

# 基于 SVM 及股价趋势的股票预测方法研究

邬春学, 赖靖文

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:**传统 SVM 算法对股票的研究主要是在数值上,通过训练后直接得到相应的值。阐述了通过 SVM 算法预测涨跌后再使用股价趋势获得相应预测值的预测方法。将历史数据输入模型作为训练集,经过 SVM 算法训练后,再与大盘走势、K 线、MACD 线、成交量等技术指标进行处理,作出拟合预测。最后通过走势方向准确度、总盈利率等指标,与基于 SVM 算法的其它预测进行比较,验证该模型的应用成效与可用度,为后续研究提供参考。

**关键词:**股票预测;SVM 算法;股价趋势;总盈利率

**DOI:**10.11907/rjdk.172448

**中图分类号:**TP301

**文献标识码:**A

**文章编号:**1672-7800(2018)004-0042-03

## Stock Prediction Method Analysis Based on SVM Model and Stock Price Trend

WU Chun-Xue, LAI Jing-Wen

(School of Optical-Electrical and Computer, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** The traditional study of SVM algorithm on stock market is centered on the numerical value and obtains the corresponding value directly after the training. This paper introduces the prediction method in which corresponding prediction value is obtained by SVM algorithm and the stock price trend is combined into consideration of the prediction of corresponding values. The historical data are inputted into the model as a training set. The data are then trained by SVM algorithm and dealt with the trend of the market, K line, MACD line, volume and other technical indicators to make a prediction. Finally, the trend direction accuracy, the total profit rate and other indicators and the newest predictions based on SVM algorithm are compared to test the application effectiveness and availability of the model, which makes a reference for the follow-up study.

**Key Words:** stock prediction; SVM algorithm; stock price trend; total profit margin

## 0 引言

股票具有变化多、非线性、可控性差等特点,2009 年、2015—2016 年的股灾更是历历在目,从争先恐后的盲目追高,到诚惶诚恐地一路阴跌,其中不乏一些赚的满盆钵、买车买房的暴发户,也有亏空跳楼的投资者。正是因为这种复杂多变的风险与收益落差极大的特点,促使许多学者毕生致力于股票研究。

股票预测研究方法很多,比较常见的是回归分析、时间序列分析两种方法。这些方法在一定程度上揭示了股票的规律,但在股票预测的稳定性、对大盘涨跌和总盈利率预测上并不理想。支持向量机(SVM)理论是以统计学为基础的统计学习理论(Statistical Learning Theory),对

于小样本的快速预测能得到全局最优解。虽然预测效果很大程度上取决于核函数的优劣,但相对于单纯使用 ARIMA、贝叶斯等建模难度大、难以收敛、容易陷入局部极小值的模型,SVM 还是比较适用的。SVM 优秀的泛化能力是其成为最常用的分类器的重要原因。

综上所述,本文决定采用 SVM 算法套用在神经网络模型上预测上证股价,通过分析本算法以及近期最新的预测算法,进行调整,为后续的股票预测研究打下坚实的基础。

## 1 SVM 及相关工作

### 1.1 选股方法

选股方法很多,有从价格直接选取估值较低的股票进

收稿日期:2017-08-30

作者简介:邬春学(1964—),男,博士,上海理工大学光电信息与计算机工程学院教授、硕士生导师,研究方向为计算机应用技术、高速网络技术、网络控制系统、无线自动巡检、智能仪器仪表及嵌入式系统、计算机监测与控制;赖靖文(1993—),男,上海理工大学光电信息与计算机工程学院硕士研究生,研究方向为软件工程。

行分析和买进的,有从交易量看炒股板块热度进行投资的,也有跟着大单买入卖出的,也有看经济新闻买入的。看盘方式形形色色,亏了赚了只有实践才知道。归纳起来有下面3种:

(1)基本面。基本面就是股票的整体情况,从该公司的经营理念策略、公司报表、分红以及未来的公司运营方向等考虑,这方面的投资者一半是中长线(超过30日或以上),相当于是一种价值投资,其收益除了股票本身的价值差外还有股票分红等,是比较稳健的投资方式。

(2)技术面。技术面就是股票的价格、成交量等数据。从K、RSI、MACD、MA均线等分析股票的价格走势。这种投资者一般是短线或超短线(日抛或周抛),相当于风险投资。收益很大,但风险也相对较高。

(3)消息面。消息面就是注意公司的各种信息。由于中国股票市场制度还不完善,很多公司与券商联合起来炒作股票,券商买入后放出消息吸引资金追捧从中获利。所以消息面和技术面是相辅相成的两个选股指标,虽然市场上虚假消息以及延后消息非常多,但毕竟是公司内部出来的消息,具有一定的价值。

本文的SVM算法基于数据统计,主要从技术层面出发,消息以及基本面暂不考虑。

## 1.2 SVM应用

SVM算法由Vladimir N Vapnik和Alexey Ya发明。1992年,Bernhard E Boser、Isabelle M Guyon和Vladimir N Vapnik提出了一种通过将内核技巧应用于最大余量超平面创建非线性分类器的方法。目前的标准化身(软裕度)由Corinna Cortes和Vapnik于1993年提出。

