

深度强化学习应用于金融市场量化交易研究综述

许波^{1,2}, 贺一峻^{1,2}, 温健城^{1,2}, 李祥霞^{1,2}

(1. 广东财经大学信息学院, 广东 广州 510320;

2. 广东省智能商务工程技术研究中心, 广东 广州 510320)

摘要: 深度强化学习 (deep reinforcement learning, DRL) 作为一种实现通用人工智能的有效学习范式, 已在一系列实际金融市场量化交易应用中取得了显著成果, 成为该领域的主流方法。首先, 对深度强化学习的基本概念和原理进行了详细介绍, 在此基础上, 系统地综述了 DRL 在金融市场量化交易中的应用和实践进展, 涵盖了基于策略的算法模型、基于价值的算法模型, 以及基于演员-评论家的算法模型等不同类型的 DRL 在金融市场量化交易中的应用。进一步探讨了 DRL 在金融市场量化交易中的优势, 指出其能够根据市场环境的动态变化调整交易策略, 以适应不同的市场环境。其次, 指出了 DRL 在金融市场量化交易中面临的挑战, 包括数据质量问题、模型稳定性问题、过拟合问题等。最后, 对 DRL 在金融市场量化交易领域未来的发展趋势进行了展望。随着算法的不断优化和计算能力的提升, DRL 将在金融市场量化交易领域发挥更加重要的作用, 为投资决策提供更加准确和可靠的支持。

关键词: 金融市场; 量化交易; 深度强化学习

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202439

Review: the application of deep reinforcement learning to quantitative trading in financial market

XU Bo^{1,2}, HE Yijun^{1,2}, WEN Jiancheng^{1,2}, LI Xiangxia^{1,2}

1. School of Information, Guangdong University of Finance & Economics, Guangzhou 510320, China

2. Guangdong Intelligent Business Engineering Technology Research Center, Guangzhou 510320, China

Abstract: As an effective learning paradigm to realize general artificial intelligence, deep reinforcement learning (DRL) has achieved significant results in a series of practical quantitative trading applications in financial market, becoming the mainstream method in this field. Firstly, a detailed introduction to the basic concepts and principles of deep reinforcement learning were provided. On this basis, a systematic review was conducted on the application and practical progress of DRL in quantitative trading in financial market, covering the application of different types of DRL, such as policy-based algorithm models, value-based algorithm models, and actor-critic algorithm models in quantitative trading in financial market. The advantages of DRL in quantitative trading in financial market were further explored, pointing out DRL could adjust trading strategies based on dynamic changes in the market environments to adapt to different market environments. Secondly, the challenges faced by DRL in quantitative trading in financial market were pointed out, including data quality issues, model stability issues, overfitting issues, etc. Finally, we outlooked the future development trend of DRL in the field of quantitative trading in financial market. It is believed that with the continuous optimization of algorithms and the

收稿日期: 2024-08-07; 修回日期: 2024-10-11

通信作者: 许波, xubo807127940@163.com

基金项目: 全国统计科学研究项目 (NO.2024LY075); 教育部人文社科项目 (NO.24YJAZH187); 广东省哲学社会科学规划项目 (NO.GD24CGL08)

Foundation Items: National Statistical Science Research Project (NO.2024LY075), The Ministry of Education of Humanities and Social Science Project (NO.24YJAZH187), Guangdong Philosophy and Social Science Planning Project (NO.GD24CGL08)

improvement of computing power, DRL will play a more important role in the field of quantitative trading in financial market, providing more accurate and reliable support for investment decisions.

Key words: financial market, quantitative trading, deep reinforcement learning

0 引言

金融市场充满挑战与机遇，吸引了众多投资者和研究者。随着信息技术的发展和市场交易的复杂化，量化交易成为金融领域的热点。量化交易通过数学模型、统计分析和计算机算法来执行交易策略，依赖于大量市场数据，如价格、成交量和交易历史。它的特点是系统性和高速性，交易决策基于预定义的规则和算法，在毫秒或微秒级别执行，要求快速计算和响应。在竞争激烈的市场中，固定策略常难以应对市场的复杂性，而深度强化学习（deep reinforcement learning, DRL）的出现为解决这一问题提供了更多可能性。

2014年以来，英仕曼集团旗下的AHL Dimension Programme基金通过运用深度学习与强化学习，取得了卓越的投资回报，其投资回报率达到行业平均水平的两倍。以强化学习为基础，2019年，乾象投资成功在中国期货市场实现了超过8的夏普比率，年化收益翻倍。2020年，该公司将研究框架扩展到中国A股和全球期货市场，初步研究结果显示，其研究框架在这两个市场也具备领先优势。类似地，美国的Wealthfront、英国的Money on Toast、德国的FinanceScout24、法国的MarieQuantier等公司成功将深度学习和强化学习技术引入投资理财，并且已在市场中占据重要地位。

深度强化学习结合了强化学习理论和深度神经网络的计算能力，为金融市场提供了一种强大的建模和决策工具，它不仅能自动化交易决策，还可以分析和预测大规模数据，有望改变传统金融交易方式通过优化交易策略、提高市场效率以及降低交易成本等，实现更智能、更高效的金融市场运作模式。金融市场的挑战包括快速变化的市场波动、新闻、事件对股价的影响，交易执行成本控制以及复杂的风险管理。这些挑战要求投资者具备智能化决策能力并迅速适应市场变化。深度强化学习可以根据实时数据和市场情绪动态调整策略，以适应不同的市场环境和风险水平。

本文综述了深度强化学习在金融市场量化交易中的应用，探讨其基本原理及在股票、期权、外

汇、加密货币市场中的应用。分析了深度强化学习在金融市场量化交易中的优势和挑战，并展望了其未来发展趋势。本文研究深度强化学习在金融市场量化交易中的应用和局限性，旨在帮助读者理解如何利用这一先进技术提高金融市场量化交易的效率、增强其可持续性，及更好地应对市场的不确定性和波动性，为金融从业者、学者和决策者提供有价值的见解，指导其利用深度强化学习应对金融市场的挑战。

1 深度强化学习的基本原理

强化学习的核心过程是让计算机从完全随机的操作开始，通过不断地尝试与试错来学习，最终发现规律并实现目标。在这个过程中，计算机不断更新自己的行为策略，逐步学会如何采取行动以获得高分。强化学习系统主要包括4个关键元素：智能体、环境、行动以及奖励。其主要目标是使智能体能够获得最大化的累计奖励。智能体与环境交互的整体工作流程包括动作选择和反馈处理。智能体根据当前的状态选择一个动作。随后，环境根据智能体所执行的动作提供反馈奖励，并更新环境状态。

如图1所示，强化学习受到行为主义心理学的启发，以马尔可夫决策过程^[1]为模型，构建出智能体与环境两大主体。智能体观察环境所处的状态，随后做出动作，获得反馈的奖励并进入下一个交互过程。在这个过程中，状态转移函数和折扣因子在智能体的最优决策中起到关键作用。状态转移函数定义了环境中状态的动态变化，指定了在特定状态和动作下系统转移到下一个状态的概率。折扣因子则平衡了即时奖励与长期回报，使智能体能够优先追求长远利益。通过与环境的持续交互，智能体以无监督方式学习并更新模型参数，目标是在每个状态下执行最优动作，从而最大化未来的累计奖励。

