

Stock Price Forecast Based on SVM-KNN

Peiqi Zhang, Haijun Liu, Dongxiao Pei, Xiyue Wang

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou Henan
Email: zhangpeiqi1217@163.com

Received: Nov. 12th, 2019; accepted: Nov. 25th, 2019; published: Dec. 2nd, 2019

Abstract

In this paper, the support vector machine (SVM) and K-nearest neighbor (KNN) algorithm were used to study the stock price forecasting problem, select the transaction data reflecting the stock changes and its technical indicators, including volume, closing price, highest price, moving average (MA), etc., forecasting the ups and downs and the closing price of the Shanghai Composite Index. Firstly, the SVM was used to predict the training set's ups and downs. Then the training set used KNN to predict the short-term (1 day), medium-term (7 days) and long-term (30 days) prices of the stock, thus forming a forecast model based on transaction data and technical indicators. Finally, the MAPE and RMSE of this model were obtained. In order to verify the validity of the model, a new investment strategy was constructed based on the model prediction results, and the real data was used for investment, and the large-cap stock index was subjected to a one-month simulation investment.

Keywords

Support Vector Machine, K-Nearest Neighbor, Stock Forecast, Technical Indicators, Simulation Investment

基于SVM-KNN的股票价格预测

张佩琪, 刘海军, 裴冬晓, 王喜月

郑州大学, 数学与统计学院, 河南 郑州
Email: zhangpeiqi1217@163.com

收稿日期: 2019年11月12日; 录用日期: 2019年11月25日; 发布日期: 2019年12月2日

摘要

本文利用支持向量机(Support Vector Machine, SVM)和K近邻(K-Nearest Neighbor, KNN)结合的算法研究股票价格的预测问题。选取反映股票变化的交易数据及其技术指标, 包括成交量、收盘价、最高价、

移动平均(MA)等, 对上证综指进行涨跌趋势及收盘价格指数的预测。首先利用SVM对训练集进行预测, 预测涨跌趋势, 其次KNN对股票短期(1天), 中期(7天)和长期(30天)的价格进行预测, 由此形成基于交易数据和技术指标的预测模型。最后得出此模型的综合平均相对误差(MAPE)和均方根误差(RMSE)。为了验证模型的有效性, 根据模型预测结果构建新的投资策略, 并利用真实数据进行投资, 对大盘股指进行为期一个月的模拟投资。

关键词

支持向量机, k近邻, 股票预测, 技术指标, 模拟投资

Copyright © 2019 by author(s) 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

1. 引言

在市场经济中, 股票市场被认为是经济发展的晴雨表。并且, 它对政治混乱、恐怖袭击和投资者心理等非经济因素最为敏感[1]。股票市场的发展受到社会经济活动发展的影响, 目前的研究主要基于复杂理论的发展。在大多数情况下, 金融数据伴随着噪声、非线性和混沌的性质, 使得数据在每个阶段会有不同的表现形式。因此, 常用的统计方法中, 例如移动平均、加权平均、卡尔曼滤波、指数平滑法、回归分析、自回归移动平均(ARMA)、集成移动平均(ARIMA), 就无法很好地对股票市场的指数进行预测。近来, 与传统的统计方法相比, 人工智能可以很好地解决股票场所具有噪声、非线性和混沌的性质, 因此本文尝试使用人工智能中的 SVM 和 KNN 方法对股票价格进行分类预测。

由于人们获取股票信息的渠道有限, 常根据股票历史数据趋势结合技术指标进行涨跌预测, 技术指标(移动平均 MA、趋势指标 MACD、指数移动平均 EMA、动量 MTM 等)具有统计和经济上显著的样本预测能力。此外, 技术指标和宏观经济变量在商业周期上提供了互补的信息; 技术指标能更好地探测到商业周期峰值附近股票风险溢价的典型下降, 而宏观经济变量更容易捕捉到周期低谷附近股票风险溢价的典型上升。随着机器学习、深度学习、数据挖掘领域的发展, 越来越多的研究者利用股票市场的数据和机器学习的方法相结合进行预测[2] [3] [4] [5], 从而更准确地把握股票涨跌趋势。研究表明这种方法具有较高的预测精度, 并且通过多种技术指标对数据进行分析, 具有良好的预测效果。

近年来, 国内外许多学者利用机器学习、人工智能方法对股票市场进行了深入研究。神经网络[6] [7]、支持向量机[8]、遗传因子、粒子群优化等方法被用于股票价格预测[9]。由于中国股票市场是一个新兴市场, 有更大的不确定性。为了更准确地把握股票市场规律, 已有学者应用了 K-means [10]、神经网络[11] [12]及其优化模式、深度学习[13]等方法研究股票深层规律, 对股票价格趋势进行拟合, 得到较好的效果。Dwiarso [14]利用基于神经网络反向传播的梯度下降算法提高股票价格预测的准确性。Dai [15]提出了将股票指数进行粒子模糊化从而利用 SVM 对之后的五个交易日的数据进行预测。因此, 目前对于股票市场的研究是采用机器学习、深度学习等人工智能方法。由于股票市场的波动性较大, 要实现较高的预测准确度有一定的难度, 并且随着方法复杂程度的提高, 容易出现过拟合的情况, 且算法运行速度较慢, 难以寻找较优的参数。

虽然研究者对股票市场规律进行了大量研究, 但真正能够揭示股票市场深层次特征的、较准确预测股票价格的并不多。因此, 为了克服以往研究的局限性, 本文提出了一种新的 SVM-KNN 算法, 利用 SVM