SVM方法通过一个非线性映射 $p$ ,把样本空间映射到一个高维乃至无穷维的特征空间中(Hilbert空间),使得在原来的样本空间中非线性可分的问题转化为在特征空间中的线性可分问题,简单说就是升维和线性化。升维,就是把样本向高维空间做映射,一般情况下会增加计算的复杂性,甚至引起“维数灾难”,因而人们很少问津。但是分类、回归等问题在低维样本空间很可能无法线性处理样本集,在高维特征空间却可通过一个线性超平面实现线性划分(或回归)。一般升维都会带来计算复杂化,SVM方法巧妙地解决了这个难题:应用核函数的展开定理,不需要非线性映射的显式表达式;由于是在高维特征空间中建立线性学习机,所以与线性模型相比,不但没有增加计算的复杂性,而且在某种程度上避免了“维数灾难”,这是核函数的展开和计算理论的扩展。

使用不同核函数,可得到不同的SVM答案,常用的核函数有以下4种:①线性核函数 $K(x, y) = x \cdot y$ ;②多项式核函数 $K(x, y) = [(x \cdot y) + 1]^d$ ;③径向基函数 $K(x, y) = \exp(-|x - y|^2 / d^2)$ ;④二层神经网络核函数 $K(x, y) = \tanh(a(x \cdot y) + b)^{[1]}$ 。

正是由于SVM算法基于小样本,有着良好的数学基底,所以具有良好的鲁棒性,因而广泛应用于各行各业。

## 2 SVM股票处理模型

### 2.1 股票选取与预处理

影响股票价格的因素很多,人为因素占非常大的比例,很多券商为了炒高股价而先期大量买入,后放出消息,再择机卖出,这样的股票分析起来很困难,经常会出现没有成交量暴跌,或成交量大及暴涨等不合理、难以分析的现象。因此本文选取盘面较大的(难以被操控的)股票。数据从2014年7月8日至2017年5月26日数据,共680组,其中30组作为训练数据,训练出模型,其余650组作为预测数据。本文以收盘价为预测对象,将数据与其它网站(网易财经、东方财富、大智慧等)搜集的数据先进行完整性处理,将残缺的数据根据东方财富、网易财经等比对应后填入数据;然后进行数据准确性处理,将券商间的信息比对应后,与其它券商不同的数据进行更改处理。

进行比较并将停盘数据删除之后的数据如图1所示。

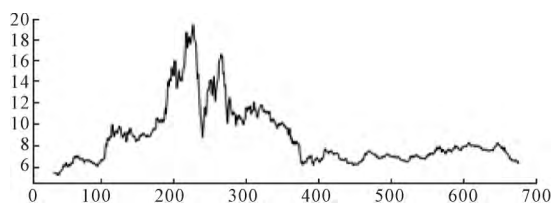


图1 中国重工(601989)收盘价

### 2.2 股价趋势分析模型

模型选择因为影响因素多,因此不能仅选取其中之一来决定股票价格,使用最多的是神经网络知识构建预测模型,如图2所示。

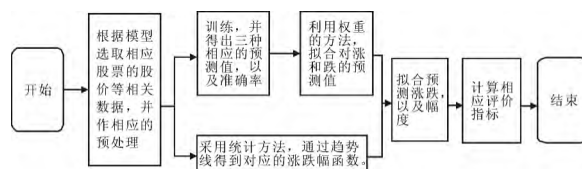


图2 模型及其算法流程

## 3 模型评价指标

### 3.1 股价走势准确率

为了分析本模型效果,选取两个指标进行评判:①走势准确率;②总盈利率。

走势准确率是和股票走势方向相同的次数与总预测次数的比值,希望知道的是股价走势和未来的方向,以此预测走势。若一个股票预测系统能正确预测大盘走势,说明这个预测模型具有很高的实用性和使用价值。

### 3.2 股票总盈利

投资者最关心的是总盈利率,该指标反映模型在单只股票中一定的交易日内,是否为投资者带来收益,是评价模型的重要指标。

#### 4 模型预测结果案例分析

本文选取中国重工(601989)从2014年7月8日至2017年5月21日的681个交易日数据作为数据源,从中划分出训练以及预测的相邻数据。以其中30组作为训练集,其余的作为预测集,之后根据权重计算相应涨跌,再根据股票的成交量与涨幅趋势计算出涨跌的相关指标,最后根据给出的结果计算相应指标,如表1所示。