深度强化学习将深度神经网络与强化学习相结合，深度强化学习主要被用来处理感知-决策问题，已经成为人工智能领域重要的研究分支^[2]。其用神经网络来近似智能体的策略或值函数。这种组合使深度强化学习不仅能够有效解析高维输入数据，还能在复杂的环境中进行高效决策，从而在许

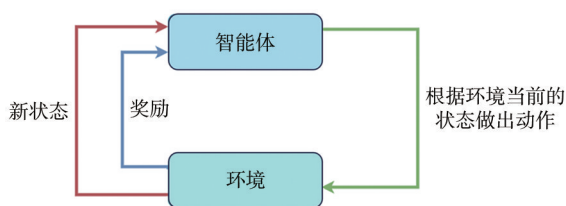


图1 智能体与环境交互

多任务上取得显著的成功。例如输入的是图像画面，则可以使用卷积神经网络，若想要考虑前后时间步之间的联系，输入时间序列相关数据，则可以使用循环神经网络。

因此，使用神经网络构建强化学习主体的方法就被称为深度强化学习。根据深度强化学习关注点的不同可以对深度强化学习进行一定程度上的分类。例如，深度Q网络（deep Q-network, DQN）算法^[3]将关注点放在回报上，使用数值 $Q^\pi(s, a)$ 表示特定状态下采取某行动的收益，即在某一状态 s 采取某一动作 a ，并将计算 Q 值的工作交由神经网络，其中， π 代表当前采用的策略，定义了智能体在每个状态下选择动作的方式，可以是确定的（即在每个状态下总是选择同一个动作）或随机的（即在每个状态下以一定的概率选择不同的动作）。其计算方式如式（1）所示。

$$Q(S_t, A_t, w) \leftarrow Q(S_t, A_t, w) + \alpha \left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a, w) - Q(S_t, A_t, w) \right] \quad (1)$$

其中， S_t 表示智能体在时间步 t 时所处的状态， A_t 表示智能体在时间步 t 时采取的动作， w 是神经网络的权重， R 表示智能体从状态 S_t 采取动作 A_t 并转移到下一个状态 S_{t+1} 后获得的即时奖励， Q 表示在当前策略下，从状态 S_t 采取动作 A_t 之后预期的长期奖励， π 代表当前采用的策略， α 为学习率（ $0 < \alpha \leq 1$ ）， r 是执行动作 a 后的及时奖励， γ 为折扣因子，且 $0 < \gamma \leq 1$ 。

在强化学习中，也可以将算法的关注点放在动作的选择上，如策略梯度算法，如果一个动作增加最终回报，则增加该动作出现的概率，反之则减小

该动作出现的概率。同时也能将两者相结合，如基于演员-评论家的算法模型，让选择动作的网络担任演员，称该网络为策略网络；关注回报的网络担任评论家，称该网络为价值网络。价值网络为策略网络的表现提供约束和指导，两个网络的更新如式（2）、式（3）所示。

$$\theta = \theta + \alpha \left(\nabla_{\theta} \log \pi_{\theta}(s_t, a_t) \right) Q(s, a) \quad (2)$$

式（2）描述了策略网络的参数 θ 通过策略函数的梯度乘以对应的 Q 值来更新。梯度 $\nabla_{\theta} \log \pi_{\theta}(s_t, a_t)$ 表示策略函数对参数 θ 的偏导数，它衡量了参数 θ 的变化对动作 a 概率的影响程度。

$$\phi = r + \gamma Q(s', a') - Q(s, a) \quad (3)$$

式（3）描述了价值网络的参数 ϕ 使用时间差分学习方法，通过比较当前估计值 $Q(s, a)$ 与目标值 $\gamma Q(s', a')$ 之间的差异进行更新，其中， s' 表示智能体在采取动作 a 后转移到的新状态， a' 表示智能体在下一个状态 s' 中可能采取的动作。

2 金融市场量化交易中的深度强化学习

金融产品在二级市场的量化交易一直是国内外学者关注的热点，而将深度强化学习与量化交易技术相结合更是备受瞩目。目前，深度强化学习可以按照不同的角度进行分类，其中一种常见的分类方法见表1。

2.1 基于策略的算法模型

在强化学习中，基于策略的算法模型是一种常见的方法，用于解决智能体在与环境交互的过程中学习最优策略的问题。基于策略的算法模型的核心思想是通过优化策略函数来寻找最优策略。策略函数是将每个状态映射到动作的概率分布。通过调整策略函数的参数，可以改变智能体在不同状态下选择动作的概率。常见的算法包括策略梯度算法等。如图2所示，策略梯度算法的输出是一种概率分布，需要通过采样方法按照给定的概率分布来获得输出的动作。

基于策略的算法模型有多种实现方式，其中最常见的是策略梯度算法，该方法通过计算策略函数

表1 DRL 分类

分类	代表算法	特点
基于策略的算法模型	策略梯度算法	通过神经网络近似策略函数
基于价值的算法模型	DQN、SARSA 算法	通过神经网络近似最优动作价值函数
基于演员-评论家的算法模型	SAC、DDPG、OAC 算法	结合了策略学习和价值学习

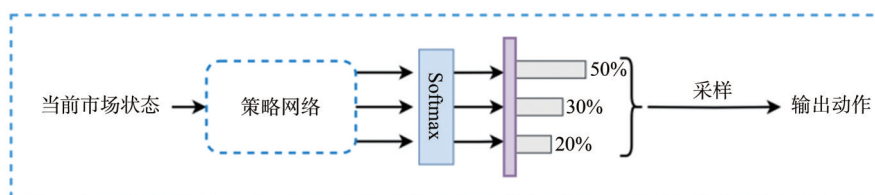


图2 策略梯度算法结构

对参数的梯度，并沿梯度方向更新参数以最大化期望回报。基于策略的算法模型可以直接处理连续动作空间，而值函数方法通常局限于离散动作空间。此外，基于策略的算法模型能学习随机策略，在探索未知环境、应对非确定性和部分可观测环境时表现更优。然而，基于策略的算法模型也存在缺陷。基于策略的算法模型直接优化策略函数，通常比值函数方法更难收敛，并且在高维状态空间下可能面临维度灾难问题。

2.1.1 投资组合优化

在金融资产分配中，滚动窗口的选择对于优化投资组合至关重要。Gašperov等^[4]首次将深度强化学习用于动态确定最小方差组合的滚动窗口大小，采用基于协方差矩阵差异范数的状态空间表示。相比传统协方差矩阵估计方法，如因子模型和收缩估计算法，该方法不是直接清除噪声，而是结合强化学习实现灵活的窗口选择，并通过遗传算法优化神经网络权重。Han等^[5]进一步研究将投资环境划分为静态市场状态和动态投资组合状态，设计了基于概率动态规划的投资回报机制，增强了深度强化学习代理的环境适应性，使其更高效地学习，降低探索成本，提升策略性能和稳健性。

Borrageiro等^[6]基于策略梯度，探讨在线回归迁移学习，使用高斯混合模型的径向基函数网络作为特征表示，应用于递归强化学习代理。Mattera等^[7]介绍一种强化学习方法，用于估计高维协方差矩阵以优化投资组合。笔者提出基于策略梯度的智能体和基于循环策略梯度的智能体，这两种智能体使用神经网络近似最优策略，即选择最佳收缩强度参数构造收缩协方差矩阵。收缩协方差矩阵是一种改进的协方差估计方法，通过将样本协方差矩阵与目标矩阵进行凸组合，降低估计误差并提高可逆性。笔者还分析了强化学习模型的可解释性和稳健性。

