电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

博士学位论文

DOCTORAL DISSERTATION



论文题目 智能量化交易策略的研究

学科专业	**计算机科学与技术**
学　　号	**202011081620**
作者姓名	**郑泽涛**
指导老师	**邵杰　教授**
学　　院	**计算机科学与工程学院（网络空间安全学院）**

中国知网 https://www.cnki.net

分类号 TP391.41 密级 公开

UDC[注1] 004.8

学 位 论 文

智能量化交易策略的研究

（题名和副题名）

郑泽涛

（作者姓名）

指导老师 **邵杰 教授**

电子科技大学 成都

（姓名、职称、单位名称）

申请学位级别 **博士** 学科专业 **计算机科学与技术**

提交论文日期 **2024.3.13** 论文答辩日期 **2024.5.26**

学位授予单位和日期 **电子科技大学 2024.6**

答辩委员会主席

评阅人

注 1：注明《国际十进分类法 UDC》的类号。

中国知网 https://www.cnki.net

Research on Intelligent Quantitative Trading Strategies

A Doctoral Dissertation Submitted to
University of Electronic Science and Technology of China

Discipline:	**Computer Science and Technology**
Student ID:	**202011081620**
Author:	**Zetao Zheng**
Supervisor:	**Prof. Jie Shao**
School:	**School of Computer Science and Engineering (School of Cyber Security)**

中国知网 https://www.cnki.net

摘　要

投资组合管理（Portfolio Management, PM）是指通过投资策略、投资组合来有效管理大规模资本的过程，其最终目的在于为客户、投资者实现可观的投资回报。资产管理的策略中主要分为主动投资（Active Investment）和被动投资（Passive Investment）两大类别。每类策略都包括两个关键任务：资产选择和资产配置。资产选择任务的目的是挑选合适的资产进行投资，而资产配置任务则着眼于在选定资产间实现最优的资金分配。在主动投资策略中，投资者或基金经理积极寻求超越市场平均水平的回报，因此资产选择和资产配置的优化目标是通过精准的市场分析和资产评估来最大化投资收益。与最大化投资收益目标不同，被动投资策略的目标是实现与整体市场同步的稳定增长。在被动投资策略中，指数化投资尤其受到青睐，其目标是通过部分指数追踪（Partial Index Tracking, PIT）精确复刻市场指数的表现。由于投资目的与主动投资策略的差异，在被动投资中，资产选择和配置的优化目标便转向了如何精确地追踪市场的趋势和表现。

本论文旨在探索如何将先进的人工智能技术，如图神经网络（Graph Neural Network, GNN）和深度强化学习（Deep Reinforcement Learning, DRL），融入到主动投资和被动投资策略中，以优化资产选择和配置这两个核心任务。具体而言，本文的主要研究内容包含以下几点：

（1）**主动投资策略–基于关系-时序图卷积网络的资产选择**：针对主动投资中的资产选择问题，本文设计了一种关系-时序图卷积网络 (Relation Temporal Graph Convolutional Network, RT-GCN) 来解决传统资产（股票）预测方法的两大局限性：（i）将每只资产视为独立个体，忽略了资产之间的相互关系。（ii）将资产预测建模为回归任务（预测股价）或分类任务（预测资产趋势），而不是优化投资的核心目标，即从交易市场中选择预期收益最高的最佳资产。首先，本文将资产的关系及其时序特征建模为一个关系-时序图。接着，应用 RT-GCN 来实现每支资产的关系-时序特征提取。最后，根据提取的特征对每支资产进行得分评估并排序，得分越高的资产代表未来的投资收益越高，投资者也可按需投资得分排名靠前的资产以获得收益。

（2）**主动投资策略–基于多视野融合的资产配置**：针对主动投资中的资产配置问题，本文设计了“多视野融合的交易方法”（Trading Approach Integrated Diverse Horizons）用以解决金融市场非平稳性带来的挑战。金融市场非平稳性的内在原因主要是由于交易者的投资视野和策略多样化导致，而现有的基于机器学习解决方

中国知网 https://www.cnki.net

案未能充分考虑到市场非平稳性的内在原因。因此，本文结合强化学习方法，通过将多种不同的投资视野融合进交易策略来适应市场的变化。首先，通过分离出特定视野的信息并利用这些信息来学习多个具有不同投资视野的子策略。这使得每个子策略能够识别其各自视野内的模式，并做出预决策。然后，学习一个跨视野策略，该策略通过综合第一步中各个子策略所作出的预决策来做出最终的交易决策。通过在决策过程中融合来自不同投资视野的决策意见，该策略提高了其适应市场变化的能力。

