股票市场在西方已经有 300 多年的发展历史, 而中国的股票市场形成于 20 世纪 90 年代初, 经过 20 多年的发展, 已经初具规模。随着经济社会的不断发展, 人们的收入越来越高, 伴随着理财观念也深入人心。但是股票价格的时间序列数据具有非线性、非平稳性、高噪音等特点(4)。因此研究股票价格的走势非常有难度。

无论是电力消费还是房价或者股票都是关系国计民生的大事, 是经济领域研究的热点, 但对这些数据进行分析预测都极具挑战性。从宏观角度而言, 分析并预测经济数据的波动行为, 有利于整体把握整个市场的整体运行特征, 为政府制定宏观经济政策提供决策依据; 从微观角度而言, 能够准确分析并预测数据的规律, 将为市场投资者或决策者的行为提供有效工具。

### 1.2 支持向量机国内外研究综述

SVM 以 VC 维统计学理论和结构风险最小原则为基础, 有效解决了小样本、非线性和高维模式识别中存在的传统困难。它的研究主要有三个方面: 基础理论、训练算法和应用。SVM 且具有良好的泛化能力, 已经在模式识别、概率密度函数估计、回归估计等领域取得了较成功的应用。



国外研究方面，Vapnik[1]系统阐述了分类和回归问题的统计学习理论以及 SVM 的概念和分类问题；Mihler 等[5]应用 SVR 进行时间序列建模的研究，拓宽了 SVM 的领域；Smola[6]详细研究了 SVM 算法中各种核函数的机理和应用，对进一步完善其算法作出了贡献；Schölkopf 等[7]提出了将核函数变换和技巧用于空间变换的新思想；Cristianini<sup>[8]</sup>详细描述 SVM 的概率解释，对寻找一个合适的核函数具有一定的指导意义；Fumitala 等[9]基于决策树的多分类 SVM，有效解决了多分类问题；Cellio 等[10]通过分解和重组提出了多分类 SVR。另外，由于 SVM 具有很好的分类性能，因此国外对 SVM 的应用研究也较多。最著名的是贝尔实验室对美国邮政手写数字库识别的实验该实验表明 SVM 比其他算法的错误率低得多，验证了 SVM 的优越性能。另外 SVM 的主要应用研究包括：Osuna[11]利用 SVM 检测人脸；Joachims<sup>[12]</sup>利用 SVM 对文本进行分类；Drezet<sup>[14]</sup>应用 SVM 进行线性和非线性的系统辨识；Suykens 等应用 LSSVM 于非线性系统的最优控制。

国内研究方面，国内对 SVM 的研究主要集中在应用方面。文献[16]综述了统计学习理论和 SVM 的知识，促进其在我国的发展。文献[17]介绍了 SVM 的基本原理和算法，并应用 SVM 实现自动化网页分类。随着 SVM 的研究在国际上逐渐成为热点，SVM 也受到国内研究者的青睐和重视，并在各个领域都得到了深入的发展。忻栋等[18]利用 SVM 识别语音；孙永宣等[19]利用 SVM 检测图像奇异性；Hwanjo 等[20]利用 SVM 对网页进行分类；刘永斌等[21]将 SVM 应用于内燃机故障诊断；王定成[22]将 SVM 应用于内膜控制；冯瑞等[23]使用 SVM 的软测量建模方法对催化裂化吸收稳定系统的汽油收率进行预测；陈念怡[24]综述了 SVM 在化学化工中的应用。

SVR 作为 SVM 的一个大的分支，目前已经成为预测数据的性能优越的工具之一。SVR 已经应用于多个领域的预测，如预测：金融时间序列[25-26]、顾客需求[27]、旅客入境人数[28]、交通流量[29]、风速的、降雨量[30]、电力市场价格[31]等。近些年来，应用的领域不断扩张，比如预测航班延误网、房价[32-38]、耕地面积[39]、粮食产量[40]、流行色[41]、水质[42]等。

除了直接应用 SVR 进行数据预测外，一般对 SVR 的应用方式进行的优化，主要优化途径包括以下三方面：一是参数的优化方法，比如 PSO、GA 等智能算法[29,43,44]；二是从核函数的角度出发，比如构造金融核[45]、多尺度小波核[46]、混合核[47]、时序核[48]、文本核[50]、图像核[51]等核函数；三是其它模型与 SVR 模型的组合[49-55]，既包含了本身的特点，又囊括了 SVR 模型的优点。

### 1.3 支持向量回归机理论

SVR 经过不断的发展和研究，目前有 SVR、 $\gamma$ -SVR、LS-SVR、二次 SVR 等，且都有线性和非线性两种模型，各有各的特点，但是使用最普遍且最常用的还是 SVR。鉴于此，本文主要使用的是 SVR，下面介绍关于它的理论和推导过程。

已知一个样本集  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，其中  $x_i$  表示特征， $y_i \in \mathbb{R}$  表示真实值。对于线性回归问题，令  $\hat{y} = a + bx$ ，如果根据样本估计出满足每对  $(x_i, y_i)$  变化关系的回归函数  $f(x)$ ，使得  $f(x_i)$  与  $y_i$  的差值都很小，那么就可以用  $f(x)$  来预测任意  $x$  所对应的  $y$  值。SVM 用于回归估计问题时称为 SVR [3]。因此 SVR 问题可以表示成公式

$$\min_{a,b} \sum_{i=1}^n |y_i - (a + bx_i)|$$

其中  $|y_i - (a + bx_i)|$  表示预测值和真实值的最大差距。

实际上，并不是所有的样本点都满足  $|y_i - (a + bx_i)| \leq \epsilon$ ，必定会有极少数样本点不满足条件，为了处理那些大于  $\epsilon$  的样本点，可以引入松弛因子  $\xi$ ，优化问题变成：

$$\begin{aligned} \min_{a,b,\xi} & \sum_{i=1}^n |y_i - (a + bx_i)| + c \sum_{i=1}^n \xi_i \\ \text{s.t.} & |y_i - (a + bx_i)| - \xi_i \leq \epsilon \\ & \xi_i \geq 0 \end{aligned} \quad (1.3.2)$$

其中  $C > 0$  表示对大于  $\epsilon$  的样本点的惩罚程度。公式(1.3.2)是一个二次规划问题，通常不直接求解，而是求解它的 Lagrange 对偶问题。公式(1.3.2)的 Lagrange 式子可表示为

$$\begin{aligned} \max_{\alpha, \beta, \gamma} & \sum_{i=1}^n \alpha_i (y_i - a - bx_i - \epsilon) - \sum_{i=1}^n \beta_i (y_i - a - bx_i - \epsilon + \xi_i) \\ & - \sum_{i=1}^n \gamma_i (\xi_i) \end{aligned} \quad (1.3.3)$$

因此求解公式(1.3.2)可以转换成求解

$$\begin{aligned} \min_{\alpha, \beta, \gamma} & L \\ \text{s.t.} & \alpha_i \geq 0, \beta_i \geq 0, \gamma_i \geq 0 \end{aligned} \quad (1.3.4)$$

由于公式(1.3.1)是凸优化函数，只要优化求解方法能够满足 KKT 条件，求解优化问题(1.3.4)就可以通过其对偶优化问题求解，对偶优化问题可表示为

$$\begin{aligned} \max_{\alpha, \beta, \gamma} & L \\ \text{s.t.} & \alpha_i \geq 0, \beta_i \geq 0, \gamma_i \geq 0 \end{aligned} \quad (1.3.5)$$

令  $W(a, \alpha) = \min_{\alpha} W(a, T)$ ，可以通过求  $\alpha$  偏导数为零得  $W(a, T)$ 。

$$\frac{\partial W}{\partial \alpha} = 0 \quad (1.3.6)$$

$$\frac{\partial W}{\partial \alpha} = 0 \quad (1.3.7)$$

$$\frac{\partial W}{\partial \alpha} = 0 \quad (1.3.8)$$

将(1.3.6), (1.3.7)和(1.3.8)代入到(1.3.5)中, 得

$$\begin{aligned} & \min_{\alpha} W(a, \alpha) = \min_{\alpha} \left[ \sum_{i=1}^m \left( \frac{1}{2} \left( \frac{y_i - \alpha}{\sigma} \right)^2 + \frac{1}{2} \left( \frac{y_i + \alpha}{\sigma} \right)^2 \right) \right] \\ & \text{s.t. } \alpha \in [0, C] \end{aligned} \quad (1.3.9)$$

求解公式 (1.3.9) 可以利用序列最小优化算法 (Sequential Minimal Optimization, SMO), 得到  $\alpha$  和  $\sigma$  的值后, 根据公式 (1.3.7) 可以得到  $\beta$  的值, 根据

$$\begin{aligned} & \frac{\partial W}{\partial \alpha} = 0 \\ & \frac{\partial W}{\partial \sigma} = 0 \end{aligned} \quad (1.3.10)$$

可以计算  $\beta$  的值。其中表示使  $\| (x_i) - z \| = 0$  的支持向量的个数,  $SV$  是使  $\| (x_i) - z \| = 0$  的支持向量的集合。

最终得到的线性回归函数为

$$f(x) = \sum_{i=1}^m \alpha_i (x_i - z) + \beta \quad (1.3.11)$$

对于非线性回归, 其基本思想是通过事先确定的非线性映射将输入向量映射到一个高维特征空间中, 然后在此高维空间中构造最优超平面。这样, 在高维特征空间的线性回归就对应于低维输入空间的非线性回归。其具体实现是通过核函数来实现的, 从而避免了在高维空间中计算复杂的点积运算问题。SVR 求解非线性回归的对偶优化问题可表示为:

$$\begin{aligned} & \max_{\alpha, \beta} W(a, \alpha) = \min_{\alpha, \beta} \left[ \sum_{i=1}^m \left( \frac{1}{2} \left( \frac{y_i - \alpha}{\sigma} \right)^2 + \frac{1}{2} \left( \frac{y_i + \alpha}{\sigma} \right)^2 \right) \right] \\ & \text{s.t. } \alpha \in [0, C] \end{aligned} \quad (1.3.12)$$

得到的非线性回归函数为:

$$f(x) = \sum_{i=1}^m \alpha_i k(x_i, x) - \beta \quad (1.3.13)$$

其中  $k(x, x')$  是核函数，满足 Mercer 条件的任意对称函数。较常见的四种核函数 包括线性核、多项式核、径向基核(Radial Basis Function, RBF)和 sigmoid 核，其 中应用最广泛的核函数是 RBF 核，表达式为

$$k(x, x') = \exp(-1/2a^2 \|x - x'\|^2) \quad (1.3.14)$$

它以非线性方式将输入空间映射到特征空间，用于处理现实中以这种方式存 在的问题。与其它核相比较，在数值计算方面，RBF 核的困难大大减弱，因此本 文选取 RBF 核作为 SVR 模型的核函数。

