

基于模糊近似支持向量回归的股价预测研究

王冰玉, 刘勇军

(华南理工大学, 广东 广州 510640)

摘要: 股价预测是投资策略形成和风险管理模型发展的基础。为了降低股价变化趋势中的噪声信息和投资者关于两种股价预测误差的不同偏好对股价预测的影响,提出了基于信噪比的模糊近似支持向量回归(FPSVR)的股价预测模型。首先构建信噪比输入变量,然后引入模糊隶属度和双边权重测量方法对支持向量回归(SVR)模型进行改进,最后借助沪深300成份股2008至2019年的股票时间序列日数据,按照股市的波动情况将其分为三个阶段(牛市、熊市、震荡市),并建立三个基准模型进行对比分析。研究结果表明:与三个基准模型相比,所提出的股价预测模型的预测误差最低;与原有的SVR模型相比,FPSVR模型可以更好地对处于牛市和震荡市阶段的股票时间序列进行股价预测。

关键词: 股价预测;支持向量回归;信噪比;输入指标;模糊近似支持向量回归

中图分类号: TP18; F224

文献标识码: A

文章编号: 1673-629X(2021)03-0014-07

doi: 10.3969/j.issn.1673-629X.2021.03.003

Research on Stock Price Prediction Based on Fuzzy Proximal Support Vector Regression

WANG Bing-yu, LIU Yong-jun

(South China University of Technology, Guangzhou 510640, China)

Abstract: Stock price forecasting is the basis of the formation of investment strategy and the development of risk management model. In order to reduce the influence of noise information in the trend of stock price change and investors' different preferences for two kinds of stock price prediction errors on stock price prediction, we propose a stock price prediction model based on fuzzy approximate support vector regression (FPSVR) based on signal-to-noise ratio. Firstly, the input variables of signal-to-noise ratio are constructed, and then fuzzy membership degree and bilateral weight measurement methods are introduced to support vector regression (SVR) model. Finally, with the help of the 2008-2019 stock time series day data of the 300 component stocks in Shanghai and Shenzhen, according to the volatility of the stock market, it can be divided into three stages (bull market, bear market and shock market), and three benchmark models are established for comparative analysis. The results show that compared with the three benchmark models, the prediction error of the proposed model is the lowest. Compared with the original SVR model, the FPSVR model can better predict the stock price in the bull market and volatility stage.

Key words: stock price prediction; support vector regression; signal to noise ratio; input index; fuzzy proximal support vector regression

0 引言

股价预测主要基于相关预测模型预测股价的未来变化趋势以捕捉相应的市场行情,从而促进股票选择。近年来,股价预测是热点问题。许多学者和业界人士都从理论和实践层面对其股价预测模型进行相关研究,其研究方法包括 GARCH^[1-3]、模糊时间序列^[4-5]、ARIMA^[6-8]等。由于计算机技术的发展和广泛应用,各种机器学习和量化投资模型也被逐渐地应用到股票

预测领域,例如支持向量机、人工神经网络、遗传算法和差分进化算法等股价预测模型。

Yu 和 Yan 基于 PSR 方法和 DL 的长短期记忆网络(LSTMs),设计了一种基于 DNN 的股价预测模型,并对不同时期的多个股票指数进行预测^[9]。Onoh 等利用协调搜索(HS)和遗传算法,建立混合人工神经网络模型,并根据数据集的统计和财务表现进行实证分析^[10]。Moghaddam 等研究人工神经网络对纳斯达克

收稿日期: 2020-04-14

修回日期: 2020-08-25

基金项目: 国家自然科学基金项目(71971086); 广东省自然科学基金-杰出青年项目(2019B151502037); 中央高校科研基本业务费重点项目(2019ZD13)

作者简介: 王冰玉(1992-),女,硕士研究生,研究方向为股票预测的机器学习算法研究;刘勇军,教授,博导,研究方向为基于智能算法的投资决策及金融风险管控。

股票日汇率的预测能力,从而对用反向传播算法训练的几种前馈神经网络进行了评估^[11]。Hájek 等将情绪与财务指标相结合,使用一个多层感知器神经网络来预测异常股票回报率^[12]。綦方中等提出一种基于 PCA-IFOA-BP 神经网络的股票价格预测模型,并借助上证指数进行预测^[13]。孟叶等选取 K-近邻、梯度提升和自适应提升这 3 个分类器,通过改进的投票算法聚合成一个新的分类器模型,对指数行情数据进行学习分类^[14]。