在高频且维度丰富的数字货币交易环境下，强化学习能够更精确地执行资产配置任务，进而提升投资回报率并增强风险管理能力。Liu等^[8]提出了

一种名为长序列特征提取器与跨资产注意力网络（long sequence representation extractor and cross-asset attention network, LSRE-CAAN）的强化学习驱动资产配置策略优化方案。该方案的核心是采用策略梯度算法进行优化计算，并通过深度挖掘历史交易数据以构造状态转移矩阵和行动价值函数，以此达到对资产组合的动态调整。其中，LSRE-CAAN创新性地融合了长期序列表示提取技术和跨资产注意力网络机制，能够有效地捕捉并刻画数字货币市场中各币种之间复杂的联动效应。

2.1.2 交易策略优化

为了减少市场噪声并提高代理表现，Ding等^[9]使用 K 线分解和聚类预处理金融数据，提取特征并通过基于策略的算法模型优化交易决策。模型利用深度神经网络从连续数据中学习动作分布，并选择最优动作。赵淼等^[10]基于动态选择预测器的强化学习模型，解决了数据源和股票预测模式单一的问题。Duan等^[11]开发了新算法OASS（optimal adaptive step size），将交易视为序列决策问题，使用多层长短期记忆网络建模环境状态的时间依赖性，并用概率动态规划处理高噪声和不确定性，无须复杂奖励设计。

与此同时，近端策略优化（proximal policy optimization, PPO）因其稳定性和广泛适用性受到学术界关注。2023年，Li等^[12]提出基于PPO的比特币交易框架，将交易过程视为动作，将回报视为奖励，将价格视为状态，利用长短期记忆网络构建策略函数，展示了加密货币交易策略的潜力。2022年，Chung等^[13]在订单堆叠框架中基于PPO开发市场交易策略，使代理适应不同市场环境。2023年，Chen等^[14]设计了适用于中国期货市场的PPO交易策略，实现从数据到交易行为的端到端决策，结果显示该策略在不同市场环境下适应性均较强。

为规避金融市场量化交易中的非对称性风险，Alameer等^[15]首次将条件风险价值引入DRL，利用条件风险价值的凸性特性保证问题的凸性。然而，

高维数据处理中仍需改进。Cui等^[16]提出结合条件风险价值和PPO的新方法，用于大尾部风险的加密货币市场投资组合构建。笔者阐述了条件风险价值和风险价值的计算方法，强调条件风险价值在捕捉尾部风险事件中的优势。通过PPO与环境交互，不断更新策略参数，实现最小化组合风险的目标。

上述文献系统性地综述了策略驱动的深度强化学习方法在金融市场量化交易领域中的应用及其最新进展，涉及策略梯度算法、长短期记忆网络、PPO等，并在不同问题上展示了它们的性能。然而一些文献虽然通过引入条件风险价值来更好地捕捉尾部风险事件，但是也可能增加一些计算和实现上的复杂性，以及对未来市场环境的假设，从而进一步增加模型对市场观测的困难。此外，PPO因其更新策略的稳定性已经被实业界证实可用于大语言模型的训练，但在面对金融市场这个噪声与波动性都远超普通对话模式的环境，PPO仍然需要进一步的研究。上述基于策略的算法模型的局限性分析见表2。

2.2 基于价值的算法模型

基于价值的算法模型的核心思想是通过优化值函数来寻找最优策略，主要关注如何评估每个状态或状态-动作对的价值，以便在决策时选择最优的动作。如图3所示，DQN算法会选择当前状态下交易动作的价值最大者作为动作决策输出，是一个

确定性的决策。常见的算法包括DQN、SARSA算法等。

值函数可以分为状态值函数（也称为 V 函数）和动作值函数（也称为 Q 函数）2种类型。 V 函数衡量在给定策略下，智能体处于某个状态的长期回报期望值。 Q 函数衡量在给定策略下，智能体在某个状态下选择某个动作的长期回报期望值。基于价值的算法模型通过学习值函数实现离线规划，无须与环境交互。然而，在高维状态空间下，基于价值的算法模型易遭遇维度灾难。此外，基于价值的算法模型通常适用于确定性环境，而在非确定性和部分可观测环境中则需进行扩展。

2.2.1 投资组合优化

为了控制资产价格的走势，并通过估计市场状态来跟随价格趋势。许杰等^[17]提出了一种融合了卷积神经网络与长短期记忆网络的端到端深度强化学习自动化交易算法，该算法能够感知股票动态市场条件并抽取动态特征，循环学习动态时间序列规律，并通过强化学习方法累积最终收益并优化投资组合。通过将经验回放机制、Double DQN算法与Dueling DQN算法相结合，以及将优势函数和价值函数分离等技术手段，提高模型的学习效率和稳定性。相比于传统的DQN算法，Dueling DQN算法在处理高维状态空间时具有更好的性能表现，因此对于市场情况更加复杂的数字货币市场，Suliman等^[18]采用了Dueling DQN算法。该算法可以分别估计状

表2 基于策略的算法模型的局限性分析

模型	模型局限性分析
自适应滚动窗口选择 ^[4] 、基于概率动态规划的投资回报机制 ^[5] 、长序列特征提取器与跨资产注意力网络 ^[8] 、基于策略的算法模型优化交易决策 ^[9] 、基于动态选择预测器的强化学习模型 ^[10] 、长短期记忆网络构建策略函数 ^[12] 、PPO算法市场策略 ^[13]	只使用了一种状态特征，而没有考虑其他可能影响协方差矩阵估计的因素，如宏观经济指标、利率数据、收益率曲线数据等
基于策略梯度探讨在线回归迁移学习 ^[6] 、OASS算法 ^[11]	可能对梯度估计的方差和噪声敏感，导致收敛速度慢或陷入局部最优
基于策略梯度的智能体和基于循环策略梯度的智能体 ^[7] 、基于条件风险价值的强化学习算法 ^[15]	使用了简单的多层感知器网络作为策略近似，可能无法捕捉高维协方差矩阵的复杂特征和结构，并且缺乏正则化或降维的机制，可能导致过拟合或维数灾难
基于端到端的PPO算法交易策略 ^[14] 、基于条件风险价值的PPO算法 ^[16]	使用了一个共享的全连接层来处理不同的输入特征，可能导致特征之间的相互干扰或信息损失

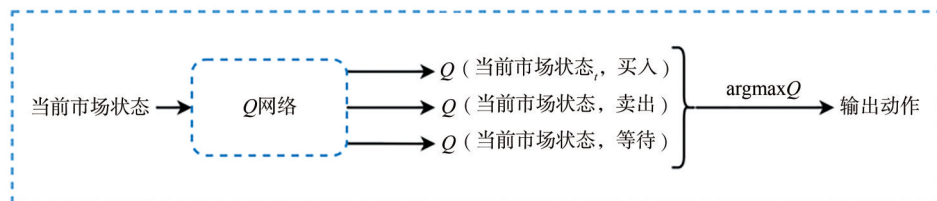


图3 DQN算法结构

态值函数和每个动作的优势函数,从而减少 Q 值的过估计问题。笔者展示了Dueling DQN算法在测试期间的表现,结果显示,DRL可以成功地产生盈利的投资组合,但是在3种加密货币上都不如基准策略。笔者认为这可能是加密货币市场的高波动性和其他外部因素造成的,并指出了未来可能的改进方向。Wang等^[19]提出了一种结合转折点分类和强化学习的股票交易策略,在转折点分类模块中,通过使用SMOTE(synthetic minority oversampling technique)进行过采样解决了样本不平衡问题,并采用轻量级梯度提升机算法进行了分类。笔者在强化学习框架中,采用了经验回放和迭代机制来解决复杂问题,并且加入了Double DQN和Dueling DQN算法来优化神经网络结构。通过转折点分类与强化学习框架相结合,能够更好地适应市场变化,提高交易收益。

