

基于深度学习的股票价格预测

叶可禾

(深圳大学, 电子与信息工程学院 2021280462)

摘要: 在股票交易方面, 深度学习近些年来一直是研究的重点。为研究国内外深度学习在量化交易方面的发展现状并研判其发展趋势, 对过去股票预测的阶段性发展历史、深度学习的主流实现算法进行了总结回顾, 对 ANN、LSTM 和 CNN 算法的基本原理进行了详细地介绍和综合评述。可以得出至今为止, 深度学习在量化交易层面上的研究越来越多, 在股票价格预测方面趋于成熟, 但是仍有研究空间。

关键词: 深度学习; 股票交易; 深度人工神经网络; 长短期记忆神经网络; 卷积神经网络

1 股票价格预测的发展历史

1.1 人为预测阶段(1611-1951)

世界上最早的股份有限公司制度诞生于 1602 年在荷兰成立的东印度公司, 股份公司这种企业组织形态出现以后, 很快为西方国家广泛应用, 成为西方国家企业组织的重要形式之一。伴随着股份公司的诞生和发展, 以股票形式集资入股的方式也得到发展, 并且产生了买卖交易转让股票的需求, 这样, 就带动了股票市场的出现和形成, 并促使股票市场完善和发展。1611 年, 东印度公司的股东们在阿姆斯特丹股票交易所就进行着股票交易, 并且后来有了专门的经纪人撮合交易, 阿姆斯特丹股票交易所形成了世界上第一个股票市场。

这个阶段的是股票和股市的发展期, 股票预测仍然停留在人为预测的阶段。

1.2 计算机化被动投资的崛起(1952-1975)

股市预测是金融和计算机科学交叉领域的经典问题。对于这个问题, 著名的有效市场假说(EMH)给出了一个悲观的观点, 并暗示金融市场是有效的(Fama, 1965), 它认为技术分析或基本面分析(或任何分析)不会产生任何一致的超平均给投资者带来利润。^[6]

1973 年, 美国经济开始陷入滞胀状态, 美股也因此出现新一轮大幅下跌。

1970 年的诺贝尔经济学奖得主萨缪尔森在这个时候发表了自己的研究成果。萨缪尔森通过数据研究发现, 绝大多数的专业投资者都没能跑赢市场, 而即便是能够跑赢, 我们也非常难判断这是因为运气因素还是因为水平因素。

1975 年, 约翰·博格推出了历史上第一只

指数基金——第一指数投资信托。之后他又将其重新命名为“先锋 500”指数, 将其变为一只完全追踪标普 500 指数的基金。这一基金吸收了从马尔基尔到艾里斯的理念, 采取完全被动的投资方式, 整个投资完全由计算机操作, 它预示着计算机化被动投资的崛起。

除了那些崛起的被动投资指数基金, 其他的机构投资者也在计算机的崛起之中越发认识到智能化的威力。交易显示电子化和交易速度的进步只是其中一个方面, 除此之外, 智能与量化也逐渐深入到各种类型的机构资产管理之中。

在马科维茨之前, 购买收益最高的证券, 把鸡蛋放在一个篮子之中然后看住它, 是投资界认可的投资收益最大化方式。这种思想来源于实业家安德鲁·卡内基。但是, 在 1952 年, 马科维茨却提出, 通过分散化, 可以在不改变投资组合预期收益的前提下降低风险。

威廉·夏普提出了更为简洁可操作的资本资产定价模型(CAPM)。根据资本资产定价模型, 市场的风险被分为两类, 一类是系统性风险, 一类是非系统性风险。在一个竞争均衡的资本市场中, 系统性风险无法消除, 但是非系统性风险却可以通过市场组合来降低到最小。与马科维茨的理论相比, 夏普的理论更有利于其在实践中的大规模应用。

马科维茨和夏普的理论可能算是 20 世纪下半场最为重要的理论, 它们为金融行业的进一步数量化建立了基础。在之后的年月中, 成型的理论配合着计算机技术的进步, 成了推动金融业革新的最主要力量。

