Finance 金融, 2020, 10(4), 334-342
Published Online July 2020 in Hans. http://www.hanspub.org/journal/fin
https://doi.org/10.12677/fin.2020.104034

Hans汉斯

A Stock Price Prediction Model Based on Stock Charts and Deep CNN

Qiao Zhou, Ningning Liu, Lingcong Shen

University of International Business and Economics, Beijing
Email: ningning.liu@uibe.edu.cn, zojoy_dario@163.com, shen_lc@outlook.com

Received: Jun. 18th, 2020; accepted: Jul. 1st, 2020; published: Jul. 8th, 2020

Abstract

The prediction of the stock market has always been a challenging issue, because many factors will cause the market uncertainty such as national policies, company financial reports, industry performance, investor sentiment, social media sentiment, and economic factors. In this paper, based on the stock charts method, the continuous time stock information is processed. According to different information richness, prediction time interval and classification method, the original data is divided into multiple categories as the training set of DCNN (Deep Convolutional Neural Network). The results show that the method has the best performance when the forecast time interval is 30 days. Moreover, this method can accurately predict the stock trend of the US NDAQ exchange for 59.7%.

Keywords

Stock Market Predicted, Convolutional Neural Network, Stock Charts

基于股票图像与CNN的股价预测模型研究

周　乔，刘宁宁，沈灵聪

对外经济贸易大学，北京
Email: ningning.liu@uibe.edu.cn, zojoy_dario@163.com, shen_lc@outlook.com

收稿日期：2020年6月18日；录用日期：2020年7月1日；发布日期：2020年7月8日

摘　要

股票市场的预测一直是一个具有挑战性的问题，其波动会受国家政策、公司财报、行业表现、投资者情绪等因素的影响。本文基于股市图像(Stock Charts)方法将股票的连续时间信息进行处理，根据不同的信息

文章引用: 周乔, 刘宁宁, 沈灵聪. 基于股票图像与 CNN 的股价预测模型研究[J]. 金融, 2020, 10(4): 334-342.
DOI: 10.12677/fin.2020.104034

丰富度以及预测时间间隔将原始数据分为了多个类别，依次作为深度卷积神经网络(Deep Convolutional Neural Network, DCNN)训练集；并利用深度卷积神经网络对股票市场进行预测，分析在不同分类方法下的精度差异。结果表明，当在标记间隔为30天，使用包含成交量的蜡烛图作为输入时，对美国NDAQ交易所的股票走势预测可以达到59.7%的准确度。

关键词

股市预测，卷积神经网络，蜡烛图

1. 引言

股票市场的涨跌与当前宏观经济形势具有对应关系，对股票市场的正确预测有利于国家及时调整宏观经济政策，维护市场、社会的稳定发展，因此正确预测股票市场的走势成为了领域内的热点问题。然而影响股票价格的因素很多，例如国家政策、公司财报、行业表现等因素，使得股票趋势预测成为了一个非常具有挑战性的问题。根据 Fama 的有效市场假说[1]，在市场信息完备的情况下，投资者仍不能通过进行基本面分析与技术分析来获得超出市场平均利润的收益，原因便是金融时间序列中的高噪声[2] [3]。在传统统计学方法中，常用统计分析[4] [5]配合股票图像(Stock Charts)的方法，对股票市场的走势进行判断，往往能取得不错的效果。如吴泽兵[6]基于蜡烛图进行量化交易并制定了“红三兵”、“牛市鲸吞线”两个量化策略，其“红三兵”策略可在 5 日与 3 日持仓的交易中达到 88.46%的胜算率，牛市鲸吞线的五日持仓收益率也有 3.85%。杜兵[7]则利用蜡烛图分析方法，对我国创业板个股进行回溯，获得了较优的平均收益率、盈亏比与胜率。