[16]良好的分类效果来准确预测股票涨跌趋势，并利用 KNN [17]算法对股票涨幅进行预测，达到较好的预测股票价格趋势的目的。实验表明具有较准确的预测效果和预测精度。本文结构如下：第二部分对数据进行基本处理，第三部分对 SVM 和 KNN 算法的理论进行说明，第四部分为实证分析，第五部分进行模拟投资，第六部分为总结。

2. 数据与指标

2.1. 数据来源

本文选用的原始数据来自于锐思数据库，选择上证综指(上海证券综合指数简称“上证指数”或“上证综指”，其样本股是在上海证券交易所全部上市股票，包括 A 股和 B 股，反映了上海证券交易所上市股票价格的变动情况)从 2008 年 1 月 2 日到 2014 年 12 月 31 日的日交易数据，包括有：开盘价、最高价、最低价、收盘价、成交量。

2.2. 数据整理

本文用到的数据指标包括前一天收盘价、开盘价、最高价、最低价、收盘价、后一天收盘价、成交量。令 $U = \langle U_{prec}, U_o, U_h, U_l, U_c, U_{nexc}, U_{volume} \rangle$ ，其中 U_p 代表前一天收盘价； U_o 代表开盘价； U_h 代表最高价； U_l 代表最低价； U_c 代表收盘价； U_n 代表后一天收盘价； U_v 代表成交量； U_i 代表每一天的数据，如公式(1)所示：

$$U = \begin{bmatrix} u_{1p} & u_{1o} & u_{1h} & u_{1l} & u_{1c} & u_{1n} & u_{1v} \\ u_{2p} & u_{2o} & u_{2h} & u_{2l} & u_{2c} & u_{2n} & u_{2v} \\ u_{3p} & u_{3o} & u_{3h} & u_{3l} & u_{3c} & u_{3n} & u_{3v} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{np} & u_{no} & u_{nh} & u_{nl} & u_{nc} & u_{nn} & u_{nv} \end{bmatrix} \quad (1)$$

因为所得到的数据中不包括类标签，所以为了预测每一天的价格涨跌情况，本文引入了一个新的属性来识别类标签，即收盘价的变化 y_i ，其中 y_i 代表的是每一天股票 u_i 的涨跌情况。计算公式如(2)所示：

$$\Delta c^i = u_{i+m,c} - u_{i,c}, \forall i = 1, 2, 3, \dots, n-1 \quad (2)$$

每天的涨跌情况 Δc^i 是以第 $i+m$ 天和第 i 天的收盘价之差作为第 i 天的收益情况。如果 $\Delta c^i > 0$ 则表示有盈利，反之，代表亏损。故 Δc^i 代表了股票的走势方向，对盈利和亏损两种状况利用标签进行表示，如公式(3)所示

$$y_i = \begin{cases} -1, & \Delta c^i < 0 \\ 1, & \Delta c^i \geq 0 \end{cases} \quad (3)$$

收集到的所有的数据都要根据公式(4)进行归一化处理。

$$u_{ij} = \frac{u_{ij} - u_{\min j}}{u_{\max j} - u_{\min j}} \quad (4)$$

其中， u_{ij} 代表第 j 天第 i 个指标的值， $u_{\max j}$ 和 $u_{\min j}$ 代表第 j 天所有指标的最大值和最小值。

2.3. 技术分析指标

HSU [18]指出，在机器学习以前，金融经济学家一般更喜欢使用计量经济学，选择技术指标利用 EMH 模型来对股市进行分析。而技术指标依赖于过去的价格和成交量确定未来的价格趋势。现有的研究一般

是基于多种技术指标构建交易策略提高盈利能力,包括过滤器规则[19]、移动平均线[20]、动量[21]和自动模式识别[22]。因此,本文在收集到的原始数据的基础上加入一些重要的技术指标构成最终的特征向量。加入的技术指标如表一所示:

根据得到的技术指标,将已有的向量 U 进行扩充,最为最终的输入向量。

$$U = \begin{bmatrix} u_{1p} & u_{1o} & u_{1h} & u_{1l} & u_{1c} & u_{1n} & u_{1v} & T_{11} & T_{12} & T_{13} & T_{14} & T_{15} & c_1 \\ u_{2p} & u_{2o} & u_{2h} & u_{2l} & u_{2c} & u_{2n} & u_{2v} & T_{21} & T_{22} & T_{23} & T_{24} & T_{25} & c_2 \\ u_{3p} & u_{3o} & u_{3h} & u_{3l} & u_{3c} & u_{3n} & u_{3v} & T_{31} & T_{32} & T_{33} & T_{34} & T_{35} & c_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{np} & u_{no} & u_{nh} & u_{nl} & u_{nc} & u_{nn} & u_{nv} & T_{n1} & T_{n2} & T_{n3} & T_{n4} & T_{n5} & c_n \end{bmatrix} \quad (5)$$

2.4. 误差分析

为了对模型的预测结果进行说明,本文选择用 MAPE 和 RMSE 来评价模型的预测能力。

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{u_{i,\text{close}} - \hat{u}_{i,\text{close}}}{u_{i,\text{close}}} \right| * 100 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (u_{i,\text{close}} - \hat{u}_{i,\text{close}})^2} \quad (7)$$

其中, $u_{i,\text{close}}$ 表示真实的收盘价, $\hat{u}_{i,\text{close}}$ 表示预测的收盘价, N 代表测试数据的个数。