1.3 深度学习与量化交易的繁荣(1976 至今)

当计算机出现在股票交易当中时, 就一发不可收拾了, 在股票交易中深深扎根的数学骨髓与计算

机相结合，诞生出了至今为止在发达国家地区规模最大的股票投资交易方法——基于深度学习的量化交易。

金融市场预测对股票市场投资和风险管理至关重要。股票市场是一个受多种因素影响的复杂非线性动态系统，如政策、社会新闻事件、公司运营投资者的心理变化。因此，预测股市走势或股价是一件极其困难的事情。^[2]所以深度学习技术逐渐被运用到股票投资和交易当中，以博取产生更高的利润。学术努力一直致力于探索新方法和新的预测指标以实现更好的预测性能。

随着人工智能的发展，机器学习算法被广泛应用于各个领域。学术研究还探讨了机器学习对金融市场预测的影响，包括贷款违约率、汇率、资产价格和股票波动率，等等。同时，学者们也发现机器学习可以用来提高股票市场的预测性能。经典机器学习模型，例如随机森林模型 (RF) 和梯度提升决策树 (GBDT) (Leippold et al., 2021; Krauss et al., 2017) 广泛用于股票市场预测。随着算法的改进，越来越多的研究开始使用深度学习模型，并得出结论认为深度学习算法比传统机器学习模型具有更好的预测性能 (Fischer & Krauss, 2018)。其中，主流的深度学习算法包括深度人工神经网络 (ANN)、长短期记忆神经网络 (LSTM) 和卷积神经网络 (CNN)。^[4]

2 量化交易中主流深度学习算法的实现方法

2.1 深度人工神经网络 (ANN)

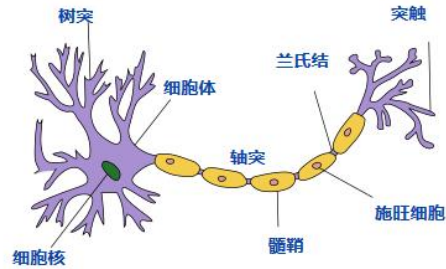
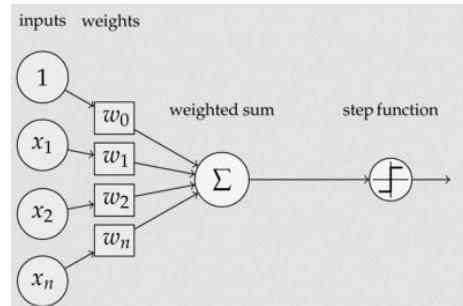
2.1.1 人工神经网络

人工神经网络 (Artificial Neural Network, ANN)，也简称神经网络，是众多机器学习算法中比较接近生物神经网络特性的数学模型。人工神经网络通过模拟生物神经网络 (大脑) 的结构和功能，由大量的节点 (或称“神经元”，或“单元”) 和之间相互连接构成，可以用来对数据之间的复杂关系进行建模。^[5]

从深度学习模型的结构来看，它可以看作是一个具有许多隐藏层的多层感知器。从深度学习模型的工作原理来看，深度学习模型经历了从低级特征到高级特征的特征归纳。深度学习模型是利用多个隐藏层，一层一层地进一步总结下层的特征，然后转移到下一个隐藏层。之后，需要继续提取和总结

高级特征。最后，可以获得更准确的特征提取，从而更准确地逼近原始数据的实际特征。^[10]

2.1.2 感知机与人工神经网络的实现



我们可以看到单个神经元在树突上收集其他神经元传递过来的信息，在信息量到达一个阈值之后，便会产生一个脉冲信号，经过轴突传送到突触，进而传递给下一个神经元。感知机正是模仿这一神经元的运行过程进行的设计。最初的感知机模型也是一种二元表示法，也就是要么有信号产生，要么没有信号产生。