#### 1.4 论文的主要成果和内容

本文的主要成果包括：

(1) 对电力、商品房价格和股票收盘价这 3 种经济数据分别进行 SVR 建模，得到了较好的预测效果。

(2) 针对各种预测方法的优点，提出基于小波分析的组合预测模型 ARIMA-SVR，比其他单一模型在股市预测更准确、更可靠，是一种有效的股票价格时间序列模型。

(3) 针对核函数的选择和构造方法，提出了多尺度小波核函数。通过与其他模型在股票收盘价预测效果性能的比较，证明了它是一种非常有前途的模型。

本文的主要内容如下：

第 1 章首先介绍本文的研究背景，目的，意义；其次介绍 SVM 的国内外研究 现状；然后介绍 SVR 的理论模型；最后介绍本文的主要成果和内容。

第 2 章介绍电力预测的方法，并对电力消费建立 SVR 模型的多变量和双变量 预测，并比较其预测效果。

第 3 章介绍了房价预测的方法，ARIMA 模型的原理，并对房价数据分别建立 SVR 模型和 ARIMA 模型，得出预测结果，并比较之。

第 4 章分别应用 ARIMA-SVR 组合模型和多尺度核函数的 SVR 模型在股票价格预测中进行研究。首先介绍了股票现有的研究方法，然后介绍了小波分析和多 尺度核函数的相关理论，最后比较 ARIMA 模型、F-SVR 模型、ARIMA-SVR 与 多尺度核函数模型的实验效果。

第 5 章是总结与展望。验证了 SVR 模型在经济数据中的良好预测效果，并指 出了其不足以及发展方向。

## 第 2 章 基于 SVR 的多变量电力消费预测

### 2.1 引言

电力作为主要能源消费品种，以其清洁高效的特质在现代经济社会发展进程 中发挥着重要作用，因此研究电力消费势在必行。不论是在工业应用中，还是在 经济运行中，它都具有举足轻重的地位。目前对电力消费进行预测所采用的方法 主要有：回归分析法、弹性系数法、灰色预测法【60】、神经网络方法等。而在对电力消费中使用 SVR

模型的则寥寥无几。本章对电力消费 建立 SVR 模型的多变量和双变量预测模型并比较其预测效果。

## 2.2 样本数据选取与来源

随着经济的持续增长,电力消费也保持着稳定增长的趋势。影响电力消费的因素诸多,包括国民经济的发展情况、产业经济结构、重工业比重、人均 GDP、工业产值、电能效率、消费习惯等。在本章的电力消费模型中,电力消费量(EC)、人均国内生产总值(PCGDP)、重工业比重(HIS)以及电能效率(EI)构成了模型多个变量。本章的数据中,EC、PCGDP和HIS都来自于2000-2012年《中国统计年鉴》与《中国能源统计年鉴》。电能效率变量则是用工业增加值和工业电力消费量计算所得。

## 2.3 数据归一化处理

以2000-2012年的EC、PCGDP、HIS和EI为SVR预测模型的输入变量,以EC作为目标向量建立SVR的电力需求预测模型。为消除各时间序列之间不同量纲的影响,采用公式(2.3.1)对数据做归一化处理,使得每一时间序列向量在区间〔0,1〕之间,归一化结果见表2·1。

凡 axfn

其中,  $x$  为原始数据;  $X$  标准化后的数据;  $x_{\min}$ 、 $x_{\max}$  分别为第  $j$  个输入变量数据序列)中的最小值和最大值。

表2·1 数据归一化一览表

年份	EC	PCGDP	HIS	EI
2000	0	0	0	0.855397
2001	0.034474	0.024967	0.016807	0.870773
2002	0.082476	0.050336	0.075630	1
2003	0.153188	0.087717	0.378151	0.955131
2004	0.234195	0.146328	0.546218	0.879916
2005	0.316006	0.206775	0.638655	0.729386
2006	0.416519	0.282402	0.840336	0.643065
2007	0.530154	0.402322	0.882353	0.478616
2008	0.580568	0.517945	0.882353	0.168055
2009	0.649203	0.580027	0.949580	0.213823
2010	0.784291	0.724055	0.882353	0.132744
2011	0.923898	0.893415	1	0.032610
2012	1	1	0.705882	0

## 2.4 电力预测实验步骤

采用以下实验步骤进行实验:

(1) 数据归一化处理;

(2) 选定电力消费量作为输出变量  $H$ , 其他的统计数据作为输入变量  $J = 1, 2, \dots, 13$ 。

在双变量预测中只选择PCGDP作为输入变量, 在多变量预测中则选择PCGDP、HIS和EI

构成的三维输入变量；

(3) 利用 LIBSVM 软件工具 `g` 的 `svmtrain` 函数，选择 `w-SVR` 算法和 `RBF` 核函数，对表 2·1 中选定的样本数据进行学习，得到回归模型。通过反复实验和 效果对比，发现参数在一定范围内取值时，不同参数下的实验结果相差很小。因此，实验表明，可以取多组不同参数都能得到相似的结果，文中实验结果所选用 参数值为： $g = 0.2$ ,  $c = 10$ ,  $\gamma = 0.005$ ,  $\epsilon = 0.001$ 。其中各项参数的意义为： $g$  表示多项式、`RBF` 和 `sigmoid` 核函数中的  $\gamma$  系数  $c$  表示 `w-SVR` 中的惩罚系数， $p$  为 `SVR` 的损失函数  $p$  的值， $Q$  为允许的终止判据；

(4) 根据得到的回归模型，利用 LIBSVM 的 `svmpredict` 函数对学习数据进行 预测，得到预测数据  $\hat{y} = 1, 2, \dots, 13$ ；

(5) 分析对比真实数据  $y$  和预测数据  $\hat{y} = 1, 2, \dots, 13$ 。

## 2.5 电力预测实验结果

分别采用实验步骤对表 2-1 的数据进行多变量和双变量实验，在 `SVR` 的参数设置相同的情况下得到的真实值与预测值的对比图，如图 2-1 和图 2-2 所示。

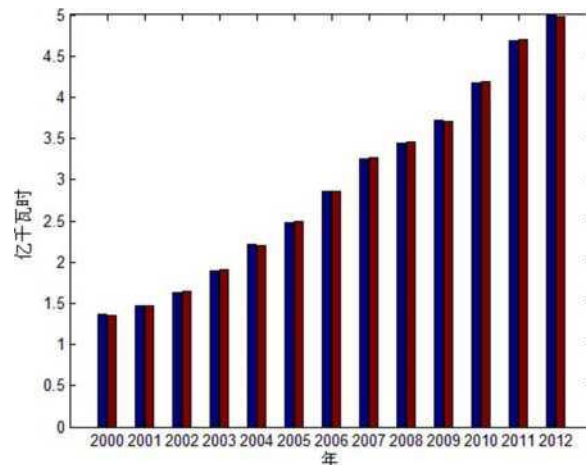


图 2·1 多变量的真实值与预测值对比

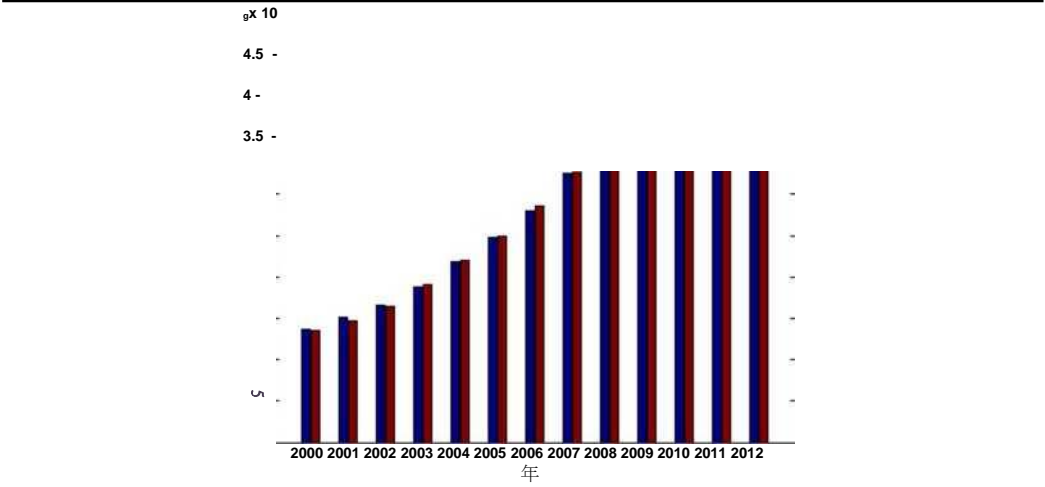


图 2·2 双变量的真实值与预测值对比

从图 2-1 和图 2-2 对比可以看出多变量预测值更接近真实值。计算两种方式下预测值与实际值之间的绝对百分比误差，其结果如表 2-2 所示。表 2·2 电力消费预测结果

年份	实际值	双变量 预测值	双预测 误差%	多变量 预测值	多预测 误差%
2000	13472.38	13615.19	1.06	13653.49	1.34
2001	14723.46	15148.00	2.89	14677.43	0.31
2002	16465.45	16650.15	1.12	16282.33	1.11
2003	19031.60	18764.29	1.41	18913.89	0.62
2004	21971.37	21850.14	0.55	22153.18	0.83
2005	24940.32	24756.87	0.74	24768.59	0.69
2006	28587.97	28034.24	1.94	28650.15	0.22
2007	32711.81	32531.98	0.55	32528.08	0.56
2008	34541.35	36244.43	4.93	34358.94	0.53
2009	37032.14	38062.20	2.78	37215.16	0.49
2010	41934.49	42035.55	0.24	41752.71	0.43
2011	47000.88	46726.80	0.58	46820.54	0.38
2012	49762.64	49941.08	0.36	49945.82	0.39

根据表 2-2,计算双变量 SVR 和多变量 SVR 模型的绝对百分比预测误差的 均值和方差，如表 2-3 所示。

表 2·3 预测结果的均值和方差		
模型	均值%	方差%
多变量 SVR 模型	0.6169	0.001
双变量 SVR 模型	1.4722	0.183

多变量 SVR 模型预测值的绝对百分比误差均值和方差都小于双变量 SVR 模型。由此可以得出影响研究对象的因素在数量上和选择的因变量上都对预测结果 产生了一定程度的影响，故在我国电力需求的预测中，多变量 SVR 模型比双变 量 SVR 模型的预测效果好。