2.2.2 交易策略优化

为提升智能体在金融市场的决策能力,杨兴雨等^[20]将行为克隆融入深度强化学习,结合Dueling DQN算法,让智能体既自主探索又模仿专家决策,增强了策略的适应性与可靠性。

为了使模型能更好地适应金融市场环境, Ma等^[21]在传统的 Q -Learning算法的基础上,提出了并行多模块深度强化学习算法。该算法通过引入两个深度神经网络分别负责动作选择和价值评估,有效解决了传统 Q -Learning算法中的过度乐观估计问题,显著提升了算法的稳定性和交易性能。这一创新不仅优化了算法结构,还强调了在复杂金融环境中精准决策的重要性。Théate等^[22]针对股票交易,设计了交易深度 Q 网络(trading deep Q -network, TDQN)算法,TDQN算法利用人工轨迹从有限的历史数据中学习,专攻夏普比率优化,利用DQN算法及深度神经网络逼近状态-动作值函数,通过人工轨迹训练,提升了交易策略在复杂市场环境中的表现。

Liu等^[23]为了进一步准确地捕捉股票市场的动态变化以及降低交易成本等,提出了一种多类型数据融合框架。该框架综合了股票数据、技术指标和蜡烛图的信息,利用不同的深度神经网络提取和整合不同数据源的时序特征,从而更好地反映股市环境状态,并学习最优的交易策略。

在风险和市场摩擦情况下,构建复制策略是关键问题。Cannelli等^[24]提出了一种基于强化学习的

AI建模方法。由于现实中连续复制不可行且成本高,笔者将对冲过程建模为动态规划问题,并使用 Q -Learning算法等强化学习求解最优策略。此外,引入风险规避概念,使用均值方差模型衡量风险并优化对冲策略。

面对复杂市场环境和非线性交易关系,以原油和天然气期货市场为例,Huang等^[25]将双分支深度 Q 网络应用于期货交易,以自动生成稳定盈利和鲁棒的交易信号。笔者综合考虑短期和长期收益,以及各因素与收益的相关性,设计了基于相关性的奖励函数,并使用折扣因子调节不同时间点的奖励值,使奖励函数反映短期和长期收益的平衡。Zhao等^[26]提出结合中位数绝对偏差方法与 Q -Learning算法的股票预测方法。通过中位数绝对偏差方法对数据预处理并合理设置奖励值,提升 Q -Learning算法的预测性能和实际效果,同时提高股票交易决策的效率和准确性。

上述研究主要集中在对传统 Q -Learning算法进行扩展,通过整合历史市场数据、技术指标、评论数据等多源信息,提出了不同的模型架构。但对于如何处理维数灾难、如何应对非确定金融市场环境都需要进一步探索和研究,以开发出更具泛化性和鲁棒性的算法模型。上述基于价值的算法模型的局限性分析见表3。

2.3 基于演员-评论家的算法模型

基于演员-评论家的算法模型是由Konda等^[27]提出的一个算法模型,它通过两个网络分别进行决策和评估,以实现对环境和学习、优化。基于演员-评论家的算法模型是一种结合策略梯度算法和时序差分学习的强化学习方法。其中,演员是指学习一个策略来得到尽量高的回报。评论家是指对当前策略的值函数进行估计,即评估演员的好坏。常见的算法包括柔性演员-评论家(soft actor-critic, SAC)算法、乐观的演员-评论家(optimistic actor-critic, OAC)算法、深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法等。DDPG算法的结构如图4所示,其中演员学习一个策略函数,评论家学习一个值函数,两者相互协作以提高决策的准确性。这种算法通常被用于解决连续动作空间的问题,因此用于金融市场中的策略生成、价值评估、实时交易决策、风险管理等行为比较合适。

表3 基于价值的算法模型的局限性分析

模型	局限性分析
基于卷积神经网络与长短期记忆网络的自动化交易算法 ^[17] 、基于动态规划的Q-Learning算法 ^[24]	模型具有一定的局限性,不能处理交易决策对市场的影响或多期规划问题
Dueling DQN算法 ^[18]	该模型的结果被买入持有策略所超越,说明该模型不能很好地识别牛市趋势
结合转折点分类和强化学习的股票交易策略 ^[19] 、基于行为克隆的Dueling DQN算法 ^[20] 、并行多模块的Q-Learning算法 ^[21] 、TDQN算法 ^[22]	模型过度依赖转折点的分类结果,如果转折点的识别不准确,可能会影响交易策略的效果
多类型数据融合框架 ^[23] 、基于中位数绝对偏差方法的Q-Learning算法 ^[26]	模型的输入数据包括股票数据、技术指标和蜡烛图,这些数据的结构和规模不同,可能导致网络结构无法充分利用多种数据源和特征
双分支深度Q网络 ^[25]	其使用二元交叉熵损失函数,这可能会导致类别不平衡,即某些网格上的到达方向出现的频率远高于其他网格上到达方向出现的频率

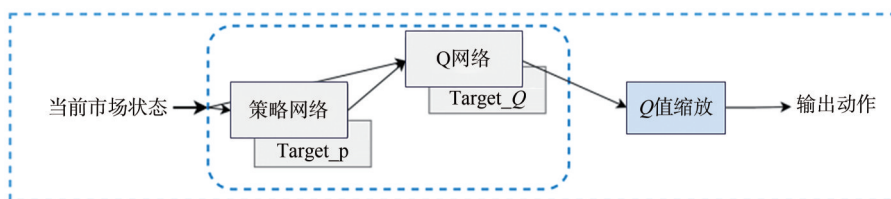


图4 DDPG算法的结构

2.3.1 投资组合优化

为了解决传统投资组合管理方法忽略系统性风险的问题, Xu等^[28]提出了一种基于配对交易和深度强化学习的动态、市场中性、风险分散的投资组合管理模型。通过将投资组合管理建模为马尔可夫决策过程, 使用深度强化学习自动调整投资组合的权重和配置, 从而有效应对市场波动, 实现收益最大化和风险最小化。此外, 该模型还能分散非系统性风险, 降低投资组合的整体风险水平。

随后, Gao等^[29]提出了类似的在线最优投资组合模型, 利用深度强化学习和DDPG算法解决了传统方法在复杂多变市场中的适应性问题, 提升了投资组合管理的效率和准确性。Lim等^[30]则探讨了通过强化学习进行动态投资组合再平衡的方法, 采用渐进式组合再平衡策略, 避免了价格预测, 使投资组合更稳定且风险更低。Koratamaddi等^[31]在DDPG算法的基础上, 引入市场情感信息并微调奖励函数, 使交易代理能够根据市场情感进行操作。这种方法通过考虑市场情绪的变化, 调整投资组合, 以适应不同的市场趋势。

为了解决资产组合中非线性相关的建模问题, Zhao等^[32]提出了一个新的资产投资政策网络, 通过使用时序卷积网络提取资产价格特征, 并在新序列上执行新的点积操作来建模资产之间的相关性, 保留时间信息并降低注意力机制的时间复杂度。Weng等^[33]则利用深度强化学习和注意力机制进行