3. 基于 SVM-KNN 的预测模型

股票的涨跌和很多因素有关系,可以将其作为一种非线性、不确定的系统,本文利用从历史数据中获得信息,以此对股票涨跌趋势及价格进行预测。主要的思路:利用 SVM 对训练数据进行预测,并根据涨跌标签形成新的训练集,其次对测试数据利用 SVM 预测涨跌趋势,并对比标签一致的训练集,随后对每一个测试数据和其对应的训练集使用 KNN 对股票的价格进行预测,由此形成基于交易数据和技术指标的预测模型。

3.1. 支持向量机分类算法

支持向量机作为一种“监督性”的分类方法,是通过找到其超平面确定支持向量而对数据进行分类。[23] [24]在线性可分的情况下,可以通过计算找到其最优超平面,如图 1 所示,其中圆形点和四边形点代表分类样本, L_1 和 L_2 代表最优超平面,二者之间的距离 r 称之为间隔(Margin)。

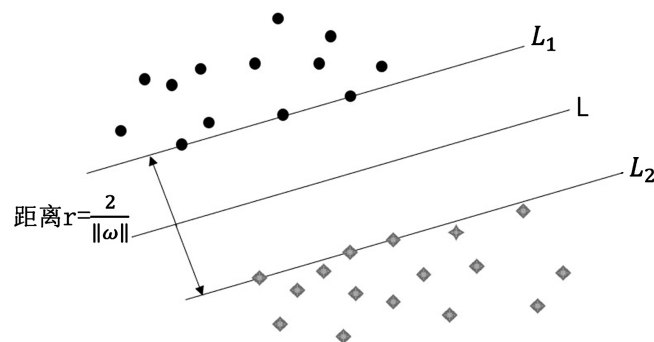


Figure 1. Linear plane separable hyperplane

图 1. 线性可分情况的超平面

对于非线性的情况,要引入核函数,将输入的数据变换到高维空间中去,在新形成的空间中,计算其最优超平面,求得支持向量机。在本文中,将数据分为训练数据和测试数据,对训练数据 (u_i, y_i) , $i=1,2,3,\dots,n-1$,利用支持向量机进行分类时,首先需要选择合适的核函数,其次利用 SMO 算法求得最优超平面和支持向量机,最后利用测试数据验证模型的准确度。

支持向量需要优化的问题目标的一般表达式如公式(7)所示:

$$\begin{aligned} \min \Phi(x) &= \frac{\|w\|^2}{2} + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad &\begin{cases} y_i (w^T x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, 3, \dots, n \end{cases} \end{aligned} \quad (7)$$

其中, C 为惩罚参数, ξ_i 为松弛变量。松弛变量说明了此分类器可以容忍分类错误的程度,当 C 越大时, ξ_i 被压制,经验风险小,表示重视分类误差。

最终的对偶规划问题如公式(8)所示:

$$\begin{aligned} \max Q(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \\ \text{s.t.} \quad &\begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, i = 1, 2, 3, \dots, n \end{cases} \end{aligned} \quad (8)$$

其中, $\langle \phi(x_i), \phi(x_j) \rangle$ 代表变换后的内积空间。分类超平面为:

$$f(x) = \text{sgn} \left\{ \left(w^T \cdot \phi(x) \right) \right\} + b = \text{sgn} \left\{ \sum_{\text{支持向量}} \alpha_i^* y_i \langle \phi(x_i), \phi(x) \rangle + b^* \right\} \quad (9)$$

本文中,选择的核函数为径向基核函数(RBF)是:

$$K(x_i, x_j) = \exp \left(-\gamma \|x_i - x_j\|^2 \right), \gamma > 0 \quad (10)$$

支持向量机的分类性能受核函数机及其参数、问题的复杂程度,及其分类面附近的噪音点的影响。

3.2. K 近邻分类算法

KNN (K-Nearest-Neighbor)分类器是通过给定一个待识别的样本集,根据在训练集中寻找到最近的 K 个近邻,通过确定 K 的个数,从而将待识别样本进行归类。图 2 中给出的是在 $k=5$ 的条件下的分类状况。

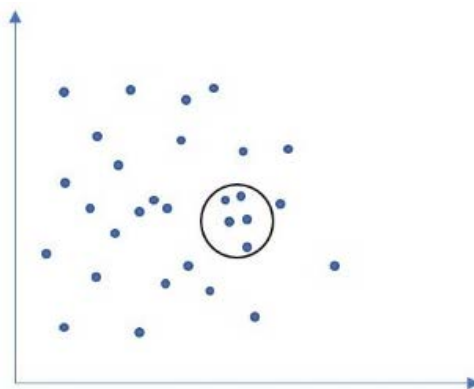


Figure 2. Classification when $K = 5$

图 2. $K = 5$ 时的分类情况

3.3. SVM-KNN 算法实现

本文所用算法的基本思路如图 3 所示，本算法共分为三个阶段。第一阶段，对收集到的数据进行预处理，依据表 1 加入技术指标，对数据进行描述，归一化处理后形成特征矩阵，并选择 80% 的数据作为训练数据，20% 的数据作为测试数据；第二阶段，利用支持向量机对模型进行训练；第三阶段，对训练数据利用支持向量机得到分类指标，+1 代表上涨，-1 代表下跌，根据涨跌趋势，将训练数据集分为上涨集 N_1 和下跌集 N_2 ，接着对每一个测试数据利用 SVM 得到分类指标 P ，并将分类指标 P 和新生成的上涨集 N_1 和下跌集 N_2 进行对比，选择指标一致的训练集 N_i ($i = 1, 2$)，利用 k 近邻对 N_i ($i = 1, 2$) 中的 Δc 计算。最终结果采用前一天收盘价加上预测的 Δc 作为当天收盘价输出。