随着信息技术的发展，计算水平的日益提高，计算机视觉技术在医学[8]、农学[9]、工学[10]等领域都有了广泛运用，其与股票市场趋势研究的结合，必然可以为这一领域带来新的可能。如 Luca Di Persio 等人[11]利用多层感知器(Multi-layer Perceptron, MLP)、卷积神经网络(CNN)、长短记忆时间递归神经网络(Long Short-Term Memory, LSTM)以及循环神经网络(RNN)对标准普尔 500 (S & P 500)指数的价格运动进行了预测，结果显示深度学习算法对股票市场的趋势拟合效果较好。其中，CNN 建模效果最好。Yang Jiao [12]基于深度网络，通过标准交叉验证、顺序验证以及单次验证方法的比较，发现利用近期信息，例如已经收盘的欧洲和亚洲指数来预测标准普尔 500 指数，可以使预测精度大幅提高。Gozde Sismanoglu [13]等人利用 1968~2018 年的 IBM 股票信息，使用 MLP 与 CNN 算法对数据库中信息的进行预测，结果证明该方法对特定股票有较好的精度，可以大概率正确预测第二日股票的涨跌的情况。2019 年，Keywan 等人[14]描述了机器学习的一些基本概念，并提供了一个简单的例子，说明投资者如何使用机器学习技术预测股票收益的横截面，同时模拟过度拟合的风险。X. Zhang 等人[15]通过输入来自历史股票的交易数据和社交媒体信息，发现通过有效分析金融新闻和用户情绪，可以达到预测股票市场动向的目的。此外，我们整理、分类和分析近几年发表的有关股票预测模型的文献，如表 1 所示。我们发现：金融科技领域中的基于股票图形趋势研究已经被证实可行[15]，而目前机器学习领域中的图像深度学习方面取得了突破进展[6]。在本论文中，我们将金融科技的股市图形趋势研究与机器学习的图像深度学习有机结合，即基于 DCNN 与股票图像结合的方法对股票市场进行预测，如图 1 所示。

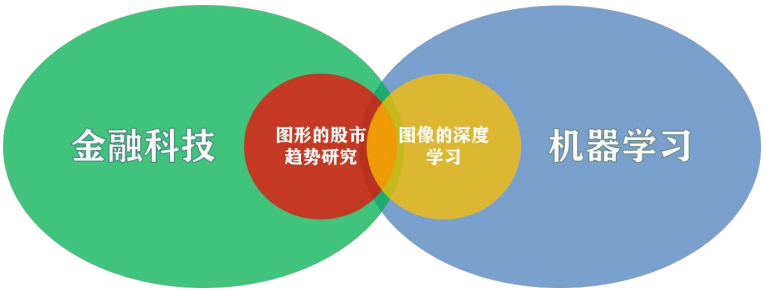


Figure 1. The cross field between machine learning and fintech
图 1. 机器学习与金融科技的交叉

基于以上分析，本文将蜡烛图方法配合新型的深度学习技术来预测 NASDAQ 的股市走势，弥补了传统非机器学习方法的不足，将深度学习算法与股票图像创新性的融合。利用金融时间序列数据处理成的信息丰富度不同两种的蜡烛图，将图片数据集作为 CNN 的输入，对比预测未来 1 天、20 天、30 天以及 60 天股票趋势的准确度，利用时间划分分类训练方式，对比两种图片的预测准确度。

Table 1. Stock prediction method and model collation
表 1. 股票预测方法及模型整理

分类	方法/模型	研究内容	作者
基于传统的非机器学习的股市预测方法	向量自回归(VAR)模型 误差修正模型(ECM) 卡尔曼滤波模型(KFM)	利用三种模型对英国股市的长期变化进行预测	Chulho Jung 等人
	GARCH 模型 QGARCH 模型 GJR 模型	研究三种模型对金融时间序列数据(中国股市)波动的预测能力和模型间相互对比	魏巍贤 周晓明
	马尔科夫预测模型	利用马尔科夫预测法对沪股东北高速收盘价进行分析，预测股价变动	李海涛
	“红三兵”、“牛市鲸吞线”量化策略	基于蜡烛图进行量化交易并制定量化策略	吴泽兵
	蜡烛图分析方法	对我国创业板个股进行回溯，追求较优的平均收益率、盈亏比与胜率。	杜兵
基于深度学习模型的股市趋势预测方法	多层感知器 CNN 长短记忆时间递归神经网络 RNN	利用多层感知器、CNN、长短记忆时间递归神经网络以及 RNN 预测标准普尔 500 指数的趋势	Luca Di Persio 等人
	标准交叉验证 顺序验证 单次验证方法	通过三种方法的比较，发现利用近期信息可以使预测精度大幅提高。	Yang Jiao
	MLP 与 CNN 算法	使用 MLP 与 CNN 算法对数据库中 IBM 股票信息的进行预测	Gozde Sismanoglu 等人
	ANN + Tree boosting	说明投资者如何使用多算法深层叠加预测股票收益的横截面，同时模拟过度拟合的风险。	Keywan 等人
	深度学习算法	输入来自历史股票的交易数据和社交媒体信息，通过有效分析金融新闻和用户情绪预测股票市场动向。	X. Zhang 等人