数字货币投资组合交易, 通过多维注意力门控机制增强模型对不同特征的关注和权重调整能力。

为进一步优化投资组合, Murray等^[34]提出了一种用于寻找任意初始投资组合和市场状态下的最优对冲策略的方法, 使用基于演员-评论家的算法模型解决一般风险厌恶的随机控制问题, 同时学习多个风险厌恶水平下的对冲策略。为避免风险偏好模型中的数值不稳定问题, 基于演员-评论家的算法模型引入了平滑的价值函数更新方法, 从而在高风险厌恶水平下提升了状态值估计的准确性。Wang等^[35]定义了股票投资组合交易的马尔可夫决策过程, 并引入熵改进深度确定性策略梯度算法。通过在优化过程中增加随机性, 促进模型对未知动作的探索, 帮助算法在探索与利用之间找到更好的平衡点, 并使训练过程更加平滑和可靠。Huang等^[36]结合DDPG算法和自适应采样策略用于连续投资组合优化, 提高采样效率, 使算法更关注对策略改进贡献大的样本数据, 加速算法收敛速度。

Jang等^[37]将现代投资组合理论与DDPG算法相结合, 通过Tucker分解解决多模态问题, 与其他投资组合算法相比, 该方法不仅考虑历史数据和技术指标, 还考虑资产间的相关性和动态变化, 从而提高预测精度和投资回报率。

为了解决集成方法中滚动窗口带来的滞后性问题, 冀中等^[38]提出了一种双层嵌套的集成深度强化学习策略。该策略对PPO、优势演员-评论家算

法和DDPG算法代理进行两层嵌套处理：第一层评估每个阶段的最优模型；第二层则通过加权并行集成，整合这3种代理在各个阶段的优势。

2.3.2 交易策略优化

为了优化黄金和比特币的投资组合决策，Yuan等^[39]利用深度强化学习分析和建模了其价格趋势。随后，Luthfianti等^[40]对LQ45指数的股票进行投资组合配置，比较了深度强化学习分别与等权策略和均值-方差相结合的应用效果，发现股票数量增多可能导致表现下降，这归因于收益模式的多样性和相互抵消。

Sun等^[41]通过动作反馈机制和监督学习标签纠正深度强化学习中的部分履行情况。为了处理股市中的非线性因素和噪声干扰，Zhu^[42]提出了一种量化交易策略——增强型深度确定性策略梯度（enhanced deep deterministic policy gradient, EDPG）算法，该算法结合了门控循环单元和DDPG算法的优点，利用深度神经网络在高维状态和动作空间中达到了良好的学习效果。Chen等^[43]开发了一种新方法，利用多尺度图案捕捉金融时间序列中的模式和趋势，并将DDPG算法应用于这些特征以优化交易策略。

在金融市场的量化交易中，基于演员-评论家的算法模型被应用于投资组合优化和交易策略优化。在投资组合优化中，研究者们利用基于演员-评论家的算法模型的强化学习框架实现动态、市场中性、风险分散的管理。通过将投资组合管理建模为马尔可夫决策过程，策略网络制定交易策略，价值网络评估策略性能，力图在不同市场环境下最大

化收益和最小化风险。

交易策略优化关注如何通过基于演员-评论家的算法模型提高交易策略性能，包括对实时市场数据的敏感处理和有效的风险管理。然而，基于演员-评论家的算法模型在金融市场的实时性要求中可能受到延迟限制，难以及时响应市场变化。此外，模型容易在历史数据上过拟合，而在真实市场中表现不佳。实际交易中的成本、滑点等细节难以被准确模拟，导致模型性能下降。上述基于演员-评论家的算法模型的局限性分析见表4。

2.4 其他类型的强化学习算法模型

深度强化学习在金融市场领域取得了显著的研究进展与应用突破，研究者们不仅在基于策略、价值和演员-评论家的算法模型中取得成果，还在探索其他算法模型以解决复杂的金融问题。多智能体强化学习能够考虑多个智能体之间的交互，在建模和理解复杂金融系统方面具有优势。Lin等^[44]提出了一种基于多智能体深度强化学习的风险转移投资组合管理方法，包括策略网络和价值网络。实验表明，该方法在风险转移方面表现优异，能通过学习和协作实现投资组合的有效分散与风险转移，增加投资组合的回报，增强其稳定性。

Ma等^[45]将趋势一致性正则化方法与多智能体强化学习相结合，以优化投资组合的决策策略，并通过趋势一致性正则化方法提高投资组合的稳定性。龙杰等^[46]提出了一种集成的深度强化学习投资组合模型，采用多智能体方法构建多个基智能体，并设计不同交易风格的奖励函数，利用集成学习方法对基智能体的策略网络进行特征融合，并嵌

表4 基于演员-评论家的算法模型的局限性分析

模型	局限性分析
基于配对交易和深度强化学习的交易模型 ^[28]	模型的网络结构较复杂，需要训练 $M+1$ 个预测器，可能导致计算效率和资源消耗较高
在线最优投资组合模型 ^[29]	提出的两种网络结构都使用了一维卷积层来处理多维状态数据，但没有考虑到不同特征之间的交互作用和复杂性
渐进式组合再平衡策略 ^[30] 、结合市场情感信息并微调奖励函数的DDPG算法 ^[31]	模型只能处理离散的状态和动作空间，而不能处理连续的状态和动作空间，限制了模型的泛化能力和灵活性
资产投资政策网络 ^[32] 、基于演员-评论家的算法模型 ^[34] 、结合DDPG算法和自适应采样策略 ^[36] 、基于动作反馈机制和监督学习标签的深度强化学习模型 ^[41]	没有考虑资产价格的多尺度特征，将环境视为一个静态的代理属性，而不考虑其可能随着状态而变化的情况
基于多维注意力门控机制的深度强化学习模型 ^[33] ，引入熵改进深度确定性策略梯度算法 ^[35] ，基于Tucker分解的DDPG算法 ^[37] ，双层嵌套的集成深度强化学习策略 ^[38] ，价格趋势模型 ^[39] ，基于等权、均值-方差的深度强化学习模型 ^[40] ，增强型深度确定性策略梯度算法 ^[42] ，利用多尺度图案捕捉金融时间序列的DDPG算法 ^[43]	没有详细说明如何计算技术指标和协方差矩阵，没有讨论模型可能面临的挑战和风险

入基于卷积块注意力模块的趋势预测网络,以引导集成策略网络进行自适应交易比重选择。

奖励函数设计是强化学习中的关键任务,因为它直接影响智能体的学习策略。在金融资产分配中,奖励函数同样重要。Belyakov等^[47]将投资组合管理视为一个部分可观察的马尔可夫决策过程,对比了两种奖励函数(投资组合收益和差分夏普比率)及两种强化学习算法(PPO和双延迟深度确定性策略梯度)在训练交易代理中的应用。Zhang等^[48]通过优化投资组合收益来训练比例投资组合网络,使用新的成本敏感奖励函数考虑交易成本和风险成本,以保证实践中具有近似最优解。在目标不明确的情况下,逆强化学习非常重要。Zhang等^[49]提出了一种基于强化学习的金融预测模型优化方法,使用逆强化学习设计奖励函数,准确估计模型利润。