虽然基于神经网络的混合模型是预测股票市场指数的有效预测方法,但是却存在着诸如黑箱技术、过拟合、收敛速度慢、陷入局部极小等局限性。为了克服这些局限性,Cortes 和 Vapnik 提出的支持向量机(SVM)方法已经成为股票市场指数预测领域的一种流行研究方法,它采用结构风险最小化的原则,以最小化泛化误差的上界。通过应用支持向量机,不太可能出现过拟合,而且最优解也可能是全局的^[15]。Nayak 等构建一种支持向量机与 K 近邻法相结合的混合框架,用于印度股市指数的短期、中期和长期预测^[16]。Chen 等引入特征加权向量,进一步提出特征加权支持向量机和特征加权 K 最近邻的基本混合框架,以有效地预测股票市场指数^[17]。Xiao 等提出一种将 SVM 和 SSA 相结合的组合模型,并建立基准模型进行对比分析^[18]。Lee 创建一种基于支持向量机(SVM)和混合特征选择方法的股市趋势预测模型^[19]。Fung 等提出一种新的近端支持向量机(PSVM)模型,并使用公开的数据集上的计算结果表明,PSVM 不仅具有与 SVM 相当的测试集正确性,而且具有相当快的计算时间^[20]。姚潇和余乐安在 PSVM 的基础上,引入模糊隶属度的思想,提出模糊近似支持向量机(FPSVM),并利用两个公开的信用数据集进行实证研究验证该模型的有效性^[21]。张贵生等针对股价的非线性特点,提出基于近邻互信息特征选择 SVM-GARCH 的股价预测模型^[22]。张冰等提出具有局部信息挖掘功能的 DNN 加权算法对 eplion-TSVR 模型进行改进,并借助上证 A 股的高频数据进行实证预测^[23]。

综上所述,首先虽然 SVR 能够在股价预测中发挥更好的表现力,但是都未考虑股票数据中所含的噪声信息对股价预测的影响,其次现有研究未曾考虑投资者对两种预测误差((1)预测值>实际值;(2)预测值<实际值)的不同偏好。有的投资者更注重收益,而有的投资者更关注损失,因此,不能简单直接地使用 SVR 模型进行股价预测,而是应该根据实际情况,针对投资者的不同目标赋予不同的偏好值,构建有效的股价预测模型进行股票价格的准确预测。

综合考虑以上研究现状和存在的不足之处,该文

构建信噪比特征向量,并借用现有文献表明的相关股价预测指标,选取历史数据、趋向、反趋向、能量、量价、波动和信噪比等其他七个方面的指标作为输入变量,考虑到投资者对预测误差的不同偏好,引入模糊隶属度和双边权重测量方法,构建基于信噪比的模糊近似支持向量回归模型进行股价预测。

1 模糊近似支持向量机

模糊近似支持向量机(FPSVM)是由姚潇和余乐安在 2012 年提出的^[21]。该模型是为了减小训练样本的奇异点和噪声对模型的干扰,在 PSVM 的基础上,引入了模糊隶属度。设含有 N 个训练样本的训练集样本对 $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in R^d$ 为第 i 个训练样本的输入向量, d 为样本空间维度, $y_i \in \{+1, -1\}$ 为对应输出值,则 FPSVM 模型的具体形式如下:

$$\begin{aligned} \min \quad & \frac{1}{2}(w^T w + b^2) + \frac{C}{2} \sum_{i=1}^N m_i \varepsilon_i^2 \\ \text{s. t.} \quad & \begin{cases} y_i(w\Phi(x_i) + b) = 1 - \varepsilon_i, i = 1, 2, \dots, N \\ \varepsilon_i \geq 0 \end{cases} \end{aligned} \quad (1)$$

其中, $C > 0$ 为错误项的惩罚参数, ε_i 为松弛变量, $\Phi(x)$ 为非线性映射函数, w^T 为特征空间维数, b 为待定的标量参数, m_i 为隶属度,表示第 i 个样本点对超平面的贡献率。 m_i 越小,则误差项所占的比例越小,在整个模型的影响就越小。当所有的 $m_i = 1$ 时, FPSVM 就退化为 PSVM。

2 股价预测输入指标的获取

由于买卖反弹,价格变化的离散性、交易规模的差异和订单流的战略组成部分等因素导致所观察到的价格过程是一个包含噪声的过程^[24]。现实生活中,一般所观察的对数价格过程包含对数有效价格和噪声两部分,相应的股价对数形式为:

$$\ln p_i = \ln x_i' + \ln n_i \quad (2)$$

其中, p_i 为股票第 i 个输入样本价格, x_i' 为第 i 个输入样本有效价格, n_i 为第 i 个输入样本噪声。从式(2)可看出,对数形式的股价包含一个加性噪声部分,则对数形式下的股票噪声与有效价格可以定义为^[24]:

$$\begin{aligned} \ln x_i' &= \ln x_i - \ln n_i \\ \ln n_i &= \ln x_i - \sum_{l \in C} \frac{1}{K} \ln x_l' \end{aligned} \quad (3)$$