第 3 章 基于 SVR 模型的房价预测

3.1 引言

居者有其屋，是中国人千百年来梦想。在市场经济条件下，具有经营资格的房地产开发公司通过出让方式取得土地使用权后经营的住宅就是商品房，所以它是80年代以后才在中国出现的。其价格组成包括成本、税金、利润、代收费用以及地段、层次、朝向、质量、材料差价等。进入21世纪以后，商品房价格不断上涨，房价已经成为关乎国计民生的问题，对商品房价格的研究也成为了学术界的一个热点。孔凡文等采用直线趋势法和一元线性回归预测上海市商品房价格，刘月婷等采用基于MATLAB建立多元线性回归模型预测兰州市商品住宅价格，武秀丽等采用时间序列预测房价，常振海等采用非参数自回归预测房价，陈森君采用BP神经网络预测房价，徐富强采用神经网络组合预测合肥市房价，陈博采用灰色GM(1,1)模型预测西安市商品房价格。本章对房价建立SVR模型并预测，并且与ARIMA模型的预测结果作比较。

### 3.2 数据的选取

1998年中国住房市场发生了变革，那么1998年就是商品房价格的一个分水岭，故从1998年的数据开始进行研究，根据供给和需求两个方面选取了8个影响上海市商品房价格的因素，通过1999-2014年的《上海市统计年鉴》(1)得到1998-2013年房地产价格相关数据，见表3-1。

表3-1 上海市1998-2013年商品房基本信息表

年份	$y$	$\&$	$\pi$	$\pi$	$\pi$	$\pi$	$K$	龙	$\xi$
1998	3758.91	3801.09	113.60	95.50	92.20	593.11	1913.55	50.60	320.66
1999	3616.07	4188.73	105.90	93.30	89.90	573.06	1950.76	45.20	324.49
2000	3565.45	4771.17	103.30	91.90	95.80	631.64	1909.11	44.50	408.82
2001	3993.30	5210.12	102.30	97.20	104.90	730.33	2434.73	43.40	439.17
2002	4284.81	5741.03	100.00	106.30	99.00	822.27	2596.95	39.40	567.76
2003	5118.41	6694.23	101.10	115.10	102.10	1195.80	3609.20	37.20	676.28
2004	6488.97	8072.83	101.60	120.30	105.53	1724.40	4672.53	36.40	900.67
2005	6842.00	9247.66	102.90	106.93	103.60	1889.25	5648.85	35.90	920.84
2006	7196.01	10572.24	102.90	101.20	104.00	2285.38	6506.41	35.60	835.65
2007	8360.98	12494.01	104.50	107.80	105.10	2524.18	6090.22	35.50	837.53
2008	8255.01	14069.87	102.50	107.90	104.64	3245.77	5723.90	36.60	843.63
2009	12839.98	15046.45	96.55	102.20	100.60	3830.53	5719.93	35.00	918.68
2010	14399.89	17165.98	103.47	118.90	104.40	4300.19	6217.15	33.50	1229.83
2011	14502.79	19195.69	105.20	117.00	106.70	4586.28	5984.74	35.50	1398.75
2012	14061.34	20181.72	102.80	104.50	105.80	4843.44	6476.07	36.80	1451.94
2013	16419.99	21602.12	102.30	110.40	105.20	5102.84	6374.25	34.90	1615.51

注：上海市生产总值(亿元) $Y$ ，上海市居民消费价格指数 $\&$ ，上海市土地交易价格指数 $\pi$ ，上海市房屋租赁价格指数上海市建筑业总产值(亿元) $X_5$ ，上海市房屋竣工面积(万平方米) $x_6$ ，上海市恩格尔系数 $X_7$ ，上海市房地产住宅投资总额(亿元) $X_8$ ，上海市商品房价格(元) $Y$ ，商品房销售额/商品房销售面积。

### 3.3 房价预测实验结果

以1998-2013年的 $Y$ 、 $\&$ 、 $\pi$ 、 $\pi$ 、 $\pi$ 、 $\pi$ 、 $\pi$ 和 $X_8$ 的时间序列为SVR预测模型的输入变量，以 $Y$ 的时间序列作为目标向量建立商品房价格预测模型。为消除各时间序列



之间不同量纲的影响,采用公式(2.3.1)对数据做归一化处理,使得每一时间序列向量在区间 $[0,1]$ 之间。与第三章的电力预测实验步骤相同,首先采用房价预测步骤对表3·1的数据进行实验,得到SVR模型的预测值和真实值的对比图和绝对百分比误差,结果见图3-1和表3·2。然后使用Eviews 7.2软件对表3·1中的Y建立ARIMA模型,得到Y的预测值和真实值的对比图和绝对百分比误差,结果见图3·1和表3·2。根据图3·1可知,两种模型的预测效果都比较好,但ARIMA模型的预测曲线波动比较大,SVR模型的预测曲线更接近真实值。

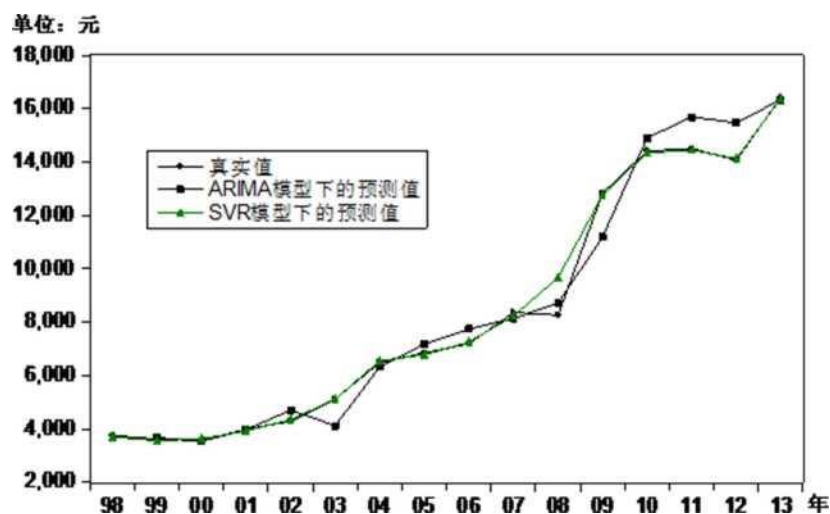


图3·1 商品房价格真实值以及两种模型下的预测值对比图

表3·2 两种模型下商品房价格的预测结果

年份	真实值	SVR 模型 预测值	ARIMA(1,2,6) 预测值	SVR 模型 预测值误 差%	ARIMA(1,2,6) 预测值误差%
1998	3758.91	3700.22	NA	1.56	NA
1999	3616.07	3548.70	3676.99	1.86	1.69
2000	3565.45	3630.62	3536.06	1.83	0.82
2001	3993.30	3925.54	3968.39	1.70	0.62
2002	4284.81	4345.60	4702.23	1.42	9.74
2003	5118.41	5112.80	411&59	0.11	19.53
2004	6488.97	6550.65	6315.28	0.10	2.68
2005	6842.00	6771.76	7188.52	1.03	5.06
2006	7196.01	7259.83	7748.11	0.89	7.67
2007	8360.98	8224.32	8140.46	1.63	2.64
2008	8255.01	9671.10	8709.04	17.15	5.50
2009	12839.98	12771.64	11198.56	0.53	12.78
2010	14399.89	14340.31	14899.97	0.41	3.47
2011	14502.79	14440.93	15679.20	0.43	8.11
2012	14061.34	14131.40	15466.30	0.50	9.99
2013	16419.99	16359.47	16331.38	0.37	0.54

注 NA 表示空值,由于进行了差分运 开头就会少一位。

根据表 3-2 可知 SVR 模型在 2008 年的商品房价格预测值与真实值相差较大，其他年份的预测值与真实值比较接近。根据查阅相关资料知造成 2008 年预测值偏差大有两个原因：一是由于 2008 年宏观经济形势恶化、房地产行业呈现衰退局面；二是 2007 年国家多次提出加息政策调控住房民生问题，在 2008 年得到了集中体现，导致房价出现空前下降。除此之外，比较两种模型的预测值绝对百分比误差，大致上可见 ARIMA 模型的预测值绝对百分比误差更大更波动一些，反之 SVR 模型的预测值绝对百分比误差更小更均匀，所以 SVR 模型的预测效果比较好的。为进一步对比 SVR 模型和 ARIMA 模型的预测效果，通过计算预测值绝对百分比误差的均值和标准差来衡量，如表 3-3 所示。

表 3-3 预测结果的均值和标准差

模型	均值%	标准差%
SVR 模型	1.97	3.97
ARIMA(1,2,6)模型	6.06	5.16

由表 3·3 可知 SVR 模型预测值的绝对百分比误差均值和标准差都小于 ARIMA 模型，说明 SVR 模型预测值更接近于真实值。因此得出基于 SVR 模型的拟合精度很高，能够较好的拟合出商品房房价的走势。该模型在处理类似小样本多影响因素的预测分析中呈现了很大的优势和参考价值，能够为我国政府部门、宏观经济工作者及房地产相关人员提供参考和指导。

## 第4章 基于组合模型和多尺度核函数模型的股票价格预测

### 4.1 引言

目前,随着人们对股市的认识不断加深,以及高级的计算机算法和技术的不断完善,在股票价格的走势研究方面,人们的热情愈发高涨,开始依赖各种预测方法。

股市是一个非线性系统,用线性模型<sup>[74]</sup>逼近容易丢失有用信息,所以线性模型不适合股票时间序列。为适应股票数据的特点,人们研究出自回归条件异方差模型(Auto Regressive Conditional Heteroskedasticity, ARCH)<sup>[75]</sup>、广义自回归条件异方差(Generalized Auto Regressive Conditional Heteroskedasticity, GARCH)<sup>[76]</sup>模型等。由于它们与实际情况更相符,从而得到了广泛的应用。但是它们也有不足之处,即基于平稳性假设、正态分布假设、线性假设等,模型的构建及参数的选择困难。当然,还有人另辟蹊径,利用状态转移法 SI 来预测股票,最典型的就 是隐马尔科夫模型(Hidden Markov Model, HMM) <sup>[78-79]</sup>。由于它 能对不可观测的信息状态及其转移过程和不可观测状态与可测关联变量的函数关系建模。但是,当高频数据的波动性大,预测效果则不显著比<sup>°</sup>。随着研究的不断深入,非线性、非参数的智能预测方法就应运而生了<sup>(8)1</sup>。