（3）**被动投资策略–基于强化学习的资产选择**：针对部分指数追踪中的资产选择任务，本文设计了基于深度强化学习的方案来选择资产，以解决传统启发式方法效率低下的问题，特别是在大规模指数追踪时。本研究将资产选择问题建模化为马尔可夫决策过程（Markov Decision Proces, MDP），并使用深度强化学习（DRL）进行优化。本文还提出关系感知的深度 Q 网络（Relevance-aware Deep Q-Network，RA-DQN）和一个成本感知的奖励函数。RA-DQN 利用领域知识智能地选择指定数量的资产，而成本感知奖励函数评估选择结果，并指导 RA-DQN 选择资产，实现最小化追踪误差目标的同时控制交易成本。

（4）**被动投资策略–基于分层强化学习的指数追踪**：针对部分指数追踪中的资产选择与配置任务，本文设计了分层指数追踪模型（Hierarchical Model for Partial Index Tracking, HIT），通过奖励函数串联起资产选择和配置这两个任务来解决两个任务间的协同优化问题。本文将部分指数追踪构建为一个层次化马尔可夫决策过程（Hierarchical Markov Decision Process, HMDP），并通过层次化强化学习（Hierarchical Deep Reinforcement Learning, HRL）进行优化。模型包含一个高层策略和一个低层策略，前者从成分股中执行资产选择任务，后者在所选资产中执行资产配置任务。此外，模型配有一个成本敏感奖励函数，作为高层策略和低层策略之间的桥梁，实现资产选择和资产配置任务的协同优化。

最后，本论文总结上述研究内容，并在此基础上对未来工作进行展望，提出可行的研究思路与方向。

关键词：资产管理；资产选择；资产配置；深度学习；金融科技

中国知网 https://www.cnki.net

ABSTRACT

Portfolio management (PM) refers to the process of effectively managing large-scale capital through investment strategies and portfolios, with the ultimate goal of achieving significant investment returns for clients and investors. Asset management strategies are mainly divided into two categories: active investment and passive investment. Each strategy includes two key tasks: asset selection and asset allocation. The purpose of the asset selection task is to pick suitable assets for investment, while the asset allocation task focuses on achieving optimal capital distribution among selected assets. In active investment strategies, investors or fund managers actively seek returns that surpass the market average, thus the optimization objectives for asset selection and allocation aim to maximize investment returns through precise market analysis and asset evaluation. Unlike maximizing investment returns, the goal of passive investment strategies is to achieve stable growth in line with the overall market. In passive investment strategies, index investing is particularly favored, aiming to precisely replicate the performance of market indices through partial index tracking (PIT). Due to the differences in investment purposes from active investment strategies, in passive investments, the optimization objectives for asset selection and allocation turn towards how to accurately track market trends and performance.

This dissertation aims to explore how advanced artificial intelligent technologies, such as graph neural network (GNN) and deep reinforcement learning (DRL), can be integrated into active and passive investment strategies to optimize the two core tasks of asset selection and allocation. This dissertation not only explores the challenges and solutions in their practical applications but also analyzes the potential of these technologies in enhancing the efficiency and accuracy of asset management. Specifically, the main research contents of this dissertation include the following points:

(1) **Active investment–asset selection based on relation temporal graph convolutional network**: For the asset selection task in active investment, this dissertation designs a relation temporal graph convolutional network (RT-GCN) to overcome two major limitations of traditional asset selection methods: (i) treating each asset as an independent entity, thereby neglecting the interrelationships between assets, and (ii) modeling asset prediction as either a regression task (predicting asset prices) or a classification task (pre-

中国知网 https://www.cnki.net

dicting asset trends) rather than selecting the assets with the highest expected return from the market. Firstly, we model the relationships among assets and their temporal features into a relation-temporal graph. Then, we apply RT-GCN to extract relation-temporal features for each asset. Finally, based on the extracted features, each asset is scored and ranked, with higher-scoring assets indicating higher future investment returns. Investors can then invest in top-scoring assets as needed to maximize returns.