其中, $\sum_{l \in C} \frac{1}{K} \ln x_l'$ 为样本有效价格 x_i' 的 k 个最近邻均值。由式(2)和式(3),可得到股价的噪声部分 $e^{\ln n_i}$ 与有效价格部分 $e^{\ln x_i'}$ 。

噪声的存在会影响股价预测。若该文能在 SVM 方程的构建中进一步消除噪声,则可提高股价预测精度。信噪比(SNR)是描述信号中有效成分与噪声成分的比例关系参数,可以有效地对噪声进行处理^[25]。该文考虑引入 SNR 降低噪声干扰。根据现有文献,可得到信噪比(SNR)的计算公式^[26]:

$$\begin{aligned} \text{SNR} &= \frac{\bar{x}}{\sigma_x} \\ \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i \\ \sigma_x &= \sqrt{\sum_{i=1}^N (x_i - 1/k \sum_{l \in \mathcal{C}} x_l)^2 / N} \end{aligned} \quad (4)$$

其中, N 为所使用样本数据的数量, \bar{x} 为有效价格均值, σ_x 为噪声方差。SNR 可作为反映股票收益波动的重要指标之一, SNR 值越大, 收益越稳定, 风险就越小。

此外, 股票市场中, 开盘价、最低价、最高价和总交易量历史数据通常被用做输入指标。最近相关学者研究表明一些技术指标有助于更好地预测股价^[17]。故该文同时借助相关的技术指标作为输入变量, 具体将这些指标分为:

- (1) 趋向指标: 移动平均线(MA)、指数移动平均线(EMA)、异同移动平均线(MACD)和动量指标(MTM);
- (2) 反趋向指标: 相对强度指数(RSI);
- (3) 能量指标: AR、BR 和成交量变异率(VR);
- (4) 量价指标: 多空比率净额(DK);
- (5) 波动指标: 平均真实范围(ATR)。

3 基于 FPSVR 的股价预测模型构建

3.1 FPSVR 模型构建

直接采用 SVR 模型进行股票预测会存在以下不足: 首先原有的 SVR 模型都未考虑投资者对股价预测误差的不同偏好情况, 默认投资者对预测误差的偏好是一致的, 使得预测结果不能准确反映投资者的投资策略, 其次 SVR 模型无法处理噪声和奇异点对模型的干扰状况, 影响模型的预测准确度。

因此, 为了解决上述问题, 该文考虑将模糊隶属度引入, 对股价预测问题进行模糊化处理, 消除样本中的奇异点和噪声对股价预测的影响, 同时分别考虑预测值大于实际值的误差项 ξ_i 和预测值小于实际值的误差项 ξ_i^* 两种情况对股价预测的影响, 分别对误差项 ξ_i 和误差项 ξ_i^* 赋予不同的权重 m_i 和 n_i , 则会得到如下形式的训练样本:

$$\{x_i, \hat{y}_i - y_i > 0, m_i\}, \{x_i, \hat{y}_i - y_i < 0, m_i\}$$

其中, x_i 为第 i 个输入向量, 即第 i 日股票的所有输入信息变量, y_i 为第 i 个输出向量, 即第 i 日股票的所有输出信息变量, \hat{y}_i 为预测值, 即第 i 日股票的所有预测输出信息变量, $\hat{y}_i - y_i > 0$ 表示第 i 日的股票预测价格大于实际价格, $\hat{y}_i - y_i < 0$ 表示第 i 日的股票预测价格小于实际价格, m_i 和 n_i 表示第 i 日股票在出现 $\hat{y}_i - y_i > 0$ 和 $\hat{y}_i - y_i < 0$ 两类情况下的隶属度。

故而, 该文得到基于模糊近似支持向量回归的股价预测模型, 其具体形式如下:

$$\begin{aligned} \min \quad & \frac{1}{2} (w^T w + b^2) + \frac{C}{2} \left(\sum_{i=1}^N m_i \xi_i^2 + \sum_{i=1}^N n_i \xi_i^{*2} \right) \\ \text{s. t.} \quad & \begin{cases} -y_i + w\Phi(x_i) + b = \varepsilon + \xi_i \\ y_i - w\Phi(x_i) - b = \varepsilon + \xi_i^* \end{cases} \end{aligned} \quad (5)$$

