

股票市场的集成预测方法研究

耿恩泽, 满溢, 胡坤澎

北方工业大学, 北京

收稿日期: 2021年11月18日; 录用日期: 2021年12月2日; 发布日期: 2021年12月20日

摘要

为了研究集中单一预测方法和组合预测方法的预测的效果。我们分别选取逻辑回归模型、决策树模型、多元神经网络三种单一的预测方法进行预测。为了提高金融股票预测价格的稳定性与有效性, 我们把单一方法进行集成, 采用随机森林和循环神经网络两种集合方法进行测定。结果表明集成预测的效果稍优于单一预测, 而选取较好的单一方法进行集合预测会达到一个更好的效果, 基于大数据的循环神经网络具有良好的自适应性和很强的学习能力, 在对比的集中方法中是最有效的。

关键词

股票价格, 集成预测, 循环神经网络

Research on Integrated Forecasting Method of Stock Market

Enze Geng, Yi Man, Kunpeng Hu

North China University of Technology, Beijing

Received: Nov. 18th, 2021; accepted: Dec. 2nd, 2021; published: Dec. 20th, 2021

Abstract

In order to study the forecasting effect of centralized single forecasting method and combined forecasting method. We select three single forecasting methods, logistic regression model, decision tree model and multivariate neural network, respectively. In order to improve the stability and effectiveness of financial stock price prediction, we integrate the single method and use random forest and recurrent neural network to measure. The results show that the effect of ensemble prediction is slightly better than that of single prediction, and the selection of a better single method for ensemble prediction will achieve a better effect. The circular neural network based on big data has good adaptability and strong learning ability, and is the most effective among the comparative centralized methods.

Keywords

Stock Price, Integrated Prediction, Recurrent Neural Network

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

股票市场的预测一直是金融分析领域的难题。对股票投资者来说,未来股价变化趋势预测越准确,对利润的获取及风险的规避就越有把握;对国家的经济建设和发展而言,股票预测研究同样具有重要作用。因此对股票价格趋势预测的研究有着十分重要的意义。股票市场的特性是高收益与高风险并存。一般来说,收盘价对于短线投资者有较大的指导意义,尤其是在尾市,如果股价出现了放量上涨的态势,股价以当天较高的价格收盘,次日往往会有更突出的表现,因此此类个股有短线机会。但对于中长线投资者来说,关注收盘价往往是为了判断趋势的变化[1]。当出现带长下影线的十字星时,一般意味着个股底部的来临,对于市场个股判断有积极的技术意义,反之则可以判断为个股走势的头部。总之,作为一种重要的价格表现,收盘价的高低是市场投资者特别是短线投资者必须重视的一个技术数值。当然,观察收盘价格的作用也要同时结合其他的价格表现,包括开盘价、最高价和最低价,有时候还要结合成交的情况来进行综合分析和判断。因为金融市场不是线性的,因此大量的分析方法都不尽如人意,所以我们引入机器学习。在机器学习引入量化领域的早期,受制于计算能力,通常使用的是较为简单的模型,例如 logistic 回归, SVM 以及决策树等,然而一个显而易见的事实是股票市场的机制过于复杂,依托于简单的模型想要取得战胜市场的效果是基本不可能的,因此更为复杂的模型开始被引入量化领域,这其中,最具代表性的就是随机森林和神经网络。所以为了研究股票市场的集成预测效果,文章首先利用逻辑回归、决策树和多元神经网络三种单一方法进行预测。然后对随机森林模型和循环神经网络模型两种非线性集成方法对股票指数进行预测。最后把单一预测的结果和组合预测的结果放在一起进行对比分析[2]。

2. 三种预测方法对上证指数的预测效果研究

本篇文章我们以上海证券交易所当成研究对象。为了研究三种方法对于各类股票预测的可靠性,我们分别选择上海证券指数, A 股和 B 股。

预测的过程为:我们采用逻辑回归模型、决策树模型和多元神经网络这三种常用的股票市场预测方法分别进行预测后,对比这三种方法的预测效果,研究得出一定的结论。下面我会将这三种方法进行简单的描述,为了简洁起见,我们只对上证指数进行研究,得出他的预测结果。A 股和 B 股的结论不用图表形式给出。