(2) **Active investment–asset allocation by integrating diverse investment horizon**: For the asset allocation task in active investment, this dissertation introduces a trading approach integrated diverse horizons to relieve the non-stationarity issues. The inherent non-stationarity of financial markets primarily arises from the diversity in traders' investment horizons and strategies, a factor not adequately considered in existing machine learning-based solutions. Therefore, this dissertation employs reinforcement learning, integrating multiple investment insights into the trading strategy to adapt to market changes. Initially, it tailor information specific to certain horizons to learn multiple sub-policies. This allows each sub-policy to recognize patterns within its respective horizon and make insightful pre-decisions. Subsequently, we develop a cross-horizon strategy that makes final trade decisions by integrating pre-decisions from the specific horizon policies learned in the first step. By incorporating pre-decisions from diverse investment horizons into the decision-making process, this strategy enhances its adaptability to market changes.

(3) **Passive investment–asset selection based on reinforcement learning**: For the asset selection task in PIT, this dissertation designs a solution based on deep reinforcement learning to select assets, addressing the inefficiencies of traditional heuristic methods, especially in large-scale index tracking. This study models the asset selection problem as a markov decision process (MDP) and optimizes it using deep reinforcement learning (DRL). The dissertation also proposes a relevance-aware deep Q-network (RA-DQN) and a cost-aware reward function. RA-DQN intelligently selects a specified number of assets using domain knowledge, while the cost-aware reward function evaluates the selections and guides RA-DQN to minimize tracking error and control transaction costs.

(4) **Passive investment–index tracking based on hierarchical reinforcement learning**: For the asset selection and allocation tasks in PIT, this dissertation introduces a hierarchical model for PIT (named HIT) to ensure the collaborative optimization between asset selection and allocation. In this study, PIT is formulated as a hierarchical markov decision process, optimized through hierarchical reinforcement learning (HRL). The model

中国知网 https://www.cnki.net

comprises a high-level policy that learns to select assets from constituents, and a low-level policy that handles the task of asset allocation among the selected assets. The model is equipped with a cost-sensitive reward function, serving as a link for the collaborative optimization of both policies. By learning separate strategies for the two tasks and utilizing the reward function as a bridge for collaborative optimization, the model ensures collaborative optimization of two tasks.

Finally, this dissertation summarizes the research content and provides a future outlook based on feasible research ideas and directions.

Keywords: Portfolio management; asset selection; asset allocation; deep learning; financial technology

中国知网 https://www.cnki.net

目 录

第一章 绪论.....1
1.1 研究背景与意义.....1
1.1.1 研究背景.....1
1.1.2 研究意义.....5
1.2 国内外研究现状.....5
1.3 本文的主要研究内容和创新.....7
1.4 本论文的结构安排.....9
第二章 基础知识.....11
2.1 时序神经网络.....11
2.1.1 循环神经网络.....11
2.1.2 时序卷积网络.....13
2.2 图神经网络.....15
2.3 强化学习.....16
2.3.1 集成强化学习.....17
2.3.2 分层强化学习.....19
2.4 本章小结.....20
第三章 主动投资策略:基于关系-时序图卷积网络的资产排序算法..21
3.1 引言.....21
3.2 相关工作.....23
3.2.1 股票预测.....23
3.2.2 图神经网络.....24
3.2.3 基于图神经网络的股票预测.....24
3.3 预备知识.....25
3.3.1 基于排名的股票预测.....25
3.3.2 关系-时序图的定义.....25
3.3.3 图卷积网络.....26
3.4 模型.....27
3.4.1 关系-时序图卷积网络.....27
3.5 实验结果与分析.....31
3.5.1 数据采集与处理.....31
3.5.2 实验设置.....33
3.5.3 实验结果.....34
3.5.4 模型消融实验.....37
3.5.5 模型参数分析.....39