其中, i 为所选取股票的第 i 个样本, N 为训练样本总数, x_i 为第 i 个输入向量, $x_i = (x_{i1}, x_{i2}, \dots, x_{i15})$ 分别代表开盘价、最高价、最低价、总交易量、移动平均线、指数移动平均线、异同移动平均线、动量指标、相对强弱指数、人气指标、意愿指标、成交量比率、多空比率额、真实波动幅度均值和信噪比, y_i 为实际股票收盘价。

通过构造拉格朗日函数对式(5)进行求解, 得到:

$$\begin{aligned} L &= \frac{1}{2} (w^T w + b^2) + \frac{C}{2} \sum_{i=1}^n m_i \xi_i^2 + \frac{C}{2} \sum_{i=1}^n n_i \xi_i^{*2} - \\ & \sum_{i=1}^n \alpha_i (-y_i + w\Phi(x_i) + b - \varepsilon - \xi_i) - \\ & \sum_{i=1}^n \beta_i (y_i - w\Phi(x_i) - b - \varepsilon - \xi_i^*) \end{aligned} \quad (6)$$

分别对 w, b, m_i, n_i 求偏导, 得:

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i \Phi(x_i) + \sum_{i=1}^N \beta_i \Phi(x_i) = 0 \\ \frac{\partial L}{\partial b} = b - \sum_{i=1}^N \alpha_i + \sum_{i=1}^N \beta_i = 0 \\ \frac{\partial L}{\partial \xi_i} = C m_i \xi_i + \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} = C n_i \xi_i^* + \beta_i = 0 \\ i = 1, 2, \dots, N \end{cases} \quad (7)$$

对式(7)求解得到:

$$\begin{aligned} b^* &= \sum_{i=1}^n (\alpha_i - \beta_i) \\ w^* &= \sum_{i=1}^n [\alpha_i \Phi(x_i) - \beta_i \Phi(x_i)] \end{aligned} \quad (8)$$

$$\alpha_i = -C m_i \xi_i$$

$$\beta_i = -C n_i \xi_i^*$$

故上述规划问题变为:

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \beta_i) \times (\alpha_j - \beta_j) (\Phi(x_i)^T \Phi(x_j))$$

$$+ 1) + \frac{1}{2C} \sum_{i=1}^N \frac{\alpha_i^2}{m_i} + \sum_{i=1}^N \frac{\beta_i^2}{n_i} \quad (9)$$

求得原问题的最优解为 (α^*, β^*) , 其中 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$, $\beta^* = (\beta_1^*, \beta_2^*, \dots, \beta_N^*)$ 。同时令 $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$, 则通过式(8)和式(9)可求得最后的预测函数为:

$$f(x) = w^* \Phi(x) + b = \sum_{i=1}^N (\alpha_i - \beta_i) K(x_i, x_j) + \sum_{i=1}^N (\alpha_i - \beta_i) \quad (10)$$

3.2 隶属度的获取

为了更准确地进行股价预测,通过将欧氏距离和信息增益结合起来,提出一种新的隶属度测量方法。首先,当样本 x_i 更贴近误差项 $\hat{y}_i - y_i > 0$ 时,表明该样本的预测值大于实际值,故对权重 m_i 赋予越大的数值,即对第一类误差项(预测值>实际值)的惩罚力度越大。类似,当样本 x_i 更贴近误差项 $\hat{y}_i - y_i < 0$ 时,表明该样本的预测值小于实际值,故对权重 n_i 赋予越大的数值,即对第二类误差项(预测值<实际值)的惩罚力度越大,其次,样本信息可以反映各样本之间的相关关系,会对预测结果产生一定影响^[17]。隶属度计算方法如下:

令 Δd^+ 为误差项 $\hat{y}_i - y_i > 0$ 的样本中心点, Δd^- 为误差项 $\hat{y}_i - y_i < 0$ 的样本中心点,则隶属度计算方法如下:

$$m_i = \frac{1}{\|x_i - \Delta d^+\| \log(\|x_i - \Delta d^+\|)} \quad (11)$$

$$n_i = \frac{1}{\|x_i - \Delta d^-\| \log(\|x_i - \Delta d^-\|)}$$

4 实证分析

4.1 数据的选取

选择沪深300成份股的股票日数据进行股票收盘价的实证分析,相应时间序列的日期为2008年1月1日至2019年12月31日,数据来源于东方财富数据库。沪深300成份股共包括300只股票,故该文总共选取300只股票作为总样本,并从中随机选取30只股票进行股票收盘价预测。所有的实证均在同一系统环境下运行,系统运行环境为PC(CPU 2.60 GHz, 4.00 GB RAM),操作系统为Windows 10,仿真软件为Matlab R2016b。此外,采用libsvm处理SVR。并将样本数据集分为训练样本数据集和测试样本数据集,选择样本数据的前80%作为训练样本,后20%作为测试样本。