Wu等^[50]提出了门控深度Q-Learning交易策略和门控确定性策略渐进交易策略两种强化学习交易策略,在动荡股市中表现出稳定的收益。Olschewski等^[51]提出了值函数组合强化学习,该方法通过构建资产变异性及相关性模型,将这些关键因素纳入强化学习框架中,以权衡不同的投资选择。Nicolas等^[52]介绍了一种深层次分层强化学习方法,用于在共享资源系统中权衡公共池资源的可持续性与伦理性。

Gašperov等^[53]提出了一种无模型深度强化学习信号框架进行市场做市,并利用信号改进交易策略。基于此,Song等^[54]提出了一个从确定性到随机的可解释无模型强化学习框架,用于投资组合优化,结合无模型强化学习和随机模型概念,通过生成收益概率分布优化投资策略,提供应对复杂市场环境的可解释性。Ye等^[55]提出了状态增强强化学习(state augmented reinforcement learning, SARL)框架,以解决金融投资中的数据异质性和环境不确定性问题。该框架结合监督学习和强化学习,通过端到端网络提取资产价格信息和外部新闻数据,增强状态表示,并通过确定性策略梯度算法优化资产配置。

图神经网络在异质信息整合与复杂关系建模方面具有独特优势。Li等^[56]提出了一种结合超图与强化学习的新颖框架,利用超图节点、边信息以及强化学习奖励信号进行投资决策策略学习,提升投资组合选择的效果和收益。卜湛等^[57]在深度图神经网络基础上,提出了动作评估反馈的深度强化学习方法,融入股票关联网络层次化表征和市场运行

状态,以优化环境状态学习。Wang等^[58]使用时序卷积网络处理长期序列数据,结合空间注意力机制和图卷积网络捕捉股票间相关性。Soleymani等^[59]提出了基于深度强化学习的投资组合管理框架,利用金融工具间的时变关系进行管理,采用图卷积神经网络提取资产间相关性特征,以更好地预测市场趋势。Shi等^[60]提出了基于图卷积网络的强化学习投资组合管理方法,通过多尺度卷积操作提取时间序列特征,引入温度参数平衡风险与收益。

Yang等^[61]研究了一种交易点感知的投资组合优化问题,提出了Strader交易策略,采用连续正则化流技术及神经常微分方程,实现价格向量到潜在变量分布的映射,以获取超额日内利润。Niu等^[62]提出了MetaTrader框架,用于投资组合优化,通过整合多个策略并学习如何选择与组合策略,优化投资组合。上述算法模型的局限性分析见表5。

3 结束语

本文介绍了深度强化学习的基本概念和相较于传统机器学习和强化学习的优势。文章从基于策略、价值和演员-评论家的算法模型3个角度总结了在金融市场量化交易中的典型深度强化学习算法,并讨论了这些算法在股票、期货、期权、外汇和加密货币市场的应用。不同算法模型被分类和比较:基于策略的算法模型能直接优化策略,适用于连续动作空间,但训练中可能面临收敛性和稳定性问题;基于价值的算法模型通过经验回放和目标网络提高训练效果,具备稳定性且易于收敛;基于演员-评论家的算法模型结合了策略优化和价值评估的优点,但需要更多的计算资源和训练样本。

每种算法模型都有其适用场景和特点,算法模型的选择取决于具体问题、市场特征和可用数据等因素。在金融市场量化交易中,深度强化学习面临诸多挑战,需要进一步研究和改进,如提高数据质量和模型泛化能力、增强鲁棒性,以及平衡收益与风险。聚焦于如何优化这些关键领域,以促进深度强化学习在金融市场的更广泛应用和提升有效性。

(1) 适用于金融市场的新型深度强化学习方法

现有的强化学习模型通常过于理想化,难以应对复杂的现实环境,因此改进深度强化学习至关重要。如通过使用联邦强化学习,可以在不共享原始数据的情况下进行模型训练,从而保护用户隐私。同时,分布式学习允许在多个设备或节点上并行处

表5 其他类型的算法模型的局限性分析

模型	模型局限性分析
多智能体深度强化学习 ^[44] 、部分可观察的马尔可夫决策 ^[47] 、门控深度 Q -Learning 交易策略和门控确定性策略渐进交易策略 ^[50] 、结合空间注意力机制和图卷积网络 ^[58] 、Strader 交易策略 ^[61]	模型虽然考虑到数据质量、数据偏差、过拟合、计算复杂度等问题,但没有提供有效的模型验证和解释方法
基于趋势一致性正则化方法的多智能体强化学习 ^[45] 、深层次分层强化学习模型 ^[52] 、可解释无模型强化学习框架 ^[54] 、SARL 框架 ^[55] 、结合超图与强化学习模型 ^[56] 、基于深度强化学习的投资组合管理框架 ^[59] 、MetaTrader 框架 ^[62]	模型忽略了股票价格之间的相关性或协整性,以及不同股票对市场波动的敏感度或贝塔值
集成的深度强化学习投资组合模型 ^[46] 、动作评估反馈的深度强化学习方法 ^[57]	模型涉及多个神经网络的构建和训练,没有对模型的复杂度和计算效率进行分析,可能需要较多的时间和资源来进行优化和更新
Proportional Portfolio Network ^[48]	模型策略梯度算法在高维连续行动空间的优化稳定性欠佳
基于强化学习的金融预测模型优化方法 ^[49]	模型上置信界算法是一种贪心的策略选择方法,它可能会忽略一些潜在的高收益但低概率的策略
值函数组合强化学习模型 ^[51] 、基于图卷积网络的强化学习资产组合管理方法 ^[60]	模型使用了简单的代理变量来近似资产的方差和相关性,对高质量数据强依赖、过拟合风险、关系复杂性捕捉不足
无模型深度强化学习信号框架 ^[53]	模型没有详细说明对抗性代理的设计原理和优化目标,也没有分析对抗性代理对市场做市者代理策略的影响和意义

理数据,完成计算任务,进一步提高训练效率和系统的可扩展性。逆强化学习通过观察行为推断代理目标,自主推定奖励函数。结合智能化金融科技创新监管工具可及时定位、跟踪、预防风险隐患,确保金融科技创新应用风险的可知、可析、可控^[63]。这些方法旨在应对传统强化学习在实际应用中的挑战,希望未来能进一步研究这些模型,以更好地满足行业需求。

(2) 高效的金融数据预处理和特征提取方法

由于金融市场中的数据具有高维、非线性、非平稳、噪声大等特点,如何有效地处理和提取数据中的有用信息,是影响深度强化学习性能和效果的关键因素。因此,未来需要研究更多的数据预处理方法和特征提取方法,例如使用自编码器、卷积神经网络、注意力机制等进行数据降维、特征抽取、特征选择等操作,以提高数据质量和表达能力,同时降低计算复杂度和内存消耗。

(3) 深度强化学习模型的测试应该增加更多实时交易测试

实时交易测试可以真实模拟市场的不确定因素,评估交易策略的鲁棒性。实际交易中,网络、交易所和执行延迟等因素会影响策略表现,这些在回溯测试中无法被完全模拟。实时测试有助于识别和优化延迟问题。此外,市场中存在冰山订单、时间加权平均价格订单、比例成交算法订单等非标准订单,导致不同市场行为和价格波动。通过实时测试,可以更好地评估和优化策略在处理非标准订单