中国知网 https://www.cnki.net

3.5.6 案例分析.....41
3.6 本章小结.....41
第四章 主动投资策略:基于多视野融合的资产配置.....43
4.1 引言.....43
4.2 相关工作.....44
4.2.1 传统的资产配置方法.....44
4.2.2 基于强化学习的资产配置.....45
4.3 预备知识.....46
4.4 模型:多视野交易策略.....47
4.4.1 视野信息分解.....47
4.4.2 策略学习.....50
4.4.3 网络结构.....51
4.4.4 中心化评论家.....52
4.5 实验结果与分析.....55
4.5.1 实验设置.....55
4.5.2 实验结果.....57
4.5.3 模型消融实验.....59
4.5.4 模型参数分析.....62
4.5.5 案例分析.....64
4.6 本章小结.....65
第五章 被动投资策略:基于强化学习的资产选择.....66
5.1 引言.....66
5.2 相关工作.....67
5.2.1 指数追踪.....67
5.2.2 强化学习在金融领域的应用.....68
5.3 预备知识.....69
5.4 模型: 基于强化学习的指数追踪模型.....70
5.4.1 资产选择.....70
5.4.2 资产配置.....74
5.5 实验结果与分析.....75
5.5.1 实验设置.....75
5.5.2 实验结果.....77
5.5.3 模型消融实验.....80
5.5.4 模型参数分析.....81
5.5.5 案例分析.....81
5.6 本章小结.....82
第六章 被动投资策略:基于分层强化学习的指数追踪.....84
6.1 引言.....84

中国知网 https://www.cnki.net

6.2 相关工作.....86
6.2.1 指数追踪.....86
6.2.2 强化学习.....87
6.3 预备知识.....88
6.4 模型:基于分层强化学习的指数追踪模型.....89
6.4.1 高层马尔可夫决策过程-资产选择.....90
6.4.2 低层马尔可夫决策过程-资产配置.....91
6.4.3 高层策略的优化.....92
6.4.4 低层策略的优化.....96
6.5 实验结果与分析.....96
6.5.1 实验设置.....96
6.5.2 实验结果.....98
6.5.3 模型消融实验.....99
6.5.4 模型参数分析.....102
6.5.5 案例分析.....103
6.5.6 与第五章节模型对比分析.....105
6.6 本章小结.....106
第七章 总结与展望.....107
7.1 总结.....107
7.2 后续工作展望.....109
参考文献.....110

中国知网 https://www.cnki.net

图目录

图 1-1 中国个人可投资资产总规模 1
图 1-2 智能资产管理主要技术框架 2
图 1-3 资产管理类别概览 3
图 1-4 本文的研究框架 7

图 2-1 循环神经网络结构 11
图 2-2 时序卷积网络中的扩张因果卷积结构 13
图 2-3 强化学习的基本框架 17
图 2-4 集成强化学习中两种常见的集成方式 18

图 3-1 公司关系影响股票价格的例子 22
图 3-2 关系-时序图的示意图 26
图 3-3 框架示意图 28
图 3-4 时序卷积示意图 31
图 3-5 不同模型的训练和测试速度对比 37
图 3-6 三种策略在不同指标下的效果图 38
图 3-7 模型在不同超参数条件下的训练效果 40
图 3-8 日收益率预测的示例 42

图 4-1 框架的示意图 48
图 4-2 分解的视野信息进行可视化 49
图 4-3 策略的输入示意图与其网络结构 51
图 4-4 时间消耗对比 59
图 4-5 累积回报可视化 59
图 4-6 分解粒度为 3 时不同策略的累计收益率 60
图 4-7 不同神经网络结构的策略所带来的累计收益率 61
图 4-8 不同设置下的学习曲线 62
图 4-9 不同超参数下的模型性能表现 63
图 4-10 模型性能可视化 64

图 5-1 模型框架概述图 69
图 5-2 RA-DQN 的三个模块 72

中国知网 https://www.cnki.net

图 5-3 映射函数与奖励函数可视化 74
图 5-4 两种不同的数据分割方法 76
图 5-5 在六个基准指数数据集上的测试时间对比 78
图 5-6 不同规模的成份股数据集上的训练效果 79
图 5-7 在六个指数数据集的追踪可视化 80
图 5-8 奖励函数分析与参数 λ 值分析结果 80
图 5-9 在 S&P 500 数据上的指数追踪可视化 82

图 6-1 模型框架概述图 88
图 6-2 映射函数与奖励函数可视化 91
图 6-3 高层策略在资产选择上的示意图 93
图 6-4 MCTS 的网络结构 94
图 6-5 蒙特卡罗搜索树与核回归功能结合示意图 95
图 6-6 平均决策时间对比 99
图 6-7 模型在小、中、大规模成份股的指数上的追踪效果 99
图 6-8 有、无 KR 情况下的回报比较 100
图 6-9 两种不同的高层策略情况下 TEO 性能的比较 101
图 6-10 分层结构效果分析 101
图 6-11 不同 K 值的情况下的追踪效果可视化 102
图 6-12 有、无成本约束情况下的交易决策可视化 103
图 6-13 模型在 S&P 500 数据集上的追踪效果可视化 105