沪深300成份股是沪深证券交易所于2005年4月8日联合发布的反映A股市场整体走势的指数,由

上海和深圳证券交易所中选取300只A股作为样本,其中沪市有179只,深市121只,综合反映深交所上市A股的股价走势。它具有业绩优于整体、对家方案较优、股改行情明显等优势,经常被研究者用于股价预测。

为了更好地反映所提出的基于信噪比的FPSVR股价预测模型的有效性,将所选取的股票时间序列按照股市的波动情况(牛市、震荡市和熊市)分为三个阶段,进行阶段式预测。为了更清晰地展示股市状况,以招商银行为例,对阶段式预测进行说明,如图1所示。

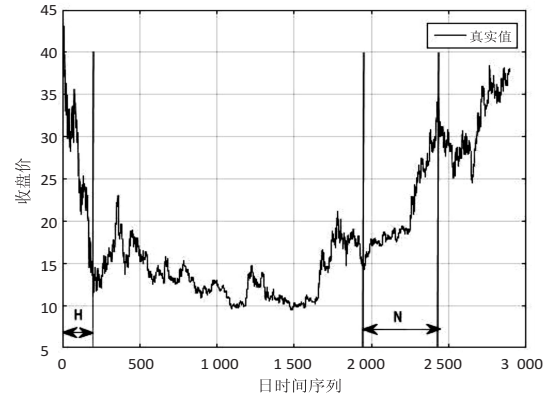


图1 招商银行2008-2019年日收盘价的变化趋势

从图1可看出,招商银行的收盘价变化趋势可分为三个阶段,其中H代表熊市阶段,为阶段1;N代表牛市阶段,为阶段2;其余时间的指数变化趋势大致代表震荡市阶段,为阶段3。

本研究为确保所有的输入特征位于相同参数范围内,以防止大范围的输入特征压倒其他输入特征,采用式(12)对数据进行标准化处理:

$$\tilde{x}_{ij} = \frac{x_{ij} - x_{\min j}}{x_{\max j} - x_{\min j}}, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, d \quad (12)$$

其中, x_{ij} 为第 i 个样本的第 j 个输入特征值, $x_{\min j}$ 为样本数据第 j 个输入特征的最小值, $x_{\max j}$ 为样本数据第 j 个输入特征的最大值。

4.2 相关性分析

Spearman 相关系数不但可以衡量存在非线性关系的相关变量之间的相关程度,而且未对数据有严格的假设要求,故本节采用 Spearman 相关系数进行输入特征变量与股票收盘价之间的相关性分析。为了更清晰地说明特征变量与股票收盘价之间的相关性,进一步地以招商银行为例展示输入特征与股票收盘价的相关性分析结果,如表1所示。

由表1可看出:

(1) 在置信度为0.01时,开盘价、最高价、最低价、MA和EMA都与股票收盘价之间的Spearman相关系数 > 0.95 ,说明这些指标都与股票收盘价存在很强的正相关性,这些输入指标数值增加时,股票收盘价也增加。