2. 方法

2.1. 股票图像

烛台图作为一种金融图表，可用于描述给定时间段内的股票价格走势。烛台图由日本大米交易商

Munehisa Hooma 开发[16]，因此也被称为日本烛台图。每个烛台通常显示一天的交易数据，因此一个月图可将 20 个交易日转换为 20 个烛台。图 2 显示了烛台图表应包含的信息，每个烛台包含了交易日信息的四个重要组成部分，即开盘价，收盘价，低价和高价。烛台通常由 3 部分组成，即上阴影线，下阴影线和实体。如果开盘价高于收盘价，则主体将填充为红色；否则，主体将以绿色填充，以此表示股票的涨跌。上下阴影表示指定时间段内的高价和低价范围。但是，并非所有烛台都有阴影。烛台图是可视化的帮助，可以辅助股票交易决策。根据烛台图，交易者将更容易理解高点和低点以及开盘价和收盘价之间的关系。因此，交易者可以确定特定时间范围内的股票市场趋势[17]。当收盘价大于开盘价时，烛台被称为看涨烛台。否则，它被称为看跌烛台。

本文中主要使用可视化图像即股票图像参与深度卷积神经网络训练过程，通过股票图像的特性和 CNN 视觉及图像处理方面的优势，来达到对股票趋势的准确预测的目的。

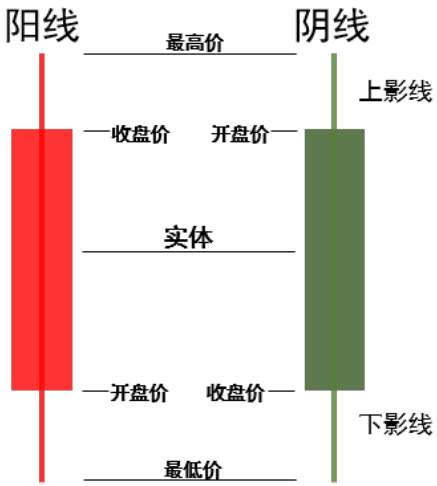


Figure 2. Candlestick chart diagram
图 2. 烛台图示意图

2.2. 深度卷积神经网络模型

在本研究中，我们将使用基于 CNN 的深度学习网络(Deep Learning Networks, DLN)即深度卷积神经网络(DCNN)对股票市场预测进行分类。卷积神经网络多用于图像的各种处理任务，如目标检测[18]、图像分类[19]、图像分割[20]等，其展示出了远超传统方法的精度。卷积神经网络要求所采集数据信息为时间连续信息，即某个像素点的值均与其临近像素点有关联，而 CNN 相比于全连接层，更易提取出纹理信息与边缘信息，从而提高预测效果。

CNN 是一种前馈人工神经网络，它包括输入层，输出层和一个或多个隐藏层，其结构如图 3 所示[21]。CNN 的隐藏层通常由池化层、卷积层和全连接层组成。卷积层负责读取小段数据并使用内核读取诸如二维图像或一维信号之类的输入，并扫描整个输入字段。池化层采用特征投影，最终池化层的输出被发送到一个或多个全连接层，这些层将解释已读取的内容并将此内部表示形式映射为类值。CNN 类似于由一组具有可学习的权重和偏见的神经元组成的普通神经网络(NN)，区别在于卷积层使用卷积运算来输入，然后将结果传输到下一层。此操作允许使用更少的参数更有效地实现前向功能。