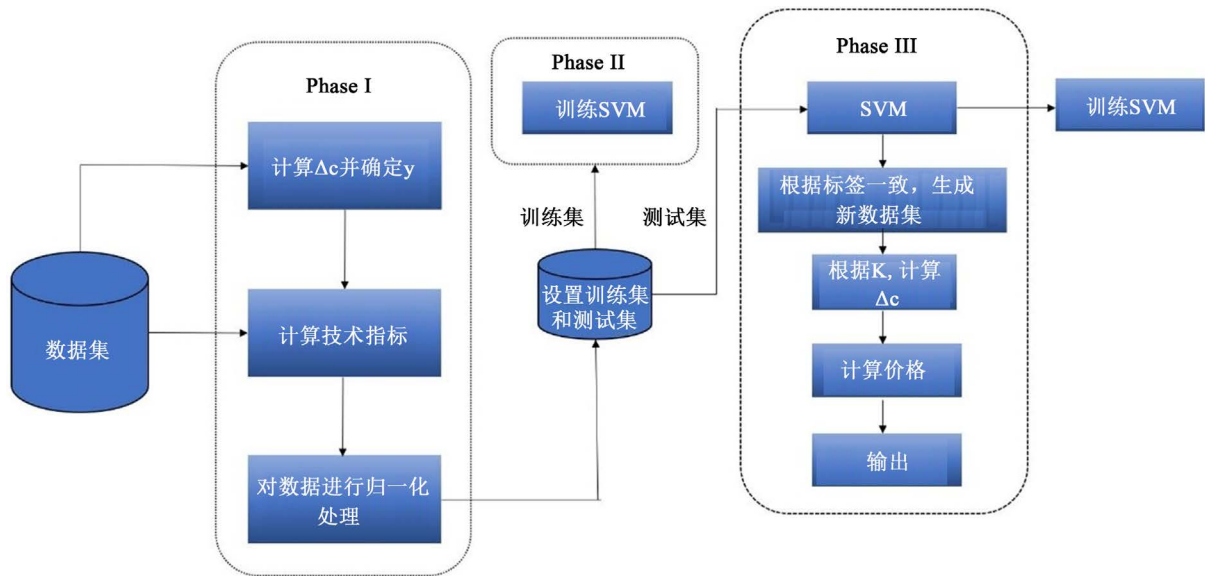


Figure 3. Algorithm flowchart
图 3. 算法流程图

Table 1. Technical index calculation formula
表 1. 技术指标计算公式

技术指标	公式
移动平均线(MA)	$MA(N) = \frac{1}{N} \sum_{t=1}^N u_{t,close}$
指数移动平均值(EMA)	$EMA(N) = \begin{cases} u_{i,close} & N = 1 \\ \frac{2}{N+1} * u_{i,close} + \frac{N-1}{N+1} * EMA(N-A) & N > 1 \end{cases}$
异同移动平均线(MACD)	$DIF(i) = EMA(N_{fast}) - EMA(N_{slow})$ $DEA(i) = \delta * DEA(i-1) + (1-\delta) * DIF(i)$ $MACD(i) = 2 * (DIF(i) - DEA(i))$
相对强弱指数(RSI)	$RSI(N) = 100 - \frac{100}{1 + \frac{EMA(N)_{up}}{EMA(N)_{down}}}$
动量指标(MTM)	$MTM(i, N) = u_{i,close} - u_{i-N,close}$

算法的运行流程如下：

SVM-KNN 算法用于股票价格涨跌趋势分析

输入： U, C, γ, k

其中 U 代表前一天收盘价、开盘价、最高价、最低价、收盘价、后一天收盘价、成交量和技术指标组成的矩阵；
 n 是输入的股票个数， k 代表了 KNN 的参数

输出：预测的价格及趋势

1. 计算 Δc 和 y
2. 进行数据归一化利用公式(4)
3. 初始化正确分类的数据个数 $M \leftarrow 0$
4. 将数据进行按照一定比率，分为测试集和训练集
5. 对训练集利用 SVM 进行训练，得到参数 C, γ ；
6. 对训练集利用已训练好的 SVM 计算涨跌精确度，并输出；
7. 将 SVM 得到的涨跌趋势 P 和训练集原有的涨跌趋势进行对比；
8. 将训练数据集分为上涨集 N_1 和下跌集 N_2 ；
9. 对每一个测试数据利用 KNN 对新生成的训练集进行价格预测；
10. 输出价格 price

4. 实例验证与结果分析

根据所选取的上证综指，利用模型预测提前 1 天、1 周和 1 个月的收盘价变动。并且利用公式(4)进行归一化处理，处理后的数据如图 4 所示。按照所选取的从 2008 年 10 月 31 日到 2014 年 12 月 31 的数据，将其分为训练集以及测试集，该模型最终利用 1200 个数据进行训练，300 个数据进行预测，所使用的技术指标如图 4~7 所示。模型的训练程度根据 MAPE 和 RMSE 进行评估。