现在,应用较为广泛的智能预测方法包括人工神经网络(Artificial Neural Network, ANN)<sup>®^</sup>、模糊神经网络(Fuzzy Neural Network, FNN)<sup>^</sup>, 85] 以及 SVM。

神经网络具有很强的非线性拟合能力,可以映射任意复杂的非线性关系,且学习规则简单,便于实现,具有局部逼近的优点<sup>盼 7]</sup>。文献[81]的研究表明,ANN 方法存在过估计问题,可能产生局部而非全局最优解的缺点。

因此,如何进一步寻找既能改进缺陷,又能准确预测的方法便成为研究的重点和热点。随着粗糙集理论<sup>朗</sup>和 SVM 的引入,对股票价格研究的新思路、新技术和新方法也随之诞生了。Cao 和 Tay<sup>[94]</sup>在金融预测方面建立 SVM 模型,得出其预测效果比神经网络更出色,并对核函数的选取和参数的选取进行了详细的研究。Trafhllis 和 Ince<sup>[95-96]</sup>用 SVR 对股票价格进行了预测研究,并和多层感知机(Multilayer Perceptron, MLP)和 ARIMA 进行对比研究得出,SVR 的预测精度远高于 MLP 和 ARIMA,如果交易策略改变的话,MLP 和 SVR 会差不多。国内的研究者们<sup>弘心]</sup>也建立 SVM 在股票价格方面进行了研究。

这些预测方法都有其各自的优点,但是股票价格的复杂性、外部因素的多变性限制了其预测的准确性,现有的预测方法在实际应用过程中的预测效果并不理想。目前,人们建立组合模型来预测股票价格,有 SVM 与 ARIMA 混合构造预测模型<sup>[52]</sup>、二进正交小波变换和 ARIMA-SVM 模型<sup>同</sup>、CAR-SVM 模型<sup>网</sup>、小波变换和 ARMA · SVM<sup>a]</sup>、ARIMA 模型与神经网络组合预测<sup>MA]</sup>、小波包与神经网络

组合预测〔105〕、神经网络与 SVM 的组合预测〔106〕、改进 GM(1, N)和优化 SVM 组合模型 MI SVM-AdaBoost 模型何】等。

另一方面,在构造核函数方面对股票价格研究却是比较少的,而在构造合适的多尺度核函数来预测股票价格的研究更是少之又少。金得宝昭建立基于金融核的二阶段预测股市模型,与传统的 SVM 预测模型相比,无论在学习能力和泛化能力上看都具有明显的优越性。文献〔109-110〕提出了基于高斯调频小波核构造小波核函数的小波支持向量机(Wavelet SVM, WSVM),其逼近精度和模式识别率均优于高斯核函数 SVM。但是文献〔111〕指出目前提出的 WSVM 不一定最好的。基于此,任世锦和吴铁军购提出了一种基于径向基小波核的多尺度 WSVM 学习算法,可提高 WSVM 的训练速度和逼近精度。蒋波的针对文献〔46〕的不足,构造了多尺度小波核,并比较了多尺度小波核 v-SVM,单尺度小波核 v-SVM,小波神经网络和 RBF 核函数 SVR 对上证指数的预测效果。

基于这两点,本章采用的是两种不同角度的基于 SVR 组合模型的股票价格预测。其中一种是通过小波变换分解和重构出股票数据的高频信息和低频信息,然后对高频信息建立 ARIMA 模型并预测,对低频信息则用 SVR 模型进行拟合,最后将两种模型的预测结果进行叠加就得到预测值。也就是一种股票价格预测的 ARIMA-SVR 组合模型。另外一种核函数的构造,众所周知,SVR 模型最关键的就是核函数的选择,因此首先构造一种特殊的核函数——多尺度核函数,然后建立基于多尺度核函数的 SVR 模型对股票价格进行预测。通过比较两种不同的改进模型与 RBF 的 SVR 模型的预测效果,证明其在股市预测领域的有效性和优越性。本章的研究丰富和拓展了当前股票价格预测的研究内容,其结论能为股民有效投资理财提供科学的参考依据,并且有着广泛的应用前景。

## 4.2 小波分析

Mallat 利用多分辨分析的特征构造了快速小波变换算法,即 Mallat 算法(1)习。其算法的分解部分如下:

假若已知  $C_j$  和  $D_j$ , 贝  $U$

$$P_j + Q_j f = \sum_k C_k \phi_k + \sum_k D_k \psi_k \quad (4.21)$$

$$C_n^{j-1} = \langle \phi_{j-1n}, P_{j-1} f \rangle = \sum_k C_k^j \langle \phi_{j-1n}, \phi_{jk} \rangle + \sum_k D_k^j \langle \phi_{j-1n}, \psi_{jk} \rangle$$

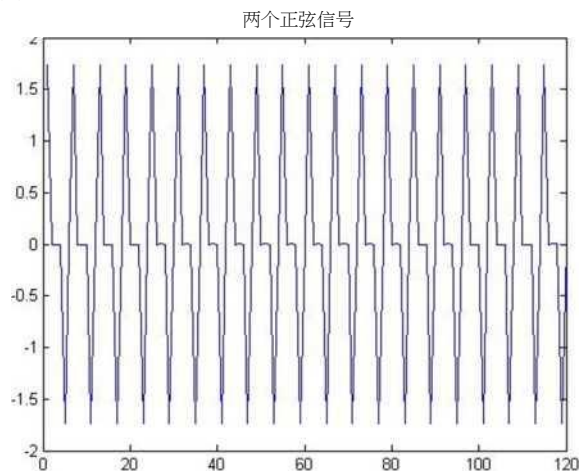
$$= \sum_k C_k^j \langle \phi_{j-1n}, \phi_{jk} \rangle \quad (4.2.2)$$

因而  
或者

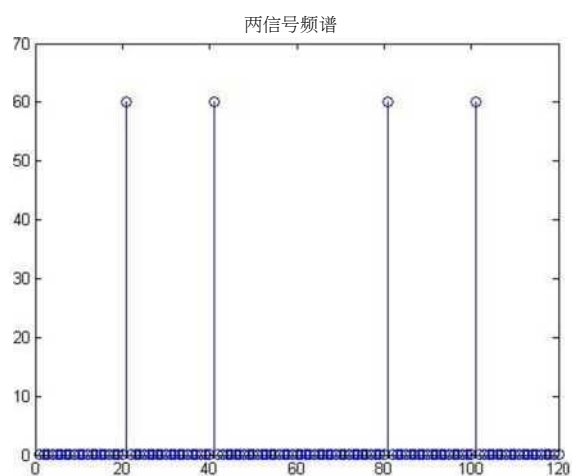
$$\sum_k D_k^j \langle \phi_{j-1n}, \psi_{jk} \rangle \quad (4.2.3)$$

重构算法也是一个树状算法,而且与分解算法用的是同样的滤波系数。以正

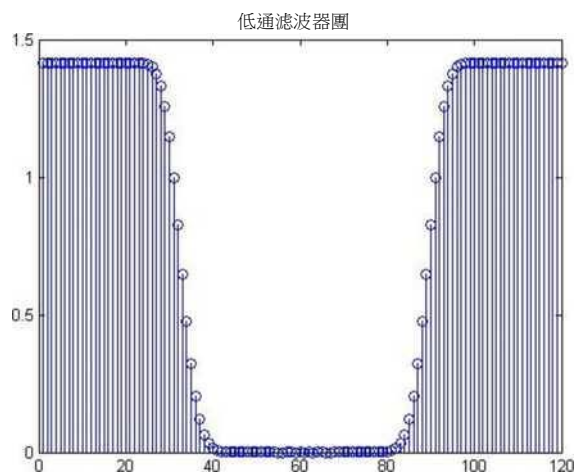
弦波为例，以便清晰理解 Mallat 算法。定义两个频率分别为 50 和 100 的正弦波，对其进行采样频率和点数，得到混合正弦波。应用 Mallat 算法对其进行小波分析，得到图 5·1,包括子图(a)混合正弦波示意图，子图(b)两信号频谱，子图(c)低高通 滤波器示意图，子图(d)高通滤波器示意图，子图(e)分解信号示意图，子图(f)分解 信号频谱示意图，子图(g)重构低频和高频信号示意图，子图(h)重构信号频谱示意图，子图(i)重构信号和原信号的比较示意图。



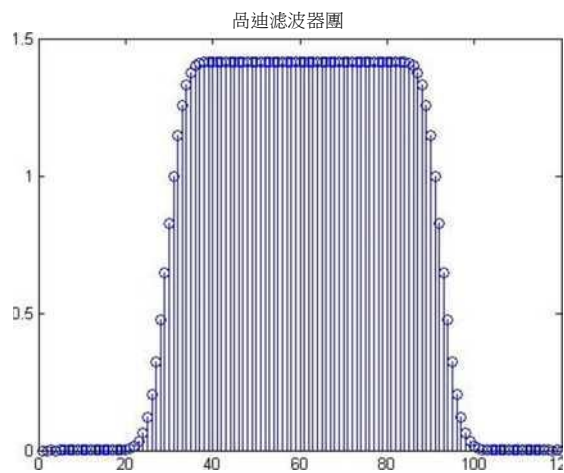
(a)



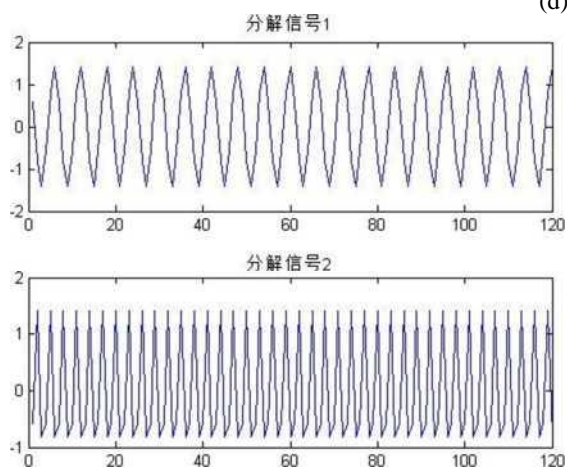
(b)