2.1. 逻辑回归模型对股票指数的预测

分别以上海证券指数、A 股和 B 股作为研究对象。把 2015 年的数据作为训练样本 2015 年的 235 个数据作为训练样本,2016 年前三个月的 60 个数据作为测试样本。用表 1 中的 12 个向量作为输入变量,把明日的收盘价当成输出变量。预测过程选择逻辑回归,过程利用 python 实现,部分数据见表 2。

2.2. 决策树模型对股票指数的预测

分别以上海证券指数、A 股和 B 股作为研究对象。把 2015 年的数据作为训练样本 2015 年的 235 个数据作为训练样本，2016 年前三个月的 60 个数据作为测试样本。用表 1 中的 12 个向量作为输入变量，把明日的收盘价当成输出变量。划分数数据集，70% 的训练数据 30% 的测试数据，算法采用决策树模型。过程利用 python 实现，部分数据见表 2。

2.3. 多元神经网络模型对股票指数的预测

分别以上海证券指数、A 股和 B 股作为研究对象。把 2015 年的数据作为训练样本 2015 年的 235 个数据作为训练样本，2016 年前三个月的 60 个数据作为测试样本。用表 1 中的 12 个向量作为输入变量，把明日的收盘价当成输出变量。多元神经网络是一个单层结构，我们设置 1000 个隐含层，学习效率设置未 0.01，最大迭代次数设置为 1000。过程利用 python 实现，部分数据见表 2。

Table 1. The input variable

表 1. 输入向量

X1	X2	X3	X4	X5	X6
今日开盘价	今日收盘价	昨日开盘价	昨日收盘价	今日最低价	今日最高价
X7	X8	X9	X10	X11	X12
昨日最高价	昨日最低价	10 天移动平均 最高价	10 天移动平均 最低价	成交量	成交额

Table 2. Results of three prediction methods

表 2. 三种预测结果

	真实值	预测值 MLR2	预测值 CART	预测值 MLP
1	3003.92	3003.888756	3009.96	3003.888756
2	3009.53	2971.61694	2960.97	2971.61694
3	3053.07	3045.037492	3009.96	3045.037492
4	3050.59	3036.846721	3005.17	3036.846721
5	3008.42	3019.666703	3038.14	3019.666703
6	2984.96	2973.624409	2960.97	2973.624409
7	3033.96	3058.908889	3009.96	3058.908889
8	3023.65	3005.198042	3005.17	3005.198042
9	3066.64	3103.910818	3141.59	3103.910818
10	3082.36	3065.226461	3038.14	3065.226461
MSE		773.0134	1626.5224	459.1378
R ²		0.464	0.2963	0.709

在对测试集进行检验的时候，我们发现逻辑回归模型的均方误差是 773.0134，拟合优度达到了 0.464；多元神经网络模型的均方误差是 459.1378，拟合优度是 0.709；而决策树模型的均方误差高达 1626.5224，拟合优度则只有 0.2963，因此在选取单一的检验方法时选择多元神经网络较好。

3. 组合预测的测定方法

本篇文章我们以上海证券交易市场当成研究对象。为了研究三种方法对于各类股票预测的可靠性，我们分别选择上海证券指数，A 股和 B 股。

预测的过程为：我们采用随机森林模型和循环神经网络模型两种集合模型来测试预测的效果。对比这两种方法的预测效果并和单一预测进行对比，研究得出一定的结论。下面我会将这两种方法进行简单的描述，为了简洁起见，我们只对上证指数进行研究，得出他的预测结果。A 股和 B 股的结论不用图表形式给出。