我们把感知机的模型进行分解，可以分为：输入权值、激活函数、输出。

输入权值：每个输入都有一个权值，另外还有一个 w_0 表示偏置，偏置的作用是让后面的阶跃函数位于原点。^[5]

激活函数：这里采用阶跃函数作为激活函数，实际应用的还有很多激活函数，诸如 Sigmoid 函数, ReLU 等等。

$$f(z) = \begin{cases} 1 & z > 0 \\ 0 & \text{otherwise} \end{cases}$$

输出：该感知器的输出由下面这个公式来计算

$$y = f(w \cdot x + b)$$

ANN 实际上就是一个庞大的神经网络，是模仿人脑构造的一种函数网络。相较于 CNN，ANN 对于庞大数据群的处理显得较为乏力，无法在较少的时间内完成训练，所以就有了 CNN 的出现。

2.2 长短期记忆神经网络 (LSTM)

2.2.1 长期记忆、短期记忆

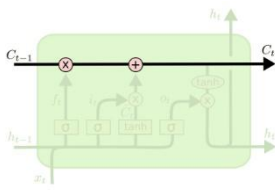


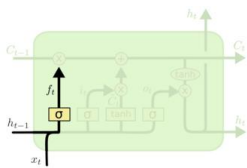
Fig.1 长期记忆线

长期记忆，也叫细胞状态，用于完成序列学习的工作，如图 1 中最上方的一条长线，并在线上连接了三道门。

短期记忆，事件信息，由门来控制添加到长期记忆中的部分，从而达到选择性通过的结果。

LSTM 是一种特殊的 RNN，能够学习长期依赖关系。它的引入是为了处理消失和爆炸梯度问题。^[8] LSTM 在每个单元模块中添加了三个重复的神经网络层，分别为：遗忘门、输入门和输出门。每个 LSTM 都是一组捕获数据流的单元。这些单元从一个模块连接到另一个模块，传输过去的数据并收集当前数据。这些门基于 sigmoid 函数的神经网络层，使这些单元能够选择性地允许数据通过或处理数据。^[9]

2.2.2 遗忘门



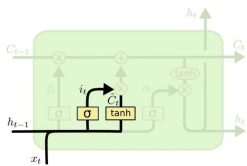
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Fig.2 遗忘门

LSTM 中的第一道门，决定了在细胞状态中我们需要舍弃多少信息。

如图 2，读入和后经过函数整合为一个 0~1 之间的数，来告诉中的数字需要保留多少，“0”是全部舍弃，“1”是全部保留。

2.2.3 输入门



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Fig.3 输入门

输入门用于计算新的细胞状态。

输入门（如图 3）由两步构成，第一步是先计算一个叫做“Input Gate Layer”的 sigmoid 层和一个 tanh 层来决定哪些信息需要更新。

然后将 $i_t \times \tilde{C}_t$ 加到 C_{t-1} 中，更新为 C_t 。^[7]

2.2.4 输出门

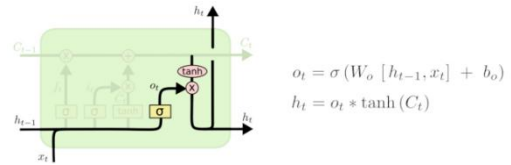


Fig.4 输出门

输出门用于计算该细胞单元的短期记忆值 h_t 。

h_t 由一个 sigmoid 层和一个 tanh 层相乘而来，其中 tanh 层完全由当前细胞状态 C_t 处理得到。^[7]

2.2.5 sigmoid 函数和 tanh 函数

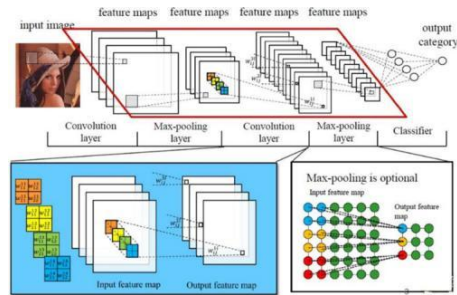
人工智能中处理数据常用的激活函数。

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

2.3 卷积神经网络 (CNN)