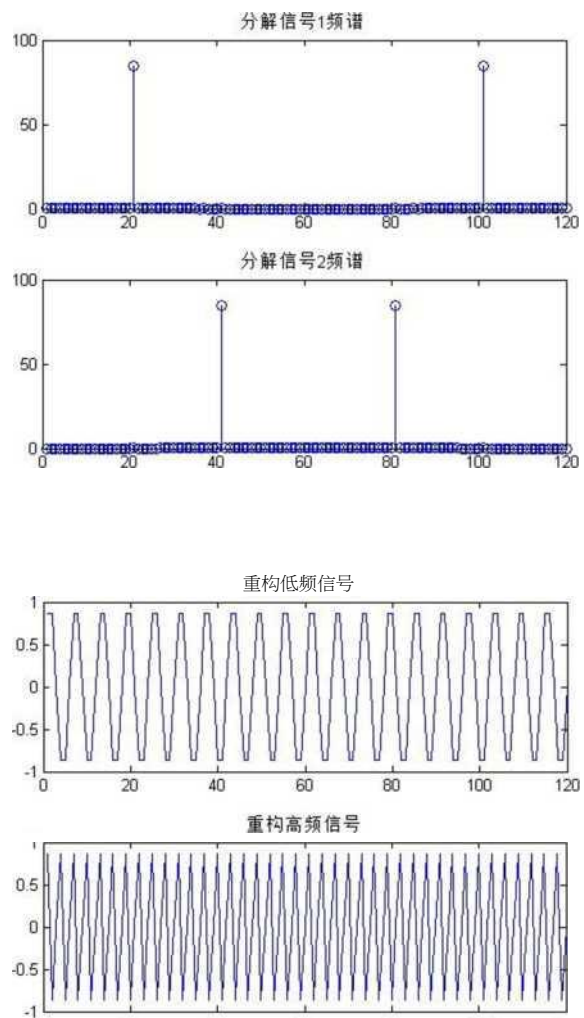


(c)



(d)

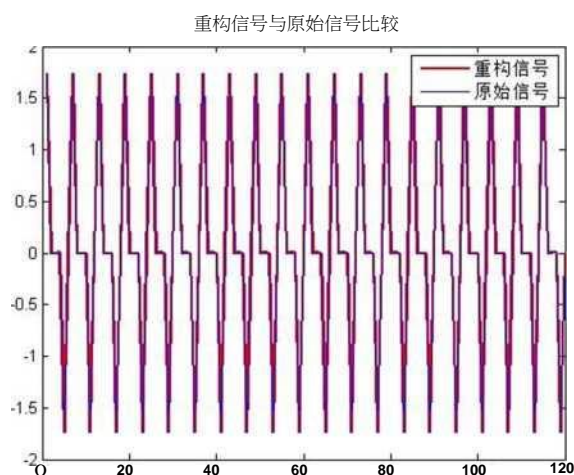




(g)



(h)



(1)

图 4-1 混合正弦波的小波分析示意图

### 4.3 多尺度核函数

定义 4.1 多尺度核函数(Multiscale Kernel)<sup>[13]</sup> 记  $Q \subset \mathbb{R}^d$  是具有紧支撑的、有界的、满足双尺度方程的函数，使得  $\forall x \in \mathbb{R}^d$ , 存在  $k \in \mathbb{Z}^d$ , 满足  $\sum_{k \in \mathbb{Z}^d} \hat{\phi}(x - k) = 1$ 。更进一步，假设双尺度方程中的序列  $\{c_k\}_{k \in \mathbb{Z}^d}$  为有限长，即仅有有限个  $c_k \neq 0$ , 使得  $\sum_{k \in \mathbb{Z}^d} c_k \delta(x - k) = 1$ 。

取定  $\mathbb{Z}^d$ , 以及  $X \subset \mathbb{R}^d$  其中  $\lambda > 0$ , 且  $\lambda \in \mathbb{R}$ , 定义  $U$  函数

$$\phi(x) = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}^d} c_k \phi_j(x - k) \quad (4.3.1)$$

是 Mercer 核函数，称为多尺度核函数。

定义 4.2 有限多尺度核函数(Finite Multiscale Kernel)<sup>[113]</sup> 设  $\phi$  和  $\psi$  与定义 4.1 中有相同的定义，则对某  $u > 1$ ,  $u \in \mathbb{Z}$ , 定义如下的有限多尺度核函数：

$$\phi : (x, y) \mapsto \sum_{j=1}^u \sum_{k \in \mathbb{Z}^d} c_k \phi_j(x - k) \psi_j(y - k) \quad (4.3.2)$$

其中， $u$  和  $\lambda$  分别表示核函数所包含尺度的上、下界。

Mercer 条件的充要条件：一个平移不变核  $k(x, z) = k(x - z)$  是一个容许 SVM 核函数的当且仅当它的傅里叶变换满足 HD

$$\int_{\mathbb{R}^d} \hat{k}(\xi) \hat{k}(\eta) e^{-i\xi \cdot \eta} d\xi d\eta > 0 \quad (4.3.3)$$

如果一维小波函数记为  $\phi(x)$ , 则  $d$  维空间中的小波函数可以表示为：

$$\phi(x) = \sum_{k \in \mathbb{Z}^d} c_k \phi(x - k) \quad (4.3.4)$$



因为 Morlet 小波函数满足 Mercer 条件，因此可以作为 SVR 的核函数，下面给出其对应的核函数的证明。

Morlet 小波函数定义为

$$\psi(x) = \cos(1.75x) \exp(-x^2) \quad (4.3.5)$$

则 Morlet 小波核函数定义为

$$K(x, z) = \frac{1}{2} \cos(1.75 \frac{x-z}{a}) \exp(-\frac{(x+z)^2}{2a^2}) \quad (4.3.6)$$

计算  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(x, z) dx dz$  的傅里叶变换可以得到

$$F(K(x, z)) = \exp(-\frac{a^2}{2}) + \exp(-\frac{a^2}{2}) \quad (4.3.7)$$

又  $K(x, z) > 0$ ，所以  $F(K(x, z)) > 0$ ，即它可作为 SVR 的核函数。

当  $K(x, z)$  作为 SVR 的核函数时，被估计的函数  $f(x)$  可以表示如下：

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) \quad (4.3.8)$$

所以多尺度小波核函数的形式为：

$$MWK(x, z) = \sum_{j=1}^n \alpha_j \cos(1.75 \frac{x-z}{a_j}) \exp(-\frac{(x+z)^2}{2a_j^2}) \quad (4.3.9)$$

其中， $a_j$  为尺度数， $\alpha_j$  为加权因子， $a_j$  为  $j$  尺度上的伸缩因子。

#### 4.4 ARIMA-SVR 模型的实验

本章选用“华谊兄弟”股票 2014 年 3 月 17 日-2015 年 4 月 7 日的每交易日的收盘价格作为研究对象，共 220 个样本，如图 4·2。本章对前 210 个的测试样本进行建模分析，然后对后 10 天的股票收盘价格进行预测，并与实际数据进行对比。

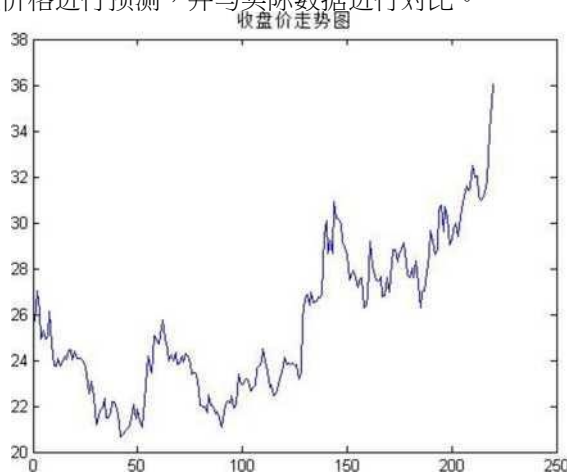
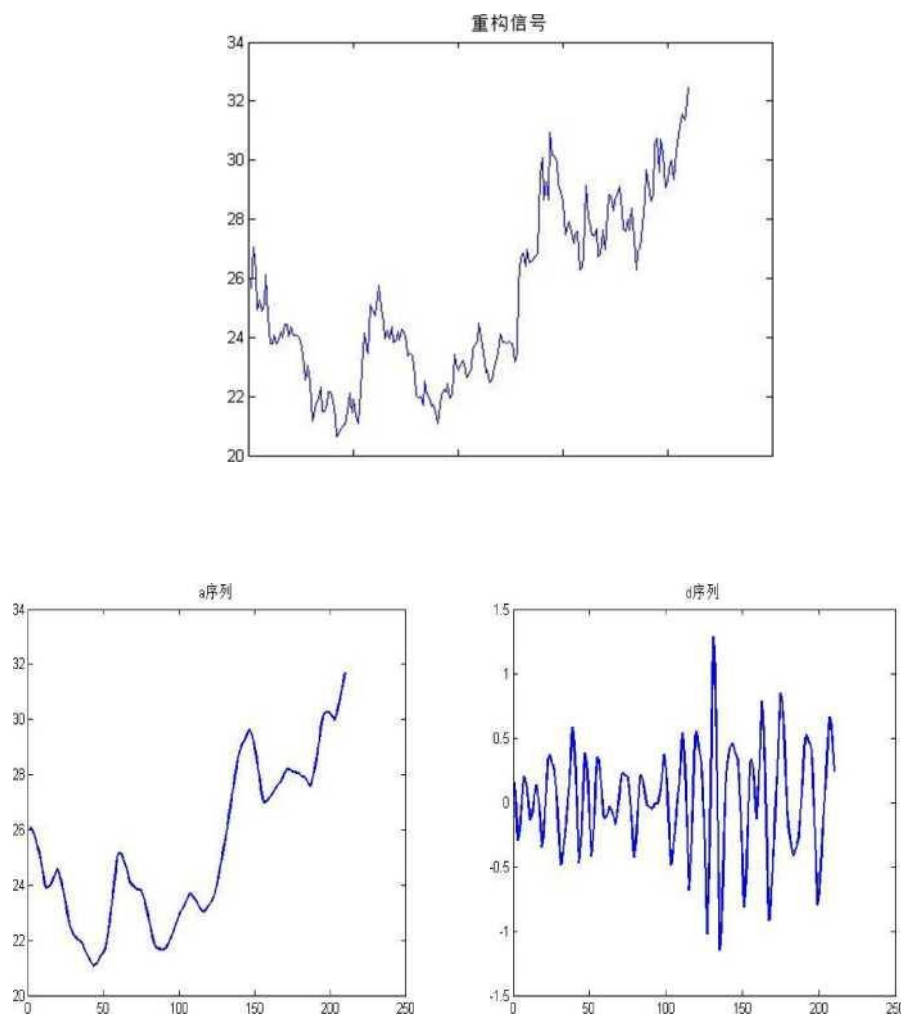