3.1. 随机森林的集合预测

分别以上海证券指数、A 股和 B 股作为研究对象。把 2015 年的数据作为训练样本 2015 年的 235 个数据作为训练样本，2016 年前三个月的 60 个数据作为测试样本。采用自抽样的方法，抽取一部分来建立模型。对特征进行处理，把建立出的很多模型集成，通过随机森林的算法来提高多样性。在这里，我们建立了 100 个子模型，参数都采用刚刚决策树的参数。数据见表 3。

3.2. 循环神经网络的集合预测[3]

分别以上海证券指数、A 股和 B 股作为研究对象。把 2015 年的数据作为训练样本 2015 年的 235 个数据作为训练样本，2016 年前三个月的 60 个数据作为测试样本。循环神经网络是一种深度学习方法，该模型存在记忆能力可以自身返回。代码主要运用 Keras Sequential 的深度学习框架，首先进行标准化，数据导入 Keras 后导入隐含层 layers 最后倒入优化函数 Adam，也就是自适应梯度消减法。首先建立一个训练模型，在里面添加了一个循环层，在这里我设置了 512 个隐层节点，然后对模型进行编译和拟合。数据见表 3。

Table 3. Comparison of results of several prediction methods

表 3. 几种预测结果对比

	真实值	预测值 MLR2	预测值 CART	预测值 MLP	预测值 RF	预测值 RNN
1	3003.92	3003.888756	3009.96	3003.888756	3012.385	3009.55531
2	3009.53	2971.61694	2960.97	2971.61694	2982.818	2981.74313
3	3053.07	3045.037492	3009.96	3045.037492	3044.474	3039.26436
4	3050.59	3036.846721	3005.17	3036.846721	3048.818	3046.42713
5	3008.42	3019.666703	3038.14	3019.666703	3025.124	3024.1518
6	2984.96	2973.624409	2960.97	2973.624409	2979.591	2977.05049
7	3033.96	3058.908889	3009.96	3058.908889	3051.42	3046.55606
8	3023.65	3005.198042	3005.17	3005.198042	3016.094	3015.20248
9	3066.64	3103.910818	3141.59	3103.910818	3103.193	3093.50961
10	3082.36	3065.226461	3038.14	3065.226461	3075.826	3074.53897
MSE		773.0134	1626.5224	459.1378	343.9756	389.9852
R ²		0.464	0.2963	0.709	0.5945	0.7287

在对集合测试集进行检验的时候，我们发现随机森林的均方误差是 0343.9756，拟合优度达到了 0.5945；循环神经网络模型的均方误差为 389.9852，拟合优度达到了 0.7287。可见运用集合方法比单一

方法要好一些，但也不是总优于单一方法。综合来看选取较好的单一方法进行集合的效果最好，循环神经网络具有良好的学习能力，效果最好。

4. 结论

通过结果我们发现，三种单一的预测方法中多元神经网络的预测结果最好，最稳定，而使用决策树的预测效果是最差的最不稳定的。虽然组合预测与单一预测相比不是总是优势，但无论是从数据的稳定性或是数据的有效性来看回归神经网络的表现都要更好。因此在进行组合预测的时候，选择效果好的单一预测方法进行组合预测可以提高预测结果的有效性和稳定性。在文章中讨论的这几种方法来看，循环神经网络的有效性和稳定性是最好的。运用循环神经网络来预测股票市场的价格和变化趋势是可行的，然而，影响股票价格的因素往往是复杂多变的，提高预测模型的泛化能力需要进一步的研究

对于怎么提高预测结果我总结出以下三点：

- 1) 从数据本身入手，提高数据的多样性；
- 2) 算法入手，改善模型；
- 3) 算法的结合方式。

参考文献

- [1] 殷光伟. 中国股票市场预测方法的研究[D]: [博士学位论文]. 天津: 天津大学, 2003.
- [2] 王国华. 中国股票市场日内波动率研究[D]: [博士学位论文]. 武汉: 中南财经政法大学, 2017.
- [3] 纪滕. 基于 BP 网络的股票预测研究[D]: [博士学位论文]. 昆明: 昆明理工大学, 2014.