

# 基于KNN算法的A股量化交易策略

◆ 冼 靖

**摘要:** 运用机器学习KNN算法,筛选了11个因子,利用历史指标以及对应的涨跌情况训练模型,使用前一天指标预测当前交易日个股的涨跌情况。以沪深300成分股票作为标的,当预测上涨概率大于给定阈值时买入股票,并设置止盈与止损点,作为卖出时机。以历史数据回测,结果表明:综合年化收益率、收益波动率、最大回撤等指标分析,KNN模型表现较好,年化收益率比原来提高了11.6%,且波动率较小,即取得该收益率的偶然性及可能面临的最大的亏损较小,表示基于KNN算法的A股量化交易策略性能好。

**关键词:** KNN算法;量化交易策略;机器学习

DOI:10.16457/j.cnki.hbhjllw.2020.01.105

## 一、前言

如今信息科技发展迅速,以智能化、数字化为标志的技术创新呈现出广泛应用、深度融合的特征。现代金融行业就是这些信息技术极具挑战性和价值的应用场景,其中最具市场前景的就是股票预测,合理运用这些信息技术,进行股票市场的研究和预测,必然将给金融领域带来新机遇,更带动新一轮经济的腾飞。

机器学习算法作为重要的数据分析方法,相比于传统数据分析方法,其可以处理更多的输入信息,能够考虑的信息面更全,信息量更大,可以达到的效果上限更高。本文主要使用机器学习算法中的KNN进行股票市场的研究和预测,并对投资策略进行性能测试。

## 二、KNN算法

KNN是一种常用的监督学习方法,它的基本思想是“近朱者赤,近墨者黑”,其工作机制非常简单:给定测试样本,基于某种距离度量找出训练集中与其最靠近的 $k$ 个训练样本,然后基于这些邻居的信息来进行预测。在分类任务中可使用投票法,即选择这 $k$ 个样本中出现最多的类别标记作为预测结果;在回归任务中时使用平均法,即将 $k$ 个样本的实值输出标记平均值作为预测结果;还可基于距离远近进行加权平均或加权投票,距离越近的样本权重越大<sup>[1]</sup>。

## 三、仓位管理

在投资领域,交易择时和风险控制是同等重要的两大模块<sup>[2]</sup>。KNN算法解决了交易择时的问题,而马科维茨投资组合则在通过仓位配比来实现风险控制。

本文以沪深300指数的成分股票作为股票池,在每个回测交易日预测各个股的涨跌情况,建立了如下的交易策略:

- (1) 选取上涨概率最大的五支个股进行买入,上涨概率越大,买入资金越多,但一个交易日内使用的资金不能超过可用资金的5%。
- (2) 卖出下跌概率大于给定阈值的个股。
- (3) 在持仓过程中,若持有的个股股价触及止盈或止损点,卖出股票。

## 四、结果分析

回测利用了2017年1月1日至2019年1月1日的沪深300指数,为了提高策略的性能,为其添加指标作为新的参数。其

中包括4个反应短期营运能力的财务指标,5个反映大众投资态度的情绪因子,以及2个测量股价是否超出正常分布范围的动量因子。

(一) 收益率。基于KNN算法的投资策略年化收益率为6.8%,而基准年化收益率为-4.8%。可见策略的年化收益率比原来提高了11.6%,策略的表现较好。

(二) 波动率。波动率是金融研究领域关注的重要指标,它直接反映了金融资产或金融指数的收益率的变异程度,也是对未来收益所面对的风险和不确定性的一种度量<sup>[3]</sup>。收益波动率为15.8%,表明取得该收益率的偶然性较小。

(三) 最大回撤。最大回撤则用来显示买入产品后可能出现的最坏情况。投资者可能面临的最大的亏损为26.3%。

(四) 阿尔法系数与贝塔系数。阿尔法系数是投资的绝对回报和按照贝塔系数计算的预期回报之间的差额。由于阿尔法为8.4%,即大于零,表示股票价格可能被低估,建议买入。

贝塔系数为0.62。无法通过分散投资来减低的风险被称为系统风险,贝塔系数就是衡量股票投资系统风险的指标。贝塔系数小于1,表示投资的系统风险比较低。

(五) 夏普比率。夏普比率的绝对值为0.21,小于1,其投资机会“质量”较低,投资者以1单位风险换取更低的回报期望。

## 五、结语

随着神经网络理论日趋成熟,它凭借自身对噪声数据的高承受力、非线性映射能力、泛化能力、自学习自适应能力等优点称为股票预测的首选。它不需要建立复杂的关系式和数学模型,从而可以克服时间序列模型在预测上的局限。同时,也能避免人为因素的影响,这使得神经网络应用于股票预测领域,有一定新的发展空间。本文运用机器学习中的KNN算法进行中国A股量化交易策略的研究,分析结果发现年化收益率比原始提高了11.6%,波动率较小为15.8%,表明取得该收益率的偶然性较小,阿尔法系数、贝塔系数、夏普比率等指标均表示策略性能较好。

## 参考文献

- [1] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [2] 王艳荣. 长记忆波动率的模型研究与实证分析[D]. 西北农林科技大学, 2013.
- [3] 赵鹏飞. 沪深300股票指数收益波动率度量研究[D]. 中国青年政治学院, 2015.

(作者单位: 华南师范大学数学科学学院)