图 4·2 “华谊兄弟”股票收盘价走势图

## 1) 数据的小波分析

由图 4·2 可知, 该曲线是一个具有趋势性的非平稳时间序列, 对上述数据用 `sym4` 小波进行 Mallat 算法的小波分析, 得到原数据的重构信号和低频信息部分  $a$  序列和低频信息部分  $Q$



序列, 如图 4-3 和 4-4 所示。

图 4·4 重构后的高频信息和低频信息序列示意图

## 2) 高频信息部分的 ARIMA 模型预测

应用 ARIMA 模型对  $d$  序列进行预测, 其预测步骤如下:

**Step1:**  $d$  序列对其进行平稳化处理。处理后发现  $d$  已经平稳, 所以设定 ARIMA 模型参数  $d = 0$

**Step2:** 通过反复对模型进行估计比较不同模型的变量对应参数的显著性来确定模型阶数, 得到  $p = 2, g = 1$  时, ARIMA 模型的预测效果最好, 预测样本的拟合优度为 93.25%。

图 4·3 小波重构后股价走势图

Step3: 利用确定的模型 ARIMA (2,0,1) 对最后 10 个样本进行预测。

### 3) 低频信息部分的 SVR 模型预测

对 Q 序列采用 K · CV 方法确定回归模型的最佳参数  $c$  和  $g$ , 在图 4 · 5 中,  $x$  轴表  $\log_2 c$  取以 2 为底的对数后的值,  $y$  轴表示  $\log_2 g$  取以 2 为底的对数后的值, 等高线图亦取相应的  $c$  和  $g$  后对应的 K-CV 方法后的最低的均方误差。可知  $c = 0.5$ ,  $g = 8$ ,  $MSE = 0.00117134$ , 并进行反复多次试验, 对比均方误差, 最后确定阶数  $M = 40$  使用最优参数下训练得到的模型对图 4 · 4 的序列进行预测。

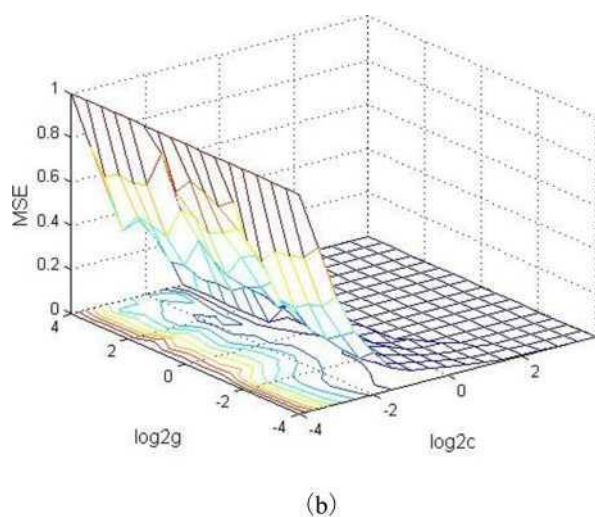
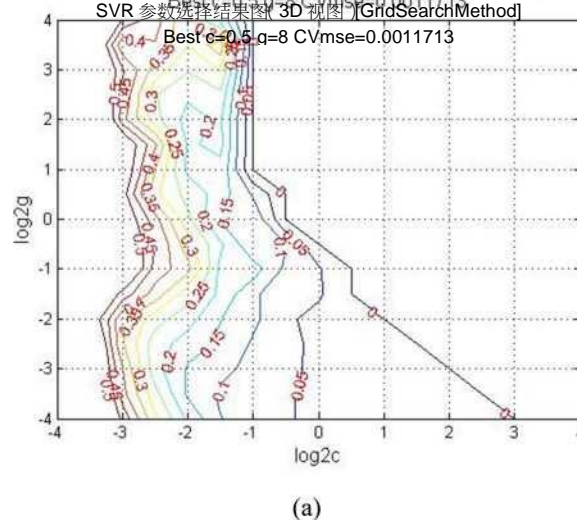


图4-5 参数选择结果图  
SVR 参数选择结果图 (等高线图) [GridSearchMethod]  
SVR 参数选择结果图 3D 视图 [GridSearchMethod]  
Best c=0.5 g=8 CVmse=0.0011713



### 4) 组合预测

把两种模型的预测结果进行相加, 就得到 ARIMA-SVR 模型的预测结果, 如图 4 · 6 所示。可见预测的总体趋势和实际股票价格的走势一致, 预测效果较好。

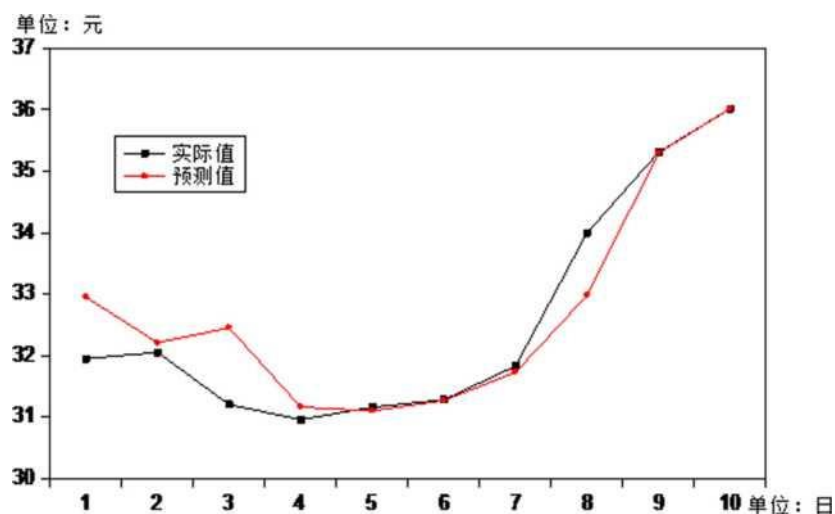


图 4·6 ARIMA-SVR 模型预测走势图

## 4.5 多尺度核函数模型的实验

为与前面 ARIMA-SVR 组合模型进行对比，本实验仍然以华谊兄弟股票收盘价作为研究对象。通过对多尺度小波核函数编译程序对样本数据训练得到模型，并预测。表 4·1 列举出四种模型的 10 个测试样本的预测数据，表 4·2 则给出了其 RMSE 结果，由表 4·1 画出图 4·7,可以清晰的看出各种模型的预测效果。

表 4·1 四种模型预测结果对比表

日期	实际值	ARIMA 模型	SVR 模型	ARIMA-SVR 模型	多尺度小波核函数模型
2015-03-24	31.95	33.54	33.54	32.95	31.97
2015-03-25	32.05	32.83	32.83	32.21	32.01
2015-03-26	31.20	32.75	32.75	32.45	31.08
2015-03-27	30.96	31.57	31.57	31.16	31.03
2015-03-30	31.16	31.65	31.65	31.10	31.20
2015-03-31	31.28	31.88	31.88	31.27	31.25
2015-04-01	31.83	32.45	32.45	31.73	31.85
2015-04-02	34.00	33.46	33.46	32.98	33.87
2015-04-03	35.31	34.27	34.27	35.30	35.36
2015-04-07	36.02	35.82	35.82	36.02	36.00

表 4-2 四种模型的 RMSE 比较表

模型	RMSE
ARIMA 模型	2.88
SVR 模型	2.67
组合模型	1.92
多尺度小波核函数模型	1.01

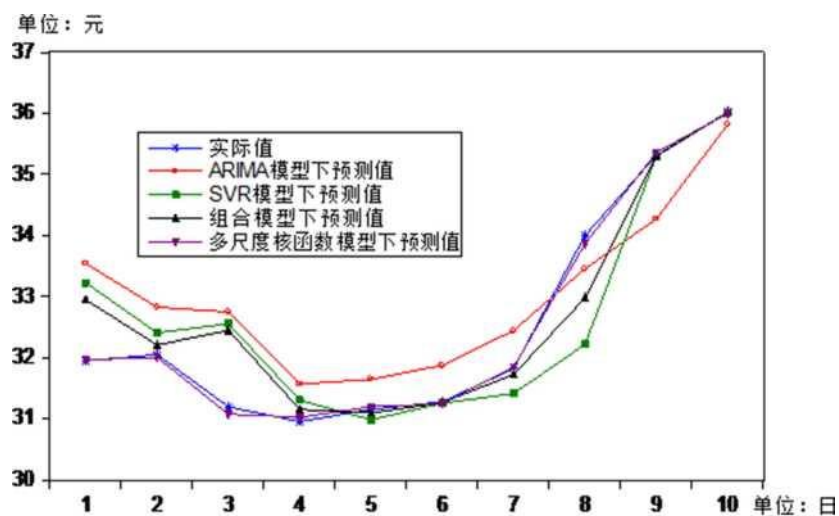


图 4-7 各种模型预测结果走势图

观察表 4·1、表 4·2 及图 4·7 发现：

- (1) 组合模型结合 SVR 模型和 ARIMA 模型二者的优点，且其预测效果明显 优于单一的 SVR 模型和 ARIMA 模型。
- (2) 多尺度小波核函数模型其非线性逼近能力明显优于 SVR 模型，并且在本 数据中它优于组合模型的预测结果，可见其是一种非常有前途的模型。

## 第5章总结与展望

### 5.1 本文总结

支持向量机是国内外研究的热点问题之一，它在数学理论和实际应用中有重要意义，而它与其他学科的结合，更是达到了取长补短的作用，发挥着更加显著的实际应用能力，在本文的经济数据预测得到深刻的验证，越来越多的研究者投身其中。

本文从 **SVR** 在经济中的应用出发，以电力、商品房价格和股票收盘价的预测作为研究对象，得到了以下成果：

(1) 对电力、商品房价格和股票收盘价这 3 种经济数据分别进行 **SVR** 建模，得到了较好的预测效果。

(2) 针对各种预测方法的优点，提出基于小波分析的组合预测模型 **ARIMA-SVR**，比其他单一模型在股市预测更准确、更可靠，是一种有效的股票价格时间序列模型。

(3) 针对目前 **SVM** 在核函数的选择和构造的基础上，提出了一个新的核函数：多尺度小波核函数。通过与其他模型在股市预测上的性能的比较，证明了它是一种非常有前途的模型。

### 5.2 今后展望

本文的研究是对目前 **SVR** 热点研究方面的初步尝试。然而对小波分析、**ARIMA** 与 **SVR** 的结合方式，以及核函数的选择和参数优化问题，依然有很多问题需要进一步解决：

(1) 本文的组合模型的预测效果还是比较好的，但在 **SVR** 模型中，本文采用了 **K·CV** 方法，可以尝试用粒子群算法、遗传算法、免疫算法等算法对模型参数进行优化。

(2) 本文选取了“华谊兄弟”收盘价 2014 年至 2015 年共 220 个样本进行研究，因此它的样本的选择是随机的，会影响预测的精度，可以进一步研究样本的选择对模型预测结果之间的关系。

(3) 本文只建立了 **MoHet** 小波的多尺度核函数 **MWK**，有待进一步研究其它小波核的预测效果。

(4) 在小波变换基础上的组合预测模型，进行对 **SVR** 模型中核函数选择的方法进行预测，至今未有人进行研究，这是对本文两种预测模型最好的完善以及更深入的探究。

## 参考文献

- [1] VAPNIK V. 统计学习理论的本质[M]. 张学工，译. 北京：清华大学出版社，2000.
- [2] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正，王猛，曾华军，译. 北京：电子工业出版社，2004.
- [3] 张丽娜. 支持向量机对股市的预测及实证分析[D]. 青岛：青岛大学，2007.
- [4] 王彦峰，高风. 基于支持向量机的股市预测[J]. 计算机仿真，2006, 23(11):256-258.