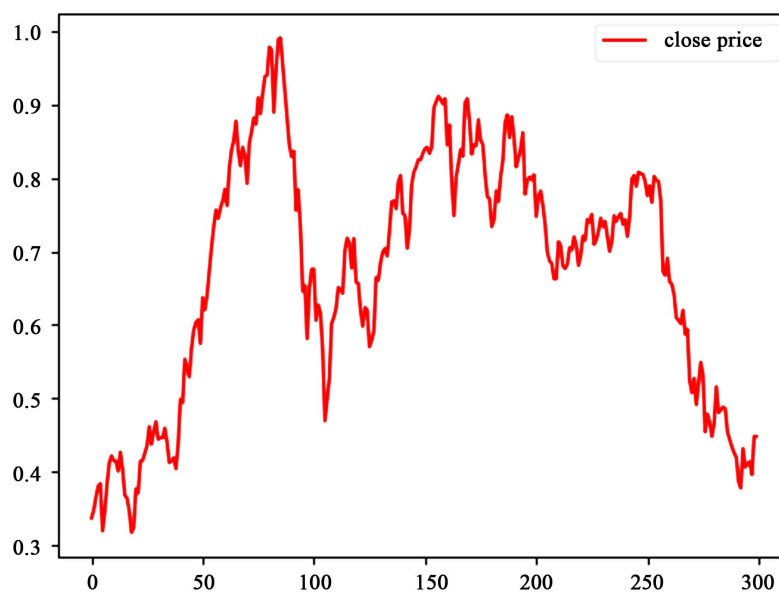


Figure 4. Normalized closing price

图 4. 归一化后的收盘价

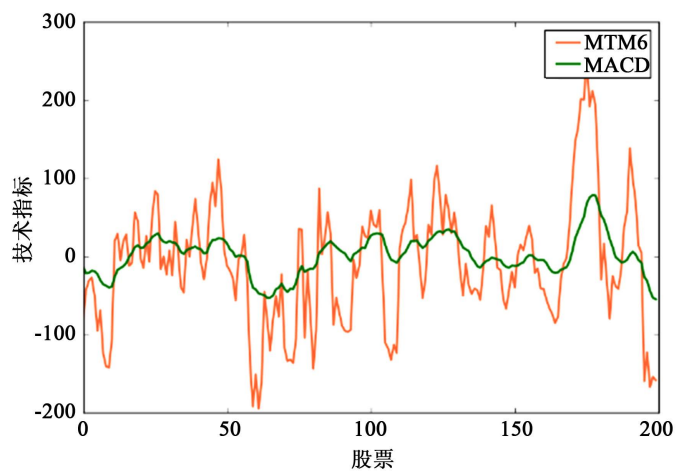


Figure 5. MTM and MACD indicators
图 5. MTM 和 MACD 指标

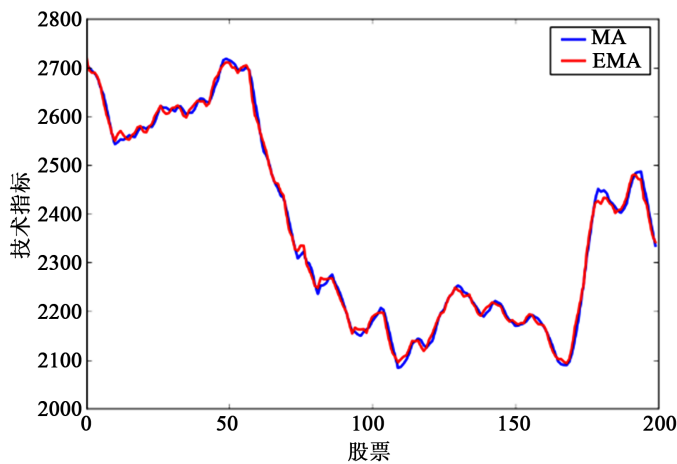


Figure 6. MA and EMA indicators
图 6. MA 和 EMA 指标

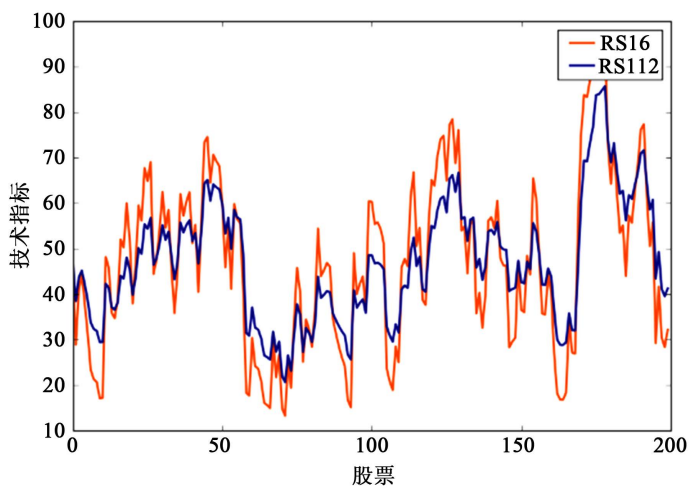


Figure 7. RSI indicators
图 7. RSI 指标

图 4 为所选股票的收盘价趋势，图 5~7 为实验所选取的技术指标特征。MA 的作用是消除股价随机波动的影响，寻找股价波动的平均趋势；EMA 是指确定时间窗口下的平均值，数据点越近权重就越大，数据点越小权重越小；MACD 是一种常见的技术分析工具，用于分析股票价格变化的强度、方向、能量以及趋势周期；RSI 包括了交叉、数值、形态和背离等多方面的判断原则，RSI 强弱指标的理论 and 实践极其适合于股票市场的短期投资；MTM 又称动量指标，通过观察股价波动的速度来揭示股票价格反转的规律变化。通过这些技术指标特征，可以对股票价格规律进行更准确的刻画，便于 SVM-KNN 更准确地预测股票涨跌趋势。

对于 SVM 分类器，对参数 C 和 γ 进行优化，并选择最优参数如表 2 所示，其中 C 为 SVM 的松弛因子， C 越小，模型的泛化能力就越强； γ 为所选用径向基核函数的参数， γ 值越小说明单个样本的影响力越大。对于 KNN，根据文献阅读最终选择 10 作为 K 值。