表 1 各指标相关性分析

特征变量	开盘价	最高价	最低价	总交易量	MA	EMA	MACD	MTM	RSI	AR	BR	VR	DK	ATR	SNR	收盘价
开盘价	1															
最高价	0.999 **	1														
最低价	0.999 **	0.999 **	1													
SNR	-0.131 **	-0.117 **	-0.138 **	1												
总交易量	0.997 **	0.996 **	0.996 **	-0.142 **	1											
MA	0.998 **	0.997 **	0.997 **	-0.140 **	1.000 **	1										
EMA	-0.006	0	0	0.063 **	-0.068 **	-0.056 **	1									
MACD	0.050 **	0.057 **	0.061 **	0.160 **	-0.016	-0.002	0.836 **	1								
MTM	0.106 **	0.114 **	0.114 **	0.154 **	0.052 **	0.064 **	0.701 **	0.817 **	1							
RSI	0.034	0.043 **	0.041 **	0.108 **	-0.002	0.005	0.550 **	0.584 **	0.692 **	1						
AR	-0.003	-0.002	-0.003	0.004	-0.003	-0.003	-0.026	-0.045 **	-0.014	-0.018	1					
BR	-0.006	-0.002	-0.002	0.084 **	-0.034	-0.028	0.360 **	0.416 **	0.554 **	0.511 **	-0.007	1				
VR	-0.02	-0.008	-0.011	0.091 **	-0.012	-0.011	0.049 **	0.128 **	0.149 **	0.097 **	-0.02	0.103 **	1			
DK	0.662 **	0.673 **	0.649 **	0.195 **	0.672 **	0.672 **	-0.063 **	-0.109 **	0.002	-0.004	0.035	-0.052 **	0.014	1		
ATR	-0.609 **	-0.611 **	-0.608 **	0.077 **	-0.610 **	-0.610 **	0.016	-0.008	-0.061 **	-0.057 **	-0.007	-0.008	-0.01	-0.455 **	1	
收盘价	0.999 **	0.999 **	0.999 **	-0.125 **	0.996 **	0.997 **	0.001	0.064 **	0.118 **	0.045 **	-0.003	0	0.009	0.661 **	0.610 **	1

注: * *. 在置信度(双测)为 0.01 时,相关性是显著的;*. 在置信度(双测)为 0.05 时,相关性是显著的。

(2) 在置信度为 0.01 时, SNR 与股票收盘价的 Spearman 相关系数 > 0.6, 说明 SVR 与股票收盘价之间存在中等程度的正相关性, 当 SNR 增大时, 股票收盘价随之增大。

(3) 不管是在置信度为 0.01 时, 还是在置信度为 0.05 时, MACD、BR、VR 和 OBV 都与股票收盘价存在非相关关系, 但是这些指标与其他指标之间存在相关关系, 说明这些指标可以影响其他指标, 从而间接影响股票收盘价的变化。

从上述发现中可得到以下结论: ① SNR 与预测变量股票收盘价之间存在直接的正相关性, 进而说明所构建的 SNR 特征变量是有效的; ② 这些输入特征之间存在紧密的相关性, 且都与预测变量股票收盘价之间存在直接或间接的相关关系。

4.3 评估准则

众所周知, 误差是指测量值与真实值或实际值之间的差值, 主要用来衡量测量结果的准确度, 其中平均绝对百分比误差因能避免误差相互抵消的问题, 可以准确反映实际预测误差的大小, 均方根误差不仅能对一组测量中的特大或特小误差反映非常敏感, 而且能够很好地反映出测量的精密度, 故经常被用于预测模型的准确度测量。该文将平均绝对百分比误差 (MAPE) 和均方根误差 (RMSE) 作为股价预测模型的评估准则, 其表达式如下所示:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_{i,\text{close}} - \hat{x}_{i,\text{close}}}{x_{i,\text{close}}} \right| \times 100 \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{i,\text{close}} - \hat{x}_{i,\text{close}})^2} \quad (14)$$

其中, N 为样本数量, $x_{i,\text{close}}$ 为第 i 日的实际股票收盘价, $\hat{x}_{i,\text{close}}$ 为所预测的股票收盘价。

考虑到上述实验条件, 对于所有 RMSE 和 MAPE, 它们的值越小, 所构建的股票收盘价预测模型的预测性能越好。

4.4 结果分析

为了验证所提出的基于信噪比的 FPSVR 股价预测模型的有效性, 在收盘价预测的过程中, 选用三种基准模型预测方法, 分别为模型 1: 未加入信噪比的支持向量回归 (未加入 SNR 的 SVR); 模型 2: 加入信噪比的支持向量回归 (加入 SNR 的 SVR); 模型 3: 未加入信噪比的 FPSVR, 并将这三种预测方法与文中所提的模型 4—基于信噪比的 FPSVR 模型进行对比分析。

由于金融时间序列的动态特性是非线性的, 文中将 FPSVR 模型应用于收盘价预测时, 使用高斯核函数作为核函数, 因为高斯核函数在一般的平滑假设下往往会有很好的性能^[27]。此外, 本节采用基于十折交叉验证的网格搜索方法, 对传统的支持向量回归参数进行了选择, 并对 FWSVR 模型也使用同样的参数, 所使用的参数如表 2 所示。

表 2 基于信噪比的 FPSVR 模型所使用的最优参数

模型	阶段	C	γ	K
FPSVR	阶段 1	$[2^{-10}, 2^{10}]$	$[2^{-10}, 2^{10}]$	10
	阶段 2	$[2^{-10}, 2^{10}]$	$[2^{-10}, 2^{10}]$	10
	阶段 3	$[2^{-10}, 2^{10}]$	$[2^{-10}, 2^{10}]$	10