2.3.1 卷积神经网络的结构



卷积神经网络由三大块组成：输入层、隐含层和输出层。

其中隐含层是卷积神经网络的核心所在，如图红框内所示，隐含层又由卷积层、池化层和全连接层组成。^[1]

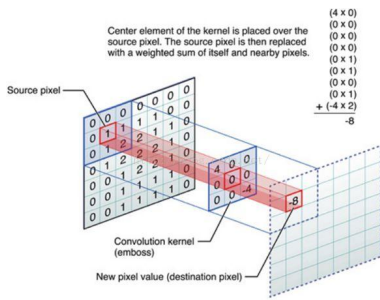
2.3.2 输入层

卷积神经网络的输入层可以处理多维数据，常见地，一维卷积神经网络的输入层接收一维或二维数组，其中一维数组通常为时间或频谱采样；二维数组可能包含多个通道；二维卷积神经网络的输入层接收二维或三维数组；三维卷积神经网络的输入层接收四维数组。由于卷积神经网络在计算机视觉领域应用较广，因此许多研究在介绍其结构时预先假设了三维输入数据，即平面上的二维像素点和 RGB 通道。^[4]

与其它神经网络算法类似，由于使用梯度下降算法进行学习，卷积神经网络的输入特征需要进行标准化处理。具体地，在将学习数据输入卷积神经网络前，需在通道或时间/频率维对输入数据进行归一化，若输入数据为像素，也可将分布于的原始像素值归一化至区间。输入特征的标准化有利于

提升卷积神经网络的学习效率和表现。^[4]

2.3.3 卷积层



如上图的操作就被称为卷积，卷积核的大小就是卷积层的感受野。

2.3.4 卷积核

卷积层的功能是对输入数据进行特征提取，其内部包含多个卷积核，组成卷积核的每个元素都对应一个权重系数和一个偏差量（bias vector），类似于一个前馈神经网络的神经元。同时可以存在多个卷积核，即多卷积核。^[4]

2.3.5 卷积层参数

卷积层参数包括卷积核大小、步长和填充，三者共同决定了卷积层输出特征图的尺寸。

其中卷积核大小可以指定为小于输入图像尺寸的任意值，卷积核越大，可提取的输入特征越复杂。卷积步长定义了卷积核相邻两次扫过特征图时位置的距离，卷积步长为1时，卷积核会逐个扫过特征图的元素，步长为n时会在下一次扫描跳过n-1个像素。

由卷积核的交叉相关计算可知，随着卷积层的堆叠，特征图的尺寸会逐步减小，例如16×16的输入图像在经过单位步长、无填充的5×5的卷积核后，会输出12×12的特征图。为此，填充是在特征图通过卷积核之前人为增大其尺寸以抵消计算中尺寸收缩影响的方法。常见的填充方法为按0填充和重复边界值填充(replication padding)。^[4]

2.3.6 权值共享

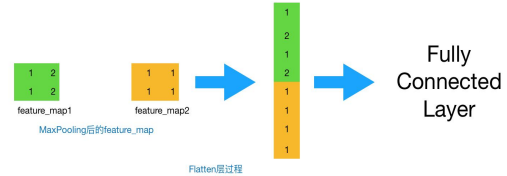
在卷积层中权值共享是用来控制参数的数量。假如在一个卷积核中，每一个感受野采用的都是不同的权重值（卷积核的值不同），那么这样的网络中参数数量将是十分巨大的。同一个深度的神经元的权值都是相同的。