表1 使用各项指标的准确率

参数	准确率(%)
5、10、20、30 均线	51.73
大盘涨跌	50.25
其余项指标	47.63
综合	49.00

综合公式如下:

$$P(i) = 1 * P_i / (P_1 + P_2 + P_3) \quad (1)$$

根据不同模型计算涨跌幅,如图3所示。

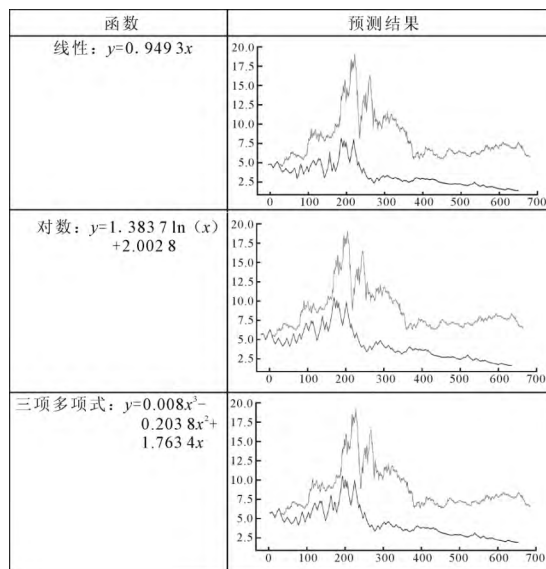


图3 各趋势线形成的股价对比

盈利率计算公式如下:

$$E = (E_{old} - E_{get}) / E_{get} \times 100\% \quad (2)$$

为使数据更有说服力,取连续上行的第3个峰作为买入点,以及连续下行的第3个峰作为卖出点,得到盈利率

为0.53。虽然该算法的盈利率在股票市场不是很理想,但相对于银行利率,预测结果还是可以接受的,具有一定的实用价值。

#### 5 结语

本文通过建立SVM模型,分别用k线、大盘以及其它指标综合预测,给出了股票涨跌的预测结果。根据价格变动幅度和换手率做出趋势方程,得到较好的结果。但SVM涨跌预测还存在较大改进空间,在评判体系以及算法改进方面还有很多需要研究的地方。

参考文献:

- [1] 百度百科. SVM 词条[EB/OL]. [http://baike.baidu.com/link? url= Ptry1cCOJxbzMe\\_yoN11b-jwJN4e4zCi-IhgCRMQN\\_SA6mRhrGn418WBc-wHhOXYKT9Qq02cE4BR-xR8zYDA2q](http://baike.baidu.com/link? url= Ptry1cCOJxbzMe_yoN11b-jwJN4e4zCi-IhgCRMQN_SA6mRhrGn418WBc-wHhOXYKT9Qq02cE4BR-xR8zYDA2q). 2017. 7. 10.
- [2] 张秀艳,徐本立. 基于神经网络集成系统的股市预测模型[J]. 系统工程理论与实践,2003,23(9):67-70.
- [3] 喻胜华,肖雨峰. 基于信息粒化和支持向量机的股票价格预测[J]. 财经理论与实践,2011,32(6):44-47.
- [4] 程砚秋. 基于支持向量机的证券价格预测方法研究[D]. 大连:大连理工大学,2007.
- [5] 张良均,王路,谭立云,等. Python 数据分析与挖掘实战[M]. 北京:机械工业出版社,2015.
- [6] WEN F H, XIAO J H, HE Z F. Fuzzy cognitive diagnosis for modelling examinee performance[J]. Procedia Computer Science, 2014(31):625-631.
- [7] 黄秋萍,周霞,甘宇健,等. SVM 与神经网络模型在股票预测中的应用研究[J]. 微型机与应用,2015,34(5):88-90.
- [8] 程昌品,陈强,姜永生. 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.
- [9] 韩瑜,刘淑环. 基于 GARCH-SVM 和 AR-SVM 的个股涨跌预测[J]. 大连海事大学学报,2016,15(3):25-30.
- [10] 贾秀倩,刘翔,周洋,等. 基于 LS-SVM 算法的旅游数量预测[J]. 科技经济导刊,2016(1):6-7.
- [11] 黄同愿,陈芳芳. 基于 SVM 股票价格预测的核函数应用研究[J]. 重庆理工大学学报,2016,30(2):89-94.
- [12] JUNG S B, KIM S W. Improvement of scanning accuracy of PZT piezoelectric actuators by feed-forward model-reference control[J]. Precision Engineering,1994,16(1):49-55.
- [13] YAN D F, ZHOU G, ZHAO X, et al. Predicting stock using microblog moods[J]. China Communications,2016,30(8):244-257.

(责任编辑:杜能钢)

(上接第41页)

- [4] EDMUND M C, ORNA G, DORON P. Model checking[M]. Cambridge: MIT Press,2000.
- [5] MANNA Z, PNUELI A. Temporal verification of reactive systems[M]. Heidelberg: Springer-Verlag,1995.
- [6] LEE E A, SESHIA S A. 嵌入式系统导论 CPS 方法[M]. 北京:机械工业出版社,2011.
- [7] GEUVERS H, KOPROWSKI A, SYNEK AUTOMATED D, et al. Machine-checked hybrid system safety proofs[M]. Heidelberg: Springer-Verlag,2010.
- [8] BERTOT Y, CATERAN P. 交互式定义证明与程序开发——Cop 归纳构造演算的艺术[M]. 北京:清华大学出版社,2010.
- [9] 李倩. 基于形式化方法的混成系统安全性验证[D]. 上海:华东师范大学,2015:12-46.
- [10] GULWANI S. Automating string processing in spreadsheets using input-output examples[J]. ACM Sigplan-sigact Symposium on Principles of Programming Languages,2011,46(1):317-330.

(责任编辑:刘亭亭)