此外, 采用遗传算法 (GA) 对 FPSVR 股价预测模型进行求解, GA 中所使用的相关参数为 $G = 2\ 000$,

popsize = 50, 其中 G 为进化代数, popsize 为种群规模, 相应的适应度函数变化情况如图 2 所示。

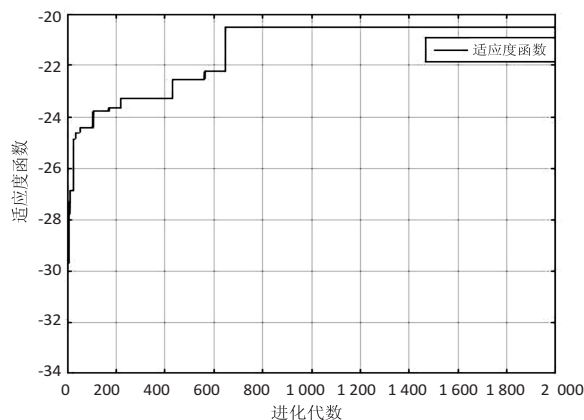


图2 GA 的适应度函数的变化情况

从图 2 中可看出, 在进化代数为 2 000 时, 随着 GA 算法的进化代数增加, 适应度函数趋于稳定, 从而说明这种情况下, GA 算法可以很好地对 FPSVR 进行求解。

下面根据上述所建立的四种不同模型对沪深 300

表3 模型对比分析结果

预测 误差	模型 1			模型 2			模型 3			模型 4		
	阶段 1	阶段 2	阶段 3	阶段 1	阶段 2	阶段 3	阶段 1	阶段 2	阶段 3	阶段 1	阶段 2	阶段 3
MAPE	0.036 3	0.571 2	0.001 3	0.035 4	0.508 5	0.001 0	0.014 7	0.105 7	0.001 0	0.011 6	0.094 3	0.000 9
RMSE	4.673 2	23.798 2	0.567 1	4.482 3	21.346 8	0.537 4	1.645 7	3.256 9	0.524 8	1.048 2	2.696 3	0.514 2

5 结束语

为了更好地实现股价的精准预测, 提出了基于信噪比的模糊近似支持向量回归 (FPSVR) 的股价预测模型, 并通过实证研究、相关性分析和对比分析三个方面实现和验证所提模型的准确性。研究表明, 与现有模型相比, 该模型不仅在震荡期, 而且在牛市和熊市期均可实现股价的精准预测。该模型通过对沪深成份股 2008 年至 2019 年的股票日数据进行研究和分析, 实现对股票的准确预测, 为投资者提供一种更有效的预测方法, 从而有助于投资者的投资决策。

参考文献:

- [1] MANASSEH C O, CHUKWU N O, ABADA F C, et al. Interactions between stock prices and exchange rates: an application of multivariate VAR-GARCH model[J]. Cogent Economics & Finance, 2019, 7(1): 1681573.
- [2] KIM H Y, WON C H. Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models[J]. Expert Systems with Applications, 2018, 103: 25-37.
- [3] 张贵生, 张信东. 基于微分信息的 ARMAD-GARCH 股价

成份股中随机选择的 30 只股票时间序列进行收盘价预测, 并给出了四种模型在不同阶段的预测误差, 如表 3 所示。

从上述实验结果中可以看出:

(1) 与模型 1 和模型 2 相比, 模型 3 和模型 4 的预测误差更低, 说明加入模糊隶属度和双边权重的 FPSVR 模型可以更好地实现股价预测。

(2) 模型 2 和模型 4 的预测误差分别低于模型 1 和模型 3 的预测误差, 从而表明加入信噪比特征变量后的股价预测模型准确度更高。

(3) 分阶段来看, 不管是 SVR 股价预测模型还是 FPSVR 的股价预测模型, 在阶段 1 和阶段 2 的预测误差均高于阶段 3 的预测误差, 说明在震荡时期, 这两种模型更适用于股价预测。

(4) 分阶段来看, 在阶段 1 和阶段 2 时期, FPSVR 的股价预测模型的预测误差要远低于 SVR 股价预测模型的预测误差, 从而表明, 该文所构建的 FPSVR 模型可以弥补 SVR 模型在阶段 1 和阶段 2 时期的股价预测误差较大的不足, 可以更好地实现股价预测。

预测模型[J]. 系统工程理论与实践, 2016, 36(5): 1136-1145.