- [5] MULLER K, SMOLA A, RATSCH G, et al. Predicting Time Series with Support Vector Machines [J]. Lecture Notes in Computer Science, 1997, 20(2) : 999-1004.
- [6] SMOLA A. Learning with Kernels [D]. Technical University of Berlin, 1998
- [7] SCHOKOPFH B, MIKA S, BURGESS C, et al. Input Space Versus Feature Space in Kernel-Based Methods [J]. Transaction on Neural Networks, 1999, 10(5): 1000-1016.
- [8] CRISTIANINI N. Bayesian Methods for Support Vector Machines: Evidence and Predictive Class Probabilities [J]. Machine Learning, 2002, 46(1): 21-52.
- [9] TAKAHASHI F, ABE S. Decision-tree-based Multiclass Support Vector Machines[C]// International Conference on Neural Information Processing. IEEE, 2002, 3:1418-1422.
- [10] CELLIO A, XAVIER P. A Support Vector Machines for Multiclass[J]. Neural Computation, 2003, 13: 57-77.
- [11] Ahmad, A.R, Khalid, M, Yusof, R. Kernel methods and support vector machines for handwriting recognition[C]//Research and Development, 2002. SCORED 2002. Student Conference on. 2002: 309-312.
- [12] OSUNA E, FREUND R, GIROSI F. Training Support Vector Machines: an Application to Face Detection[C]//Pro. Computer Vision & Pattern Recognition, 1997:130-136.
- [13] JOACHIMS T. Text Categorization with Support Vector Machines[R]. Fakultaten, 1999.
- [14] Drezet P M L, Harrison R F. Support Vector Machines for System Identification [J]. 1998, 1(1): 688-692.
- [15] SUYKENS J A K. Nonlinear Modelling and Support Vector Machines[C]//IEEE Instrumentation & Measurement Technology Conference. 2001, 1: 287-294.
- [16] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
- [17] 黄发良, 钟智. 用于分类的支持向量机[J]. 广西师范学院学报, 2004, 21(3): 75-78.
- [18] 忻栋, 杨莹春, 吴朝晖. 基于 SVM-HMM 混合模型的说话人确定[J]. 计算机辅助设计与 图形学学报, 2002, 14(11): 1080-1082.
- [19] 孙永宣, 谢昭, 高隽. 图像奇异性检测的核分类新方法[J]. 光学学报, 2013, 33(10).
- [20] HWANJO Y, JIAWEI H, et al. PEBL: Web Page Classification without Negative Example[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(1): 70-81.
- [21] 刘永斌, 何清波, 孔凡让, 等. 基于 PCA 和 SVM 的内燃机故障诊断[J]. 振动、测试与诊断, 2012, 32(2): 250-255.
- [22] 王定成, 方廷健. 一种基于支持向量机的内膜控制方法[J], 2004, 21(1): 83-88
- [23] 冯瑞, 张浩然, 邵惠鹤. 基于 SVM 的软测量建模[J]. 信息与控制, 2002, 31(6): 567 · 571.
- [24] 陈念贻, 陆文聪. 支持向量机算法在化学化工中的应用[J]. 2002, 19(6): 673-676.
- [25] 张拥华. 基于支持向量机的金融时间序列研究[D]. 湖南: 湖南大学, 2000

- [26] 李祥飞.基于误差同步预测的 SVM 金融时间序列预测方法[J].天津大学学报, 2014, 47(1): 86-94.
- [27] 黄爱华,蒲洪彬,李伟光,等.基于人工免疫机理和 LS-SVM 的顾客需求重要度影响[J].华南理工大学学报,2013,41(1): 89-94.
- [28] 张朝元.基于 LS-SVM 的大理州入境游客流量组合预测模型[J].科技通报, 2013, 29(9):15-1 &
- [29] 翁钢民,李凌雁.游客流量预测:基于季节调整的 PSO-SVR 模型研究[J].计算机应用研究, 2014,31(3):692-695.
- [30] 姚卫红,方仁孝,张旭东.基于混合人工鱼群优化 SVR 的交通流量预测[J].大连理工大学学报,2015, 55(6): 632-637.
- [31] 汤鑫.基于贝叶斯推理的短期风速预测[D].北京:华北电力大学,2013.
- [32] 罗伟,习华勇.基于最小二乘支持向量机的降雨量预测[J].人民长江, 2008, 39(19): 29-31.
- [33] 翟鸿雁,曾晋明,曾纪霞.基于支持向量机的电力市场价格预测中的核函数比较[J].计算机技术与自动化,2011,30(2): 30-33.
- [34] 罗赞骞,陈志杰,汤锦辉,等.采用支持向量机回归的航班延误预测研究[J].交通运输系统工程与信息,2015, 15(1): 143-149, 172.
- [35] 李志峰,李金伟,吉伟.基于支持向量机的上海市宝山区房价的判别与预测[J].湖北师范学院学报,2011,31(4): 60-65.
- [36] 申瑞娜,曹昶,樊重俊.基于主成分分析的支持向量机模型对上海房价的预测研究[J].数学的实践与认识,2013,43(23): 11-16.
- [37] 张彦周,贾利新.基于网格寻优 SVR 房价预测模型——以郑州市为例[J].河南科学, 2014, 32(8): 1659-1663.
- [38] 刘蓉.城市住房价格 PSO-LSSVR 预测模型研究[D].重庆:重庆大学,2014.
- [39] 王霞,王占岐,金贵,等.基于核函数支持向量回归机的耕地面积预测[J].农业工程学报,2014,30(4):204-211.
- [40] 辛霄.基于贝叶斯支持向量机的粮食产量预测的研究[D].北京:首都师范大学,2012.
- [41] 金超,牟艳,马塑,等.灰色模型和支持向量机组合的预测模型及其应用[J].微型电脑应用, 2015,31(1): 25-28.
- [42] 梁勇.基于支持向量回归的水质预测研究[D].湖北:武汉理工大学,2012.
- [43] 林献坤,李郝林,袁博.基于 PSO-SVR 的数控平面磨削表面粗糙度智能预测研究[J].系统仿真学报,2009, 21(24): 7805-780&
- [44] 王宏伟,张鑫,邱俊楠,等.基于 GA-SVR 的中长期径流预报[J].西北农林科技大学学报,2012, 40(2):201-206.
- [45] 金得宝.基于支持向量机的股市预测[D].浙江:浙江大学,2010.
- [46] 任世锦,吴铁军.基于径向基小波核的多尺度小波支持向量机[J].电路与系统学报, 2008, 13(4): 70-76,80.



- [47] 蒋波.基于混沌优化的多尺度小波核一支持向量机及其在股票市场中的应用[D].甘肃:兰州大学,2012.
- [48] 王佳,徐蔚鸿.基于动量粒子群的混合核 SVM 参数优化方法[J]·计算机应用,2011, 31(2): 501-503.
- [49] 王平,王文剑.基于时序核函数的支持向量回归机[J]·计算机辅助工程,2006, 15(3): 35·3&
- [50] 江泳.基于混合核方法的上下维语义抽取[D].上海:复旦大学,2013.
- [51] 朱孝开.基于核方法的图像目标识别技术研究[D].湖南:国防科学技术大学,2009.
- [52] PAI P, LIN C. A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting [J], Omega, 2005, 33(6): 497-505.
- [53] 程昌品,陈强,姜永也基于 ARIMA-SVM 组合模型的股票价格预测[J].计算机仿真, 2012, 29(6): 343-346.
- [54] 王晴.组合模型在股票价格预测中应用研究[J].计算机仿真,2010, 27(12):361-364
- [55] 丁玲娟.基于小波分析和 ARMA-SVM 模型的股票指数预测分析[D].上海:华东师范大学,2012.
- [56] MOHSEN B, KEYVAN A, MORTEZA E, et al. Generalization Performance of Support Vector Machines and Neural Networks in Runoff Modeling[J]. Expert Systems with Applications, 2009, 36(4): 7624-7629.
- [57] 伏进.重庆市电力消费与经济增长的关系分析及电力产业政策建议[D].重庆:重庆大学, 200&
- [58] 董智,程春萌.河南省电力消费弹性系数变动状况分析[J].电力系统保护与控制,2010, 38(5): 60-63.
- [59] 左金金.北京市电力消费与经济增长的动态关系研究[D].北京:北京工业大学,2013.
- [60] 陈万林.江苏省电力消费分析及预测研究[D].江苏:南京财经大学,2012.
- [61] 邓鸿鹄.北京市能源消费预测方法比较研究[D].北京:北京林业大学,2013.
- [62] CHANG C, LIN C J. LIBSVM: a Library for Support Vector Machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011,2(1):27:1—27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [63] 孔凡文,李忠利,张春雨.沈阳市商品房价预测分析[J].沈阳建筑工程学院学报, 2001, 17(3): 232-238.
- [64] 刘悦婷,郑卓.基于 MATLAB 的兰州市商品住宅价格变动分析预测[J].甘肃科学学报, 2011,23(3): 155-15 &
- [65] 武秀丽,张锋.时间序列分析法在房价预测中的作用[J].科学技术与工程,2007, 7(21): 5631-5635.
- [66] 常振海,刘薇.基于非参数自回归模型的房价预测[J].天水师范学院学报,2010,20(2): 56·5&