2.3.7 池化层

通常在连续的卷积层之间会周期性地插入一个池化层。它的作用是逐渐降低数据体的空间尺寸，这样的话就能减少网络中参数的数量，减少计

算资源耗费，也能有效控制过拟合。

普通池化（General Pooling）：除了最大池化，池化单元还可以使用其他的函数，比如平均池化（average pooling）或L-2 范式池化（L2-norm pooling）。平均池化历史上比较常用，但是现在已经很少使用了。因为实践证明，最大池化的效果比平均池化要好。^[4]



2.3.8 全连接层

卷积神经网络中的全连接层等价于传统前馈神经网络中的隐含层。全连接层位于卷积神经网络隐含层的最后部分，并只向其它全连接层传递信号。特征图在全连接层中会失去空间拓扑结构，被展开为向量并通过激励函数。^[4]

3 对于 ANN、LSTM 和 CNN 的总结

3.1 ANN

ANN 是一种基于人脑神经元的构造函数，能够较好地拟合一些数据，实现较为简单。但是仅适用于线性条件，并且对于大规模的数据集，显得极为吃力，准确率不足。

3.2 LSTM

LSTM 是一种设计地极为精巧的针对于时间序列分析的算法，能够较好地预测时间序列下的数据，并且相较于 ANN 算法，LSTM 更不吃资源。但是 LSTM 却是一个较为费时的算法，如果 LSTM 的时间跨度很大，并且网络又很深，这个计算量会很大，很耗时。

3.3 CNN

CNN 是一种基于 ANN 算法的优化，通过共享卷积核来优化计算量，对高维数据处理无压力，且卷积层可以提取特征。但是需要调参，所需样本量较大，且池化时容易丢失大量有价值的信息。

4 结束语

在世界经济增速发展的今天，深度学习在股票价格预测方面发展日益成熟，一定会在未来充当股票投资的主力军，对未来量化交易产生重要的影响。

[参考文献]

- [1] Ian Goodfellow, Yoshua Bengio 深度学习[M]. 人民邮电出版社, 2017. 7, 20220617.
- [2] Meiyao Tao, Shanshan Gao, Deqian Mao, Hong Huang, Knowledge graph and deep learning combined with a stock price prediction network focusing on related stocks and mutation points, Journal of King Saud University - Computer and Information Sciences, 2022, ISSN1319-1578, <https://doi.org/10.1016/j.jksuci.2022.05.014>.
- [3] Chenyao Ma, Sheng Yan, Deep learning in the Chinese stock market: The role of technical indicators, Finance Research Letters, Volume49, 2022, 103025, ISSN1544-6123, <https://doi.org/10.1016/j.fr1.2022.103025>.
- [4] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016. 01, 20220619.
- [5] Martin T. Hagan, Howard B. Demuth. 神经网络设计[M]. 机械工业出版社, 2018. 12, 20220619.
- [6] Weiwei Jiang, Applications of deep learning in stock market prediction: Recent progress, Expert Systems with Applications, Volume 184, 2021, 115537, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.115537>.
- [7] Hochreiter S, J. Schmidhuber, Long Short-term Memory[J], Neural Computation, 9 (8) (1997), pp. 1735-1780.
- [8] Rui Liu, Yong Jiang, Jianwu Lin, Forecasting the Volatility of Specific Risk for Stocks with LSTM, Procedia Computer Science, Volume202, 2022, Pages111-114, ISSN1877-0509, <https://doi.org/10.1016/j.procs.2022.04.015>.
- [9] Fengrui Liu, Yang Li, Baitong Li, Jiixin Li, Huiyang Xie, Bitcoin transaction strategy construction based on deep reinforcement learning, Applied Soft Computing, Volume 113, Part B, 2021, 107952, ISSN1568-4946, <https://doi.org/10.1016/j.asoc.2021.107952>.
- [10] Meihua Xie, Haiyan Li, Yuanjun Zhao, Blockchain financial investment based on deep learning network algorithm, Journal of Computational and Applied Mathematics, Volume 372, 2020, 112723, ISSN0377-0427, <https://doi.org/10.1016/j.cam.2020.112723>.