为了使预测结果更加精准，本文训练了 CNN 模型。本文结果亦证明 CNN 在针对计算机视觉和图像处理方面的问题非常有效。本节针对需要处理的金融时间序列数据构成的图片构建了基于 CNN 的网络模型，网络模型由 4 个 2d 卷积层，4 个 2d 最大池化层和 3 个输出层组成，如表 2 所示。

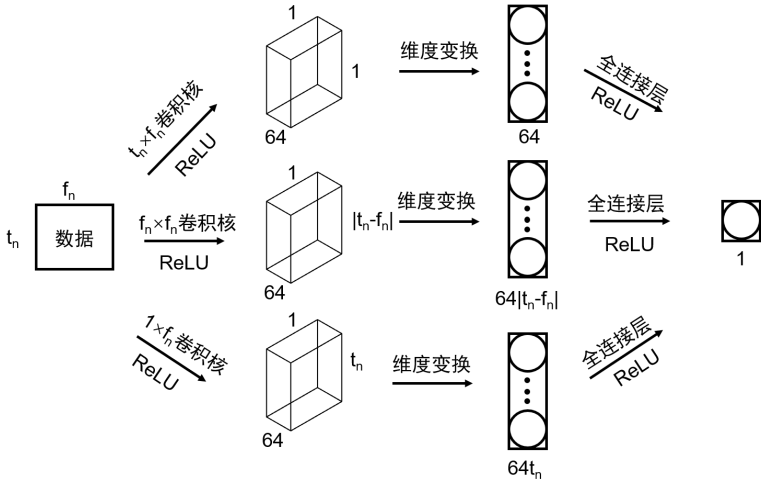


Figure 3. Schematic diagram of CNN network structure
图 3. CNN 网络结构示意图

Table 2. CNN structure for stock trend prediction
表 2. 针对股票走势预测的 CNN 结构

Input
Conv2D-32 *ReLU*
Max-poolling
Conv2D-48 *ReLU*
Max-poolling
Dropout
Conv2D-64 *ReLU*
Max-poolling
Conv2D-96 *ReLU*
Max-poolling
Dropout
Flatten
Dense-256
Dropout
Dense-2

3. 实验

3.1. 样品选取和实验设置

正确获取待测数据是模型能够成功预测的基础。本文基于 Yahoo!的应用程序接口(API)服务，采集了 NDAQ 交易所 100 只股票的交易数据，所收集数据的日期如表 3 所示。需要注意的是，交易日并非是连续的(星期一到星期五是交易日，节假日不交易)，因此，在数据爬虫时应对数据日期进行筛选，避免使用空白图片训练或测试。

Table 3. Data type and division
表 3. 数据类型及划分

Stock Data	Training Data		Testing Data	
	Start	End	Start	End
NDAQ	2014/12/31	2018/12/31	2019/1/1	2019/12/31

3.2. 数据分类方式

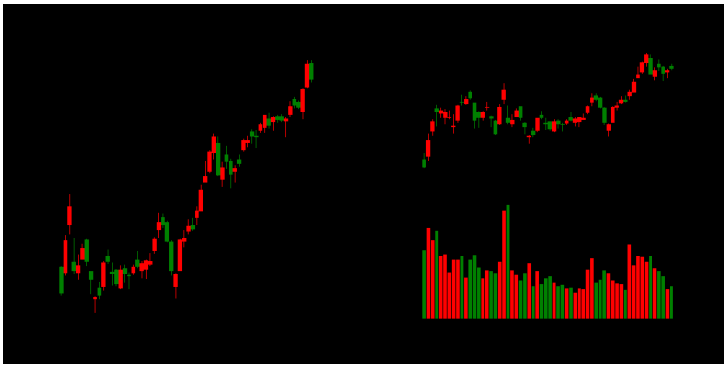
本研究中对股票时间序列数据采用时间划分方法，按照预设时间划分数据将经过筛选的股票数据划分为训练集与测试集。

时间划分是利用金融时间序列的时间特性，按照每只股票的交易时间(Date)，设定截止日期，将测试日期后的数据设定预测集，即最终的测试集中包含每只股票的部分数据。本文共收集了 NDAQ 交易所 100 只股票 5 年的交易数据，利用股票数据的时间特性，抽取 2019/1/1 至 2019/12/31 之间的数据作为测试集。