时的表现。

(4) 深度强化学习模型的泛化能力和鲁棒性

由于深度强化学习模型通常涉及多个神经网络和参数,如何设计合适的网络结构和参数设置以适应不同的金融市场环境和任务,是提高模型泛化能力和鲁棒性的关键。因此,未来需要研究更多的模型设计方法和优化方法,例如使用元学习、多任务学习、迁移学习、对抗学习等进行模型初始化、模型更新、模型适应、模型对抗等操作,以提高模型的学习效率和收敛速度,同时避免过拟合和欠拟合等问题。

参考文献:

- [1] FILAR J, VRIEZE K. Competitive markov decision processes[M]. New York: Springer, 1996.
- [2] 刘朝阳, 穆朝絮, 孙长银. 深度强化学习算法与应用研究现状综述[J]. 智能科学与技术学报, 2020, 2(4): 314-326.
LIU Z Y, MU C X, SUN C Y. An overview on algorithms and applications of deep reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 314-326.
- [3] HESTER T, VECERIK M, PIETQUIN O, et al. Deep Q -learning from demonstrations[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1).
- [4] GAŠPEROV B, ŠARIĆ F, BEGUŠIĆ S, et al. Adaptive rolling window selection for minimum variance portfolio estimation based on reinforcement learning[C]//Proceedings of the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). Piscataway: IEEE Press, 2020: 1098-1102.
- [5] HAN L, DING N, WANG G, et al. Efficient continuous space policy optimization for high-frequency trading[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2023: 4112-4122.

- [6] BORRAGEIRO G, FIROOZY N, BARUCCA P. Reinforcement learning for systematic FX trading[J]. *IEEE Access*, 2022, 10: 5024-5036.
- [7] MATTERA G, MATTERA R. Shrinkage estimation with reinforcement learning of large variance matrices for portfolio selection[J]. *Intelligent Systems with Applications*, 2023, 17: 200181.
- [8] LIU F R, LI Y, LI B T, et al. Bitcoin transaction strategy construction based on deep reinforcement learning[J]. *Applied Soft Computing*, 2021, 113: 107952.
- [9] DING F Q, LUO C. An adaptive financial trading system using deep reinforcement learning with candlestick decomposing features[J]. *IEEE Access*, 2020, 8: 63666-63678.
- [10] 赵淼, 谢良, 林文静, 等. 基于动态选择预测器的深度强化学习投资组合模型[J]. *计算机科学*, 2023: 1-14.
ZHAO M, XIE L, LIN W J, et al. A deep reinforcement learning portfolio model based on dynamic selectors[J]. *Computer Science*, 2023: 1-14.
- [11] DUAN Z J, CHEN C, CHENG D W, et al. Optimal action space search: an effective deep reinforcement learning method for algorithmic trading[C]//*Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. New York: ACM, 2022: 406-415.
- [12] LI J H, ZHANG Y, YANG X Y, et al. Online portfolio management via deep reinforcement learning with high-frequency data[J]. *Information Processing & Management*, 2023, 60(3): 103247.
- [13] CHUNG G, CHUNG M, LEE Y, et al. Market making under order stacking framework: a deep reinforcement learning approach[C]//*Proceedings of the Third ACM International Conference on AI in Finance*. New York: ACM, 2022: 223-231.
- [14] CHEN X M, GUO H R. A futures quantitative trading strategy based on a deep reinforcement learning algorithm[C]//*Proceedings of the 2023 IEEE 8th International Conference on Big Data Analytics (ICBDA)*. Piscataway: IEEE Press, 2023: 175-179.
- [15] ALAMEER A, AL SHEHRI K. Conditional value-at-risk for quantitative trading: a direct reinforcement learning approach[C]//*Proceedings of the 2022 IEEE Conference on Control Technology and Applications (CCTA)*. Piscataway: IEEE Press, 2022: 1208-1213.
- [16] CUI T X, DING S S, JIN H, et al. Portfolio constructions in cryptocurrency market: a CVaR-based deep reinforcement learning approach[J]. *Economic Modelling*, 2023, 119: 106078.
- [17] 许杰, 祝玉坤, 邢春晓. 基于深度强化学习的金融交易算法研究[J]. *计算机工程与应用*, 2022, 58(7): 276-285.
XU J, ZHU Y K, XING C X. Research on financial trading algorithm based on deep reinforcement learning[J]. *Computer Engineering and Applications*, 2022, 58(7): 276-285.
- [18] SULIMAN U, VAN ZYL T L, PASKARAMOORTHY A. Cryptocurrency trading agent using deep reinforcement learning[C]//*Proceedings of the 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. Piscataway: IEEE Press, 2022: 6-10.
- [19] WANG J J, JING F, HE M L. Stock trading strategy of reinforcement learning driven by turning point classification[J]. *Neural Processing Letters*, 2023, 55(3): 3489-3508.
- [20] 杨兴雨, 陈亮威, 郑萧腾, 等. 考虑行为克隆的深度强化学习股票交易策略[J]. *系统管理*, 2023: 1-17.
YANG X Y, CHEN L W, ZHENG X T, et al. Stock trading strategy via deep reinforcement learning with behavior cloning[J]. *Journal of Systems & Management*, 2023: 1-17.
- [21] MA C, ZHANG J S, LIU J M, et al. A parallel multi-module deep reinforcement learning algorithm for stock trading[J]. *Neurocomputing*, 2021, 449: 290-302.
- [22] THÉATE T, ERNST D. An application of deep reinforcement learning to algorithmic trading[J]. *Expert Systems with Applications*, 2021, 173: 114632.
- [23] LIU P P, ZHANG Y F, BAO F X, et al. Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading[J]. *Applied Intelligence*, 2023, 53(2): 1683-1706.
- [24] CANNELLI L, NUTI G, SALA M, et al. Hedging using reinforcement learning: Contextual k-armed bandit versus Q -Learning[J]. *The Journal of Finance and Data Science*, 2023, 9: 100101.
- [25] HUANG Z, GONG W Y, DUAN J W. TBDQN: a novel two-branch deep Q -network for crude oil and natural gas futures trading[J]. *Applied Energy*, 2023, 347: 121321.
- [26] ZHAO P C, YIN S Y, KUMAR B V D, et al. Improving quantitative stock trading prediction based on MAD using Q -learning technology[C]//*Proceedings of the 2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*. Piscataway: IEEE Press, 2023: 57-60.
- [27] KONDA V, TSITSIKLIS J. Actor-critic algorithms[J]. *Advances in Neural Information Processing Systems*, 1999, 12.
- [28] XU F C, TAN S. Dynamic portfolio management based on pair trading and deep reinforcement learning[C]//*Proceedings of the 2020 3rd International Conference on Computational Intelligence and Intelligent Systems*. New York: ACM, 2020: 50-55.
- [29] GAO N, HE Y Y, JIAO Y M, et al. Online optimal investment portfolio model based on deep reinforcement learning[C]//*Proceedings of the 2021 13th International Conference on Machine Learning and Computing*. New York: ACM, 2021: 14-20.
- [30] LIM Q Y E, CAO Q, QUEK C. Dynamic portfolio rebalancing through reinforcement learning[J]. *Neural Computing and Applications*, 2022, 34(9): 7125-7139.
- [31] KORATAMADDI P, WADHWANI K, GUPTA M, et al. Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation[J]. *Engineering Science and Technology, an International Journal*, 2021, 24(4): 848-859.
- [32] ZHAO T L, MA X, LI X M, et al. Asset correlation based deep reinforcement learning for the portfolio selection[J]. *Expert Systems with Applications*, 2023, 221: 119707.
- [33] WENG L G, SUN X D, XIA M, et al. Portfolio trading system of digital currencies: a deep reinforcement learning with multidimensional attention gating mechanism[J]. *Neurocomputing*, 2020, 402: 171-182.
- [34] MURRAY P, WOOD B, BUEHLER H, et al. Deep hedging: continuous reinforcement learning for hedging of general portfolios across multiple risk aversions[C]//*Proceedings of the Third ACM International Conference on AI in Finance*. New York: ACM, 2022: 361-368.
- [35] WANG Z Y, JIN S R, LI W. Research on portfolio optimization based on deep reinforcement learning[C]//*Proceedings of the 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*. Piscataway: IEEE Press, 2022: 391-395.
- [36] HUANG S H, MIAO Y H, HSIAO Y T. Novel deep reinforcement algorithm with adaptive sampling strategy for continuous portfolio optimization[J]. *IEEE Access*, 2021, 9: 77371-77385.
- [37] JANG J, SEONG N. Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory[J]. *Expert Systems with Applications*, 2023, 218: 119556.
- [38] 冀中, 张文嘉. 集成深度强化学习在股票指数投资组合优化中的应用分析[J]. *计算机科学与探索*, 2024: 1-10.
JI Z, ZHANG W J. Application analysis of integrated deep reinforcement learning in stock index portfolio optimization[J]. *Computer*