- [4] RUBIO A, BERMUDEZ J D, VERCHER E. Forecasting portfolio returns using weighted fuzzy time series methods[J]. International Journal of Approximate Reasoning, 2016, 75: 1-12.
- [5] 刘智, 张铁, 董莹, 等. 关于股票价格的二阶模糊时间序列[J]. 东北大学学报: 自然科学版, 2019, 40(2): 300-304.
- [6] KHASHEI M, HAJIRAHIMI Z. A comparative study of series arima/mlp hybrid models for stock price forecasting[J]. Communications in Statistics-Simulation and Computation, 2019, 48(9): 2625-2640.
- [7] ONOH J O, PETER G. Stock market performance of firms in the nigerian petroleum sector using the ARIMA model approach[J]. World Journal of Finance and Investment Research, 2019, 4(1): 1-19.
- [8] STOEAN R, STOEAN C, SANDITA A. Evolutionary regressor selection in ARIMA model for stock price time series forecasting[C]//International conference on intelligent decision technologies. [s.l.]: Springer, 2017: 117-126.
- [9] YU P, YAN X. Stock price prediction based on deep neural networks[J]. Neural Computing and Applications, 2020, 32(6): 1609-1628.

- [10] GÖÇKEN M, ÖZÇALICI M, BORU A, et al. Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection [J]. *Neural Computing and Applications*, 2019, 31(2): 577–592.
- [11] MOGHADDAM A H, MOGHADDAM M H, ESFANDYARI M. Stock market index prediction using artificial neural network [J]. *Journal of Economics, Finance and Administrative Science*, 2016, 21(41): 89–93.
- [12] HÁJEK P. Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns [J]. *Neural Computing and Applications*, 2018, 29(7): 343–358.
- [13] 綦方中, 林少倩, 俞婷婷. 基于 PCA 和 IFOA-BP 神经网络的股价预测模型 [J]. *计算机应用与软件*, 2020, 37(1): 116–121.
- [14] 孟 叶, 于忠清, 周 强. 基于集成学习的股票指数预测方法 [J]. *现代电子技术*, 2019, 42(19): 115–118.
- [15] CORTES C, VAPNIK V. Support-vector networks [J]. *Machine Learning*, 1995, 20(3): 273–297.
- [16] NAYAK R K, MISHRA D, RATH A K. A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices [J]. *Applied Soft Computing*, 2015, 35(10): 670–680.
- [17] CHEN Y, HAO Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction [J]. *Expert Systems with Applications*, 2017, 80(9): 340–355.
- [18] XIAO J, ZHU X, HUANG C, et al. A new approach for stock price analysis and prediction based on SSA and SVM [J]. *International Journal of Information Technology & Decision Making*, 2019, 18(1): 287–310.
- [19] LEE M. Using support vector machine with a hybrid feature selection method to the stock trend prediction [J]. *Expert Systems With Applications*, 2009, 36(8): 10896–10904.
- [20] FUNG G, MANGASARIAN O L. Proximal support vector machine classifiers [C]//*Proceedings of the seventh ACM SIGKDD international conference of knowledge discovery and data mining*. San Francisco, California, USA: IEEE, 2001: 77–86.
- [21] 姚 潇, 余乐安. 模糊近似支持向量机模型及其在信用风险评估模型中的应用 [J]. *系统工程理论与实践*, 2012, 32(3): 549–554.
- [22] 张贵生, 张信东. 基于近邻互信息的 SVM-GARCH 股票价格预测模型研究 [J]. *中国管理科学*, 2016, 24(9): 11–20.
- [23] 张 冰, 王传美, 贺素香. 改进的 TSVR 模型在股市高频数据上的预测 [J]. *计算机工程与设计*, 2019, 40(11): 3241–3246.
- [24] BANDI F M, RUSSELL J R. Separating micro-structure noise from volatility [J]. *Journal of Financial Economics*, 2006, 79(3): 655–692.
- [25] SIEGLE A F, TRAPP O. Improving the signal-to-noise ratio in gel permeation chromatography by Hadamard encoding [J]. *Journal of Chromatography A*, 2016, 1448: 93–97.
- [26] WANG J, WANG Q, TIAN L, et al. A low-noise, high-SNR balanced homodyne detector for the bright squeezed state measurement in 1–100 kHz range [J]. *Chinese Physics B*, 2020, 29(3): 34205.
- [27] OZTEKIN A, KIZILASLAN R, FREUND S, et al. A data analytic approach to forecasting daily stock returns in an emerging market [J]. *European Journal of Operational Research*, 2016, 253(3): 697–710.