3.3. 数据预处理

获取数据后，需要对数据进行预处理，提取数据信息，本文选用了金融数据的可视化方法，即烛台图对股票数据预处理。将历史时间序列数据，使用 Python 中的 Matplotlib 库[22]将其转换为烛台图。本研究中使用的烛台图如图 4 所示，其基于不同的标记进行对比实验。为了分析不同预测间隔与预测准确度之间的关系，本文基于不同的预测间隔(1 个、20 个、30 个、60 个和 90 个交易日后的 close 差值)对每张蜡烛图进行标记，计算方法如式(1)，当 $close_{i+d} < close_i$ 时，$target_i = 0$，当 $close_{i+d} > close_i$ 时，$target_i = 1$。本文所用烛台图均含 60 天信息量[23]。

$$t\arg et_i = sign\left(close_{i+d} - close_i\right) \tag{1}$$



(a) 第一类蜡烛图　　(b) 第二类蜡烛图

Figure 4. The example of candle chart
图 4. 蜡烛图示例

4. 结果与讨论

在本节中，首先根据建立的 CNN 模型计算预测准确度，之后对预测间隔和图片丰富度两个指标对预测准确度的影响进行分析，对比获得何种数据分类方式更优，以便未来进一步研究。

4.1. 绩效评估

绩效评估有一些统计方法，可以通过测量灵敏度(真实阳性率或召回率)、特异性(真实阴性率)、准确

性和马修相关系数(MCC)来评估所有分类器的结果。通常，TP 为真阳性或正确识别，FP 为假阳性或错误识别，TN 为真阴性或正确剔除，FN 为假阴性或错误剔除。对应公式如下：

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specitivity = \frac{TN}{TN + FP} \tag{3}$$

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}。 \tag{5}$$

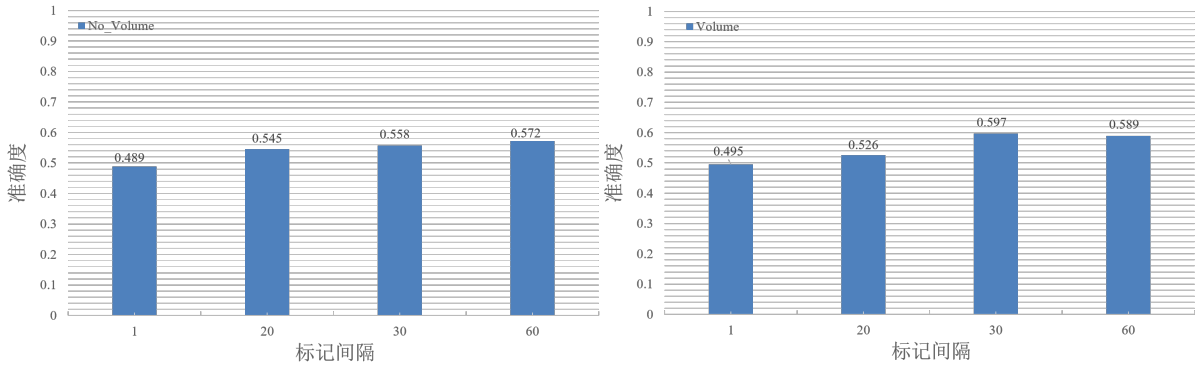
4.2. 预测间隔对准确度的影响

图 5 为两类图片在不同预测间隔下的测试结果。

由图 5(a)可得，作为信息丰富度较少的第一类图，其预测准确度随时间间隔的变大而提高，在间隔为 1 时，准确度为 0.489；20~60 天准确度增长为 5%；直到间隔 60 天时准确度超过 0.57，认为此划分方法可以承担起股票市场预测的任务。

由图 5(b)可得，带有成交量信息的第二类图在不同间隔的准确度由间隔为 1 时的 0.495 升至间隔为 30 时的 0.597 再到间隔为 60 时的 0.589，存在极大值。且测试间隔范围内的最大准确度高于第一类图。