Table 2. Parameter selection of SVM

表 2. SVM 的参数选取

参数选取	时间长度		
	1 天	7 天	30 天
C	29	36	52
γ	0.2	0.7	0.7

表 3 说明了 SVM 对于短期、中期和长期的上证综指涨跌预测的情况。可以看出，SVM 对于股票价格 1 天，7 天，30 天的涨跌预测准确率均在 90% 以上。

Table 3. SVM's forecast of the ups and downs of the Shanghai Composite Index on different periods

表 3. SVM 对于不同周期上证综指的涨跌预测情况

	时间长度		
	1 天	7 天	30 天
上证综指	0.96	0.92	0.91

图 8~10 所表示的为根据 SVM-KNN 对于 1 天、7 天和 30 天的价格预测情况。可以看出，股票价格趋势拟合良好，所提算法能很好的预测股票价格趋势。表 4 是根据 MAPE 和 RMSE 计算的预测误差。

Table 4. MAPE and RMSE calculation errors

表 4. MAPE 和 RMSE 计算误差

		时间长度		
		1 天	7 天	30 天
上证综指	MAPE	0.20	0.23	0.32
	RMSE	0.0051	0.0057	0.0102

表 4 是对模型进行评估，MAPE 代表了平均绝对百分比误差，误差值越小，代表模型的拟合度越好，RMSE 代表了均方根误差，是均方误差的算术平均值，评价数据变化的程度，RMSE 的值越小，说明预测模型描述实验数据具有更好的精确度。

本文所提出的模型完全依赖于 SVM 的分类标签和 KNN 算法计算具有相似类别标签的 K 个邻域，将其均值与前一天的指数相加即可预测第二天的指数。该过程过滤了不需要参与预测的数据，从而降低了误差。

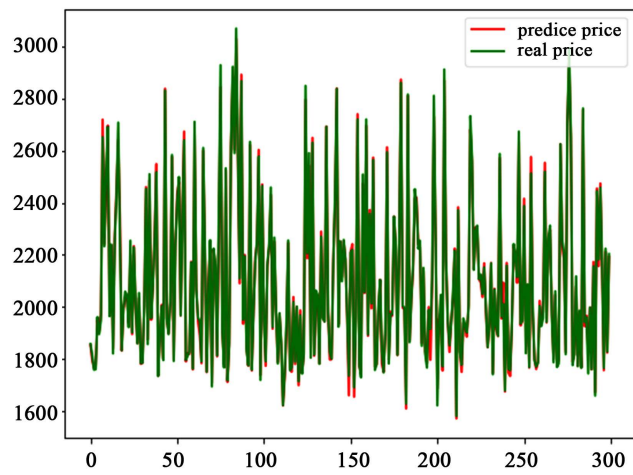


Figure 8. One-day ups and downs forecast
图 8. 一天的涨跌预测情况

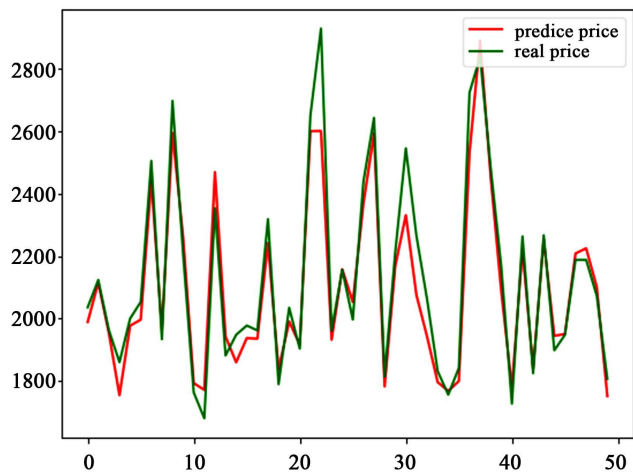


Figure 9. Seven-day ups and downs forecast
图 9. 七天的涨跌预测情况

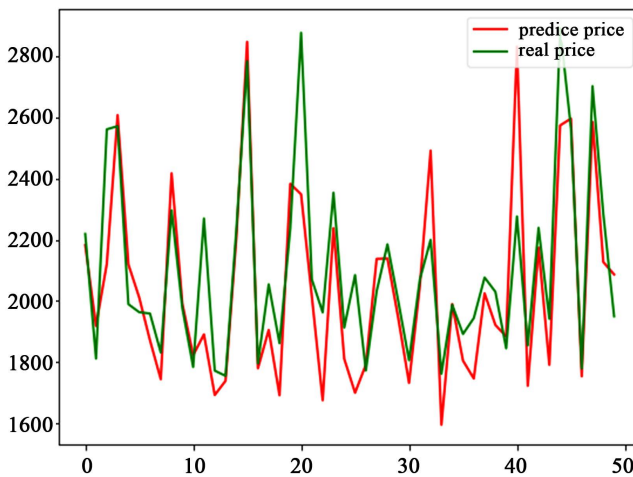


Figure 10. Thirty-day ups and downs forecast
图 10. 三十天的涨跌预测情况

5. 模拟投资

为了更加充分的验证模型的有效性, 本文利用对上证综指 2014 年 11 月 19 日到 2014 年 12 月 31 日的涨跌趋势进行预测, 根据涨跌结果根据其收盘价进行模拟投资。