- [67] 陈森君.中国商品房价格的影响因素分析及价格预测[D].湖北:华中科技大学,2006.
- [68] 徐富强.神经网络组合预测模型及其在房价趋势中的应用[D].安徽:安徽大学,2011.
- [69] 陈博.西安市商品房价格趋势的研究及其预测[D].陕西:西安建筑科技大学,2012.
- [70] 曹玉.我国商品房价格影响因素研究[D].云南:云南大学,2011.
- [71] 董艳芳.我国商品房价格形成的主要因素及形成机制分析[J].软科学, 2012,26(6): 96 · 9&
- [72] 上海市统计局.上海市统计年鉴[M].北京:中国统计出版社,1999-2014.
- [73] 王黎明,王连,杨楠.应用时间序列分析[M].上海:复旦大学出版社,2012.
- [74] CH AMPERN O WNE S D G. Sampling Theory Applied to Autoregressive Sequences [J]. Journal of the Royal Statistical Society, 1948, 10(2): 204-242.
- [75] Engle R F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation [J]. Econometrical, 1982, 50(4): 987-1008.
- [76] BOLLERSEY T. Generalized Autoregressive Conditional Heteroscedasticity[J]. Journal of Econometrics, 1986, 31(3): 307-327.
- [77] 郑挺国.基于有限混合状态空间的金融随机波动模型及应用研究[D].吉林:吉林大学, 2009.
- [78] CHATFIELD C, WEIGEND AS. Time series prediction: Forecasting the future and understanding the past[J]. International Journal of Forecasting, 1994, 10(1): 161-163.
- [79] Hassan M R, Nath B. Stock Market Forecasting Using Hidden Markov Model: A New Approach[C]//International Conference on Intelligent Systems Design and Applications, 2005. Isda \*05. Proceedings. 2005:192-196.
- [80] 吴漫君.基于隐马尔科夫模型的股价走势预测[D].广东:华南理工大学,2011.
- [81] 陈诗一.非参数支持向量回归和分类理论及其在金融市场预测中的应用[M].北京:北京大学出版社,2008.
- [82] HANSEN J V, NELSON R D. Neural networks and Traditional Time Series Methods: A Synergistic Combination in State Economic Forecasts [J]. IEEE Transactions on Neural Networks, 1997, 8(4): 863-873.
- [83] ZHANG D, ZHOU L. Discovering Golden Nuggets: Data Mining in Financial Application [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 2004, 34(4): 513-522.
- [84] Chang P C, LIU C H. A TSK Type Fuzzy Rule Based System for Stock Price PredictionfJ]. Expert Systems with Application, 2008, 34(1): 135-144.
- [85] ZARANDI M H F, REZAEE B, TURKSEN I B. A Type-2 Fuzzy Rule Based Expert System Model for Stock Price Analysis [J]. Expert Systems with Application, 2009, 36(1): 139-154.
- [86] MATSUB A, I. Application of Neural Sequential Associator to Long-term Stock Price PredictionfJ].

- Japanese Railway Engineering, 2006, 2(46): 1196-1201..
- [87] 马千里, 郑启伦, 彭宏, 等. 基于相空间重构理论与递归神经网络相结合的股票短期预测方法[J]. 计算机应用研究, 2007, 24(4): 239-245.
- [88] PAWLAK 乙 Rough Sets and in Intelligent Data Analysis[J]. Information Sciences, 2002, 147(14): 1-12.
- [89] TAY F, CAO L. Improved Financial Time Series Forecasting by Combining Support Vector Machines with Self-organizing Feature Map [J]. Intelligent Data Analysis, 2001, 5(4): 339-354.
- [90] CAO L, TAY F. Financial Forecasting Using Support Vector Machines [J]. Neural Computing&Applications, 2001, 10(2): 184-192.
- [91] TAY F, CAO L. Application of Support Vector Machines in Financial Time Series Forecasting [J]. Omega, 2001, 29: 309-317.
- [92] TAY F, CAO L. s-Descending Support Vector Machines for Financial Time Series Forecasting [J]. Neural Processing Letters, 2002, 15(2): 179-195.
- [93] TAY F, CAO L. Modified Support Vector Machines in Financial Time Series Forecasting [J]. Neurocomputing, 2003, 4& 847-862.
- [94] CAO L, TAY F. Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting [J]. IEEE Transactions on Neural Networks, 2003, 14(6): 1506-1511 &
- [95] TRAFALIS T, INCE H. Support Vector Machine for Regression and Applications to Financial Forecasting [J]. IEEE, 2000: 348-353.
- [96] INCE H, TRAFALIS T. Short Term Forecasting with Support Vector Machines and Application to Stock Price PredictionfJ], International Journal of General Systems, 2008, 37(6): 677-687.
- [97] 李拥军, 奉国和, 齐德星. 快速增量加权支持向量机预测证券指数[J]. 控制理论与应用, 2006, 23(5): 805-509.
- [98] 李立辉, 田翔, 杨海东, 等. 基于 SVR 的金融时间序列预测[J]. 计算机工程与应用, 2005, 41(30): 221-22.
- [99] 曲文龙, 樊广俭, 杨炳儒. 基于支持向量机的复杂时间序列预测研究[J]. 计算机工程, 2005, 31(23): 1-3.
- [100] 张玉川, 张作泉. 支持向量机在股票价格预测中的应用[J]. 北京交通大学学报, 2007, 31(6): 73-76.
- [101] 杨一文, 杨朝军. 基于支持向量机的金融时间序列预测[J]. 系统工程理论方法应用, 2005, 14(2): 176-181.
- [102] 阎威武, 常俊林, 邵惠鹤. 基于滚动时间窗的最小二乘支持向量机回归估计方法及仿真[J].

- 上海交通大学学报, 2004, 38(4): 524-526.
- [103] 张晨希, 张燕平, 张迎春, 等. 基于支持向量机的股票预测[J]. 计算机技术与发展, 2006, 16(6): 35-37.
- [104] 甘昕艳, 张铨玲. 基于股价指数预测的仿真研究[J]. 计算机仿真, 2010, 27(10): 297-300.
- [105] 陈振伟, 郭拯危. 小波神经网络预测模型的仿真实现[J]. 计算机仿真, 2008, 25(6): 147-150.
- [106] 张玉川, 张作泉. 支持向量机在股票价格预测中的应用[J]. 北京交通大学学报, 2007, 31(6): 73-76.
- [107] 张克宜. 基于改进 GM(1,N) 和优化 SVM 组合模型的股票价格预测[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [108] 詹财鑫. 基于 SVM-AdaBoost 模型的股票涨跌实证研究[D]. 广东: 华南理工大学, 2013.
- [109] ZHANG L, ZHOU W D, JIAO L C. Wavelet Support Vector Machine [J]. IEEE Transaction on Systems, man, and Cybernetics-part B: cybernetics, 2004, 34(1): 34-39.
- [110] 林继鹏, 刘君华. 基于小波的支持向量机算法研究[J]. 西安交通大学学报, 2005, 39(8): 816-819.
- [111] CHEN G, DUDEK G. Auto-Correlation Wavelet Support Vector Machine and Its Applications to Regression[C]//2013 International Conference on Computer and Robot Vision. IEEE Computer Society, 2005:246-252.
- [112] 胡昌华, 李国华, 周涛. 基于 MATLAB 7.x 的系统分析与设计——小波分析[M]. 陕西: 西安电子科技大学出版社, 2008.
- [113] OPFER R. Multiscale Kernels [J]. Advances in Computation Mathematics, 2006, 25(4): 357-380.

## 致谢

当毕业论文完成就意味着宣告 3 年硕士研究生生涯的落幕，真挚的感谢从我的内心深处油然而生！

首先要感谢我的导师郑伯川教授，在论文写作期间，他对我的悉心指导和论文的耐心修改是我完成论文的充要条件。3 年来，郑老师不仅传授了我丰富的专业知识，成为我学术上的导师，而且人生态度和方向上给予我以深刻的理解，成为我人生的导师。他认真严谨的做学问方式深深地感染了我，使我受益匪浅。正是这些使我在学习上不断努力和进取，顺利完成学业。在日常生活中，郑老师也是非常平易近人的，像亲人和朋友似的与我相处，使我在这个陌生而熟悉的地方扎根，现在的我对这个地方真是恋恋不舍。

同时感谢西华师范大学数学与信息学院的领导和老师，他们不仅在学术上的探索精神始终感染着我们，让我们在 3 年的求学路上获益匪浅，而且为我们创造了良好的学习环境，让我们在实验室里认真学习，陶醉在知识的海洋中。

感谢各位师兄师姐和师弟师妹在学习上给予我的巨大帮助和支持，让我在学校组织的项目活动中，取得圆满的成果。

感谢各位舍友在生活上对我的诸多照顾，让我很快融入这个地方，在平常的相处中容纳我小小的脾气，接受我成为宿舍大家庭的一员，让我全心全意的搞学问。

最后，感谢始终支持我的亲朋好友，是他们对我的爱和理解让我在人生道路上披荆斩棘，不断前行。

关于学位论文使用授权的声明

本人在导师指导下完成的本硕士学位论文，知识产权归本人和西华师范大学共有。

本人完全了解西华师范大学有关收集、保存、使用学位论文的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅。

本人授权西华师范大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本学位论文。

本人离校后发表本学位论文中的成果时，第一署名单位应为西华师范大学。

保密论文在解密后应遵守此规定。

论文作者签名：换粧芳 \_\_\_\_\_

指导教师签名：剖倫 W 日 期：年£月纾日

关于学位论文原创性的声明

本人郑重声明：本人所呈交的学位论文，是在导师的指导下独立 进行研究  
所取得的成果。学位论文中凡引用他人已经发表或未发 表的成果、数据、观点  
等，均已明确注明出处。除文中已经注明 引用的内容外，不包含任何其他个人  
或集体已经发表或撰写过的作品 或科研成果。对本文的研究成果做出重要贡献  
的个人和集体，均已在 文中以明确方式标明。

论文作者签  
名：

郑 伟 川

2016 年 5 月 28 日

本声明的法律责任由本人承担。

## 在学期间的科研情况

### 一、发表的学术论文

1. 袁秀芳，焦伟超，郑伯川.成都市电力消费与经济增长的实证分析[J].西 华师范大学学报,2014, 35(4): 393-397.
2. 袁秀芳，张征，郑伯川，焦伟超.基于 **SVR** 的多变量电力消费预测[J].西 华师范大学学报,2015,36(3): 289-294.
3. 袁秀芳，郑伯川，焦伟超.基于 **SVR** 的上海市商品房价格预测[J].甘肃科 学学报,2016, 28(1): 25-2&