综上，在图片丰富度最高的第二类图的情况下，卷积神经网络 CNN 可以提取到相比第一类蜡烛图更多的特征值来训练预测。因此在实验数据量相同的情况下，使用 30 天间隔的带有成交量信息的蜡烛图训练同样会得到不错的结果。划分间隔可以通过细分比较得出更精确的结果，本文只追求粗略区间，不继续对比。



(a) 第一类蜡烛图准确度对比　　(b) 第二类蜡烛图准确度对比

Figure 5. Accuracy comparison of two types of candle charts at different intervals
图 5. 两类蜡烛图不同间隔下准确度对比

4.3. 图片丰富度对准确度的影响

我们可以基于信息的丰富程度对两类蜡烛图排序，即第二类大于第一类。

基于 4.2 中的分析，综合所有数据，我们可以明显看出，即使是第二类图最高的预测准确度只与第一类图的最高准确度相差 0.025。结果显而易见，第二类图在除间隔为 20 天之外的间隔条件下准确度均高于第一类图，并且在间隔为 30 天时达到最高值 0.597。因此，在本文中的研究范围内，考虑各种因素，

第二类图在标记间隔为 30 天时作为数据集训练得到的预测结果最好。两类图片训练准确度对比如图 6 所示。

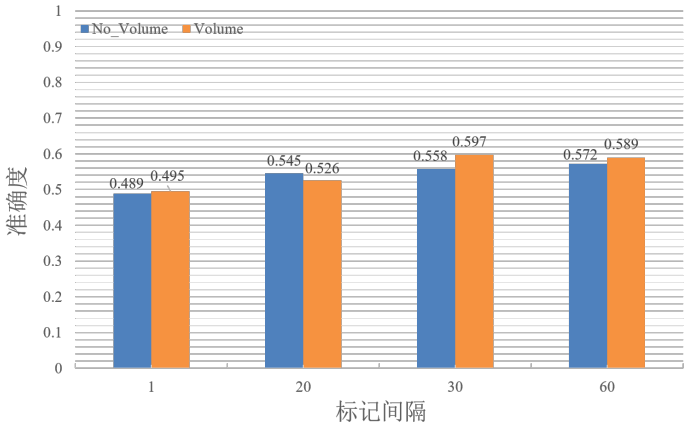


Figure 6. CNN model prediction accuracy
图 6. CNN 模型预测准确度

5. 总结与展望

本文采集了 2014/12/31~2019/12/31 之间的 NDAQ 100 只股票的时间序列交易数据，基于卷积神经网络算法构建了股票市场预测模型，结合统计学指标对准确度进行量化分析，结论如下：

1) 根据 CNN 模型预测结果，可知包含成交量(Volume)的蜡烛图在时间间隔为 30 天时拥有 0.597 的准确度，是所有实验条件的较佳组合。

2) 在图片丰富度方面的比较可以明显看出包含成交量信息的蜡烛图准确度较高，因此在类似实验中，提高蜡烛图信息丰富度也是提高预测准确度的一个方向。

3) 在未来研究中，基于不同图片丰富度，可在现有数据上可以尝试探索加入均线或换手率等数据来进行对比实验。

致　谢

感谢我的导师刘宁宁老师，在整个论文写作过程中给予我的大力帮助。刘老师的悉心指导贯穿了论文写作的方方面面，在他的指导下我认识到了自己很多不足，并在这一过程中取得进步。

基金项目

这项工作得到了国家青年科学基金资助(批准号：61806056)，北京市社会科学青年基金资助(批准号：17YYC015)，中央高校基本科研业务专项资金资助(批准号：CXTD10-05)。

参考文献

[1] Malkiel, B.G. and Fama, E.F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, **25**, 383-417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

[2] Jung, C. and Boyd, R. (1996) Forecasting UK Stock Prices. *Applied Financial Economics*, **6**, 279-286. https://doi.org/10.1080/096031096334303

[3] 魏巍贤, 周晓明. 中国股票市场波动的非线性 GARCH 预测模型[J]. 预测, 1999(5): 47-49.

[4] 李海涛. 运用马尔科夫预测法预测股票价格[J]. 统计与决策, 2002(5): 25-26.