5.1. 构建投资策略

首先, 根据 SVM 对上证综指 2014 年 11 月 20 日到 2014 年 12 月 31 日的涨跌趋势进行预测, 再利用 KNN 输出 30 天的收盘价。接着, 确定投资策略: 规定在 11 月 20 日对上证综指以当日收盘价买入一股, 随后观察涨跌趋势, 在“当日上涨, 明日下跌”的前一天卖出股票, 并在“当日下跌, 明日上午”的前一天再次买入股票。如表 5 所示, 买入的时间点应该为: 11 月 20 日、11 月 28 日、12 月 20 日和 12 月 22 日; 卖出的时间点应该为: 11 月 25 日、12 月 19 日、12 月 22 日和 12 月 23 日。最后, 根据中国股市的“T + 1 交收”交易规则, 本次买卖股票规定必须持有两天以上才可以卖出, 因此, 在模拟的 30 天内只进行两次交易。

Table 5. Thirty-day rise and fall
表 5. 30 天的涨跌情况

日期	涨跌	日期	涨跌	日期	涨跌
11/20	1	12/4	1	12/18	1
11/21	1	12/5	1	12/19	1
11/24	1	12/8	1	12/22	-1
11/25	1	12/9	1	12/23	1
11/26	-1	12/10	1	12/24	-1
11/27	-1	12/11	1	12/25	1
11/28	-1	12/12	1	12/26	-1
12/1	1	12/15	1	12/29	-1
12/2	1	12/16	1	12/30	-1
12/3	1	12/17	1	12/31	-1

5.2. 计算收益

图 11 为预测 30 天的价格走势图。由图可知, 30 天内基本走势为上升, 只有两次大规模的回落, 为了保证盈利, 可以在两次回落点前进行买卖。

t 日实施投资并持有两天以上后将获得的股票当日收益率为:

$$R_t = \frac{u_{t+n} - u_t}{u_t} \quad (11)$$

其中, u_{t+n} 为卖出时的收盘价, u_t 为买入时的收盘价。

根据公式(11), 得到在两次交易中分别卖出股票后的当日收益率, 根据公式(12)可以计算得出模拟投资期间的累计收益率。

$$\text{CumR} = \prod_{t=1}^n (1 + R_t) - 1 \quad (12)$$

其中, R_t 表示 t 日收益率, n 表示做出投资操作的次数。

本文中, 根据投资策略, 30 天内, 第一次在 11 月 19 日买入股票, 11 月 25 日卖出股票, 其收益率为 0.053; 第二次在 11 月 28 日买入股票, 12 月 23 日卖出股票, 其收益率为 0.15。因此, 两次投资以后获得的累计收益率为 0.21。

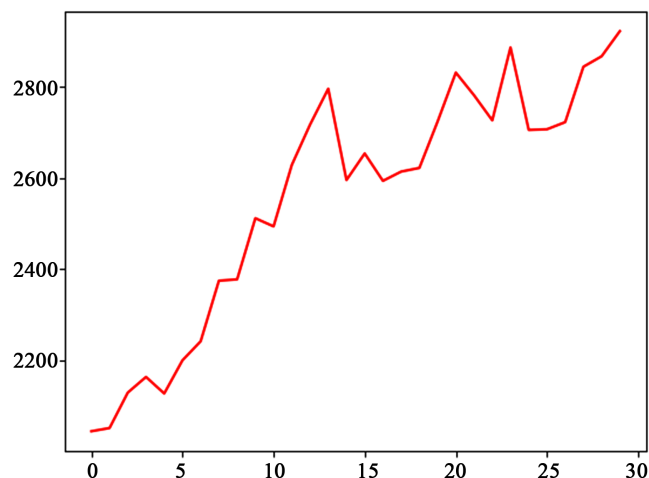


Figure 11. Thirty-day closing price

图 11. 30 天收盘价

5.3. 结果分析

夏普比率也称为夏普指数，是可以同时衡量收益和风险的综合指标，表示投资者每多承担一单位风险可获得的额外收益，计算公式如(13)所示：

$$\text{Sharperatio} = \frac{E(R_p) - R_f}{\sigma_p} \quad (13)$$

其中， $E(R_p)$ 表示投资组合的期望收益， R_f 表示无风险利率， σ_p 表示投资组合的标准差。如果夏普比率为正值，说明在衡量期内股票的平均净值增长率超过了无风险利率，在以同期银行存款利率作为无风险利率的情况下，说明投资股票比银行存款要好。夏普比率越大，说明股票的单位风险所获得的风险回报越高。夏普比率为负时，按大小排序没有意义。夏普比率是以资本市场线作为评价基准，对投资绩效做出评估。

本文根据公式(13)计算得出所构建的投资组合的夏普比率为 1.618，由此说明该投资方案要好过银行存款，并且单位内获得正回报。

6. 总结

本文提出了基于 SNM-KNN 算法的股票价格预测方法，并结合相关技术指标于对上证综指短期、中期和长期的涨跌趋势进行预测。本算法首先通过技术指标对股票涨跌趋势进行初步刻画形成特征矩阵，然后利用 SVM 良好的分类能力预测股票涨跌趋势，同时结合 KNN 算法预测涨幅，实现准确预测股票价格的目的。最后，实验结果验证本文所提算法对股票市场指数具有较好的预测能力。

参考文献

- [1] Abdoh, T.H. and Jouhare, H. (1996) The Investigation of Efficiency of Stock Price Index of TSE. *Journal of Financial Research*, **13**, 11-12.
- [2] Chang, P.C., Liu, C.H., Lin, J.L., et al. (2009) A Neural Network with a Case Based Dynamic Window for Stock Trading Prediction. *Expert Systems with Applications*, **36**, 6889-6898. <https://doi.org/10.1016/j.eswa.2008.08.077>
- [3] Oliveira, F.A.D., Nobre, C.N. and Zárte, L.E. (2013) Applying Artificial Neural Networks to Prediction of Stock Price and Improvement of the Directional Prediction Index-Case Study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, **40**, 7596-7606. <https://doi.org/10.1016/j.eswa.2013.06.071>