- Science and Exploration, 2024: 1-10.
- [39] YUAN X K, JIANG H Q, HUANG L Y. DRL-based quantitative algorithms for gold and Bitcoin portfolio decisions[C]//Proceedings of the 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). Piscataway: IEEE Press, 2022: 878-881.
- [40] LUTHFIANTI B A, SAEPUDIN D, IHSAN A F. Portfolio allocation of stocks in index LQ45 using deep reinforcement learning[C]//Proceedings of the 2022 10th International Conference on Information and Communication Technology (ICoICT). Piscataway: IEEE Press, 2022: 205-210.
- [41] SUN Q Z, SI Y W. Supervised actor-critic reinforcement learning with action feedback for algorithmic trading[J]. Applied Intelligence, 2023, 53(13): 16875-16892.
- [42] ZHU Z Z. Research on quantitative timing trading strategy based on deep reinforcement learning[C]//Proceedings of the 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). Piscataway: IEEE Press, 2022: 633-638.
- [43] CHEN J C, CHEN C X, DUAN L J, et al. DDPG based on multi-scale strokes for financial time series trading strategy[C]//Proceedings of the 2022 8th International Conference on Computer Technology Applications. New York: ACM, 2022: 22-27.
- [44] LIN Y C, CHEN C T, SANG C Y, et al. Multiagent-based deep reinforcement learning for risk-shifting portfolio management[J]. Applied Soft Computing, 2022, 123: 108894.
- [45] MA C, ZHANG J S, LI Z X, et al. Multi-agent deep reinforcement learning algorithm with trend consistency regularization for portfolio management[J]. Neural Computing and Applications, 2023, 35(9): 6589-6601.
- [46] 龙杰, 谢良, 徐海蛟. 集成的深度强化学习投资组合模型[J]. 计算机应用, 2024, 44(1): 300-310.
- LONG J, XIE L, XU H J. Integrated deep reinforcement learning portfolio model[J]. Journal of Computer Applications, 2024, 44(1): 300-310.
- [47] BELYAKOV B, SIZYKH D. Deep reinforcement learning task for portfolio construction[C]//Proceedings of the 2021 International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2021: 1077-1082.
- [48] ZHANG Y F, ZHAO P L, WU Q Y, et al. Cost-sensitive portfolio selection via deep reinforcement learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 236-248.
- [49] ZHANG W P, ZHANG N, YAN J C, et al. Auto uning of price prediction models for high-frequency trading via reinforcement learning[J]. Pattern Recognition, 2022, 125: 108543.
- [50] WU X, CHEN H L, WANG J J, et al. Adaptive stock trading strategies with deep reinforcement learning methods[J]. Information Sciences, 2020, 538: 142-158.
- [51] OLSCHESKI S, DIAO L N, RIESKAMP J. Reinforcement learning about asset variability and correlation in repeated portfolio decisions[J]. Journal of Behavioral and Experimental Finance, 2021, 32: 100559.
- [52] NICOLAS P G, PAUL-ANTOINE B. Deep hierarchical reinforcement learning to manage the trade-off between sustainability and profitability in common pool resources systems[C]//Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2021: 1-7.
- [53] GAŠPEROV B, KOSTANJČAR Z. Market making with signals through deep reinforcement learning[J]. IEEE Access, 2021, 9: 61611-61622.
- [54] SONG Z T, WANG Y N, QIAN P, et al. From deterministic to stochastic: an interpretable stochastic model-free reinforcement learning framework for portfolio optimization[J]. Applied Intelligence, 2023, 53(12): 15188-15203.
- [55] YE Y N, PEI H Z, WANG B X, et al. Reinforcement-learning based portfolio management with augmented asset movement prediction states[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 1112-1119.
- [56] LI X J, CUI C R, CAO D L, et al. Hypergraph-based reinforcement learning for stock portfolio selection[C]//Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 4028-4032.
- [57] 卜湛, 张善凡, 李雪延, 等. 基于深度强化学习的自适应股指预测研究[J]. 管理科学学报, 2023, 26(4): 148-174.
- BU Z, ZHANG S F, LI X Y, et al. Adaptive stock index prediction based on deep reinforcement learning[J]. Journal of Management Sciences in China, 2023, 26(4): 148-174.
- [58] WANG Z C, HUANG B W, TU S K, et al. DeepTrader: a deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(1): 643-650.
- [59] SOLEYMANI F, PAQUET E. Deep graph convolutional reinforcement learning for financial portfolio management-DeepPocket[J]. Expert Systems with Applications, 2021, 182: 115127.
- [60] SHI S, LI J J, LI G H, et al. GPM: a graph convolutional network based reinforcement learning framework for portfolio management[J]. Neurocomputing, 2022, 498: 14-27.
- [61] YANG M, ZHENG X, LIANG Q, et al. A smart trader for portfolio management based on normalizing flows[C]//Proceedings of the IJCAI. [S.l.: s.n.], 2022: 4014-4021.
- [62] NIU H, LI S Y, LI J. MetaTrader: an reinforcement learning approach integrating diverse policies for portfolio optimization[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: ACM, 2022: 1573-1583.
- [63] 柴洪峰, 王帅, 涂晓军, 等. 智能化金融科技创新监管工具: 理念、平台框架和展望[J]. 智能科学与技术学报, 2020, 2(3): 214-226.
- CHAI H F, WANG S, TU X J, et al. Intelligent innovative regulatory tools on financial technology: concept, platform framework, and prospects[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 214-226.

[作者简介]



许波 (1982-), 男, 博士, 广东财经大学信息学院副教授, CCF 高级会员, 计算机应用专委会执行委员, 主要研究方向为机器学习、多智能体等。



贺一峻 (1998-), 男, 广东财经大学信息学院硕士生, CCF 学生会员, 主要研究方向为机器学习、多智能体。



温健城 (2001-), 男, 广东财经大学信息学院硕士生, CCF 学生会员, 主要研究方向为机器学习、多智能体。



李祥霞 (1988-), 女, 博士, 广东财经大学信息学院副教授, 主要研究方向为机器学习、深度学习等。