[5] Prado, H.D., Ferneda, E., Morais, L.C.R., *et al*. (2013) On the Effectiveness of Candlestick Chart Analysis for the Bra-

zilian Stock Market. *Procedia Computer Science*, **22**, 1136-1145. https://doi.org/10.1016/j.procs.2013.09.200

[6] 吴泽兵. 基于A股市场的K线形态量化分析[J]. 广西质量监督导报, 2018(11): 63-64.

[7] 杜兵. 基于蜡烛图的择时策略在我国创业板股票中的研究[J]. 科技经济导刊, 2019, 27(3): 16-17.

[8] Litjens, G., Kooi, T., Bejnordi, B.E., *et al*. (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[9] Kamilaris, A. and Prenafeta-Boldú, F.X. (2018) Deep Learning in Agriculture: A Survey. *Computers and Electronics in Agriculture*, **147**, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[10] 孙剑, 张一豪, 王俊骅. 基于自然驾驶数据的分心驾驶行为识别方法[J/OL]. 中国公路学报: 1-17. https://kns.cnki.net/kcms/detail/61.1313.u.20200512.1104.002.html, 2020-07-02.

[11] Honchar, O. and Di Persio, L. (2016) Artificial Neural Networks Approach to the Forecast of Stock Market Price Movements. *International Journal of Economics and Management Systems*, **1**, 158-162.

[12] Jiao, Y. and Jakubowicz, J. (2017) Predicting Stock Movement Direction with Machine Learning: An Extensive Study on S&P 500 Stocks. 2017 *IEEE International Conference on Big Data* (*Big Data*), Boston, 11-14 December 2017, 4705-4713. https://doi.org/10.1109/BigData.2017.8258518

[13] Sismanoglu, G., Onde, M.A., Kocer, F., *et al*. (2019) Deep Learning Based Forecasting in Stock Market with Big Data Analytics. 2019 *Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science* (*EBBT*), Istanbul, 24-26 April 2019, 1-4. https://doi.org/10.1109/EBBT.2019.8741818

[14] Rasekhschaffe, K.C. and Jones, R.C. (2019) Machine Learning for Stock Selection. *Financial Analysts Journal*, **75**, 70-88. https://doi.org/10.1080/0015198X.2019.1596678

[15] Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B. and Philip, S.Y. (2018) Improving Stock Market Prediction via Heterogeneous Information Fusion. *Knowledge-Based Systems*, **143**, 236-247. https://doi.org/10.1016/j.knosys.2017.12.025

[16] Morris, G.L. and Litchfield, R. (1995) Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures. McGraw-Hill, New York.

[17] Lu, T.-H., Shiu, Y.-M. and Liu, T.-C. (2012) Profitable Candlestick Trading Strategies—The Evidence from a New Perspective. *Review of Financial Economics*, **21**, 63-68. https://doi.org/10.1016/j.rfe.2012.02.001

[18] Wang, J., Zheng, T., Lei, P., *et al*. (2019) A Hierarchical Convolution Neural Network (CNN)-Based Ship Target Detection Method in Spaceborne SAR Imagery. *Remote Sensing*, **11**, 620. https://doi.org/10.3390/rs11060620

[19] Sharifzadeh, F., Akbarizadeh, G. and Seifi Kavian, Y. (2019) Ship Classification in SAR Images Using a New Hybrid CNN-MLP Classifier. *Journal of the Indian Society of Remote Sensing*, **47**, 551-562. https://doi.org/10.1007/s12524-018-0891-y

[20] Bao, S. and Chung, A.C.S. (2018) Multi-Scale Structured CNN with Label Consistency for Brain MR Image Segmentation. *Computer Methods in Biomechanics and Biomedical Engineering*: *Imaging & Visualization*, **6**, 113-117. https://doi.org/10.1080/21681163.2016.1182072

[21] 乔若羽. 基于神经网络的股票预测模型[J]. 运筹与管理, 2019, 28(10): 132-140.

[22] Hunter, J.D. (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, **9**, 90-95. https://doi.org/10.1109/MCSE.2007.55

[23] 程智胜. 股票市场突发行为的数据挖掘[D]: [硕士学位论文]. 武汉: 华中师范大学, 2019.