- [4] Doeksen, B., Abraham, A., Thomas, J. and Paprzycki, M. (2005) Real Stock Trading Using Soft Computing Models. *International Conference on Information Technology: Coding & Computing*, Las Vegas, NV, 4-6 April 2005, 162-167. <https://doi.org/10.1109/ITCC.2005.238>
- [5] Kara, Y., Boyacioglu, M.A. and Baykan, K. (2011) Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, **38**, 5311-5319. <https://doi.org/10.1016/j.eswa.2010.10.027>
- [6] Kamijo, K. and Tanigawa, T. (1990) Stock Price Pattern Recognition—A Recurrent Neural Network Approach. 1990 *IJCNN International Joint Conference on Neural Networks*, San Diego, CA, 17-21 June 1990, 215-221. <https://doi.org/10.1109/IJCNN.1990.137572>
- [7] Guresen, E., Kayakutlu, G. and Daim, T.U. (2011) Using Artificial Neural Network Models in Stock Market Index Prediction. *Expert Systems with Applications*, **38**, 10389-10397. <https://doi.org/10.1016/j.eswa.2011.02.068>
- [8] Sui, X.S., Qi, Z.Y., Yu, D.R., et al. (2008) A Novel Feature Selection Approach Using Classification Complexity for SVM of Stock Market Trend Prediction. 2007 *International Conference on Management Science and Engineering*, Harbin, 20-22 August 2007, 1654-1659.
- [9] Chakravarty, S. and Dash, P.K. (2012) A PSO Based Integrated Functional Link Net and Interval Type-2 Fuzzy Logic System for Predicting Stock Market Indices. *Applied Soft Computing*, **12**, 931-941. <https://doi.org/10.1016/j.asoc.2011.09.013>
- [10] Shao, X.L., Ma, D., Liu, Y., et al. (2018) Short-Term Forecast of Stock Price of Multi-Branch LSTM Based on K-Means. 2017 *4th International Conference on Systems and Informatics*, Hangzhou, 11-13 November 2017, 1546-1551. <https://doi.org/10.1109/ICSAI.2017.8248530>
- [11] Liu, Y.C. and Yeh, I.C. (2017) Using Mixture Design and Neural Networks to Build Stock Selection Decision Support Systems. *Neural Computing & Applications*, **28**, 1-15. <https://doi.org/10.1007/s00521-015-2090-x>
- [12] Kai, C., Yi, Z. and Dai, F. (2015) A LSTM-Based Method for Stock Returns Prediction: A Case Study of China Stock Market. 2015 *IEEE International Conference on Big Data*, Santa Clara, CA, 29 October-1 November 2015, 2823-2824.
- [13] Lee, J.W., Hong, E. and Park, J. (2004) A Q-Learning Based Approach to Design of Intelligent Stock Trading Agents. 2004 *IEEE International Engineering Management Conference*, Singapore, 18-21 October 2004, 1289-1292.
- [14] Utomo, D. (2017) Stock Price Prediction Using Back Propagation Neural Network Based on Gradient Descent with Momentum and Adaptive Learning Rate. *Journal of Internet Banking and Commerce*, **22**, No. 3.
- [15] Dai, H., Zhang, Y. and Wang, D. (2011) Price Prediction of Stock Index Futures Based on SVM. 2011 *Fourth International Conference on Business Intelligence and Financial Engineering*, Wuhan, 17-18 October 2011, 54-57. <https://doi.org/10.1109/BIFE.2011.96>
- [16] Yu, H., Chen, R. and Zhang, G. (2014) A SVM Stock Selection Model within PCA. *Procedia Computer Science*, **31**, 406-412. <https://doi.org/10.1016/j.procs.2014.05.284>
- [17] Nayak, R.K., Mishra, D. and Rath, A.K. (2015) A Naïve SVM-KNN Based Stock Market Trend Reversal Analysis for Indian Benchmark Indices. *Applied Soft Computing*, **35**, 670-680. <https://doi.org/10.1016/j.asoc.2015.06.040>
- [18] Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T. and Johnson, J.E. (2016) Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists. *Expert Systems with Applications*, **61**, 215-234. <https://doi.org/10.1016/j.eswa.2016.05.033>
- [19] Fama, E.F. and Blume, M.F. (1966) Filter Rules and Stock Market Trading. *Journal of Business*, **39**, 226-241. <https://doi.org/10.1086/294849>
- [20] Brock, W., Lakonishok, J. and LeBaron, B. (1992) Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *Journal of Finance*, **47**, 1731-1764. <https://doi.org/10.2307/2328994>
- [21] Conrad, J. and Kaul, G. (1998) An Anatomy of Trading Strategies. *Review of Financial Studies*, **11**, 489-515. <https://doi.org/10.1093/rfs/11.3.489>
- [22] Lo, A.W., Mamaysky, H. and Wang, J. (2000) Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. *Journal of Finance*, **55**, 1705-1765.
- [23] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [24] Press, H.W., Teukolsky, A.S., Vetterling, T.W. and Flannery, B.P. (2007) Section 16.5. Support Vector Machines, in: *Numerical Recipes: The Art of Scientific Computing*. 3rd Edition, Cambridge University Press, New York.