中国知网 https://www.cnki.net

表目录

表 1-1 主动投资策略与被动投资策略的对比 3

表 2-1 分层强化学习方法总结 19

表 3-1 常用符号总结 27
表 3-2 实验数据情况统计表 32
表 3-3 维基关系数据跟产业关系数据情况统计 33
表 3-4 所有模型实验结果对比表 35
表 3-5 与 STHAN-SR 和 RSR 模型对比表 36
表 3-6 在维基关系与产业关系两个关系数据集上的实验结果表现 39
表 3-7 模型在关系卷积和时序卷积两个模块的消融实验结果 39
表 3-8 特征组合的类型 41

表 4-1 常用符号总结 48
表 4-2 数据情况统计 56
表 4-3 所有模型的实验效果对比 57
表 4-4 不同集成模型的效果对比 57
表 4-5 不同数量的子策略下模型的性能表现 60

表 5-1 基准数据集情况统计 75
表 5-2 所有模型在六个基线数据集上的表现 77
表 5-3 训练参数表格 79
表 5-4 S&P 500 指数上不同 K 值情况下的指数追踪效果 82

表 6-1 分层强化学习在金融领域的研究总结 87
表 6-2 基准数据集的介绍 96
表 6-3 所有模型在 6 个基准数据集上的表现 98
表 6-4 在有/无核回归以及有/无成本约束的情况下模型的表现 100
表 6-5 不同模型在 S&P 500 数据集上的表现 105
表 6-6 两个指数追踪模型对比 106

中国知网 https://www.cnki.net

第一章 绪 论

1.1 研究背景与意义

1.1.1 研究背景

1.1.1.1 经济背景

随着中国居民私人财富的快速增长，资产管理（Portfolio Management, PM）正在成为一个备受关注的热点。招商银行的《2023 中国私人财富报告》指出，未来两年，宏观经济温和复苏有望推动中国私人财富市场稳中有进，个人可投总资产年复合增速有望达到 8%-9%，可投总资产有望于 2024 年突破 300 万亿规模。随着居民投资信心回升，资本市场产品[①]将实现 12% 的复合增速，推动总资产增长，如图 1-1[②]所示。这样的财富增长极大地推动了人们对高效资产管理的需求。资产管理的重要性不仅体现在推动中国实体经济发展、改善社会金融结构和增加居民财产性收入方面，更在于其对实现共同富裕、平衡收入分配和再分配方面的积极作用。在这一背景下，资产管理行业正经历着前所未有的发展黄金时期。然而，宏观经济的不确定性以及金融市场波动性成为了当前经济市场的新常态，投资者逐渐将理财重点从单纯的财富增长转向更注重安全和保值。这一转变不仅提升了对多元化投资组合的需求，同时也提高了对专业资产管理知识和技能的需求。

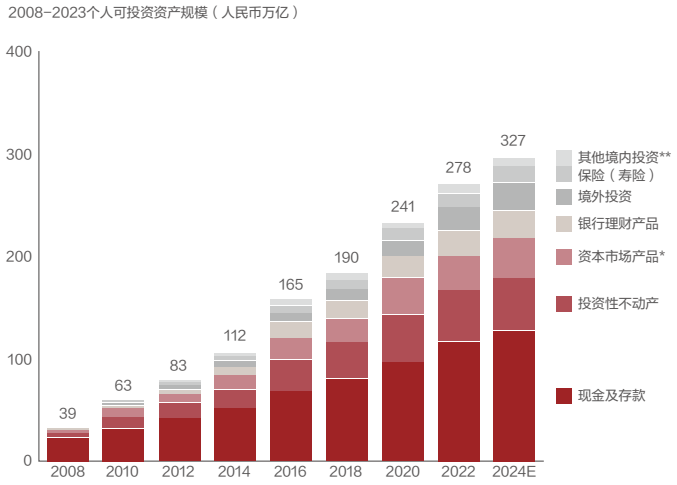


图 1-1 中国个人可投资资产总规模。

① “资本市场产品”包含个人持有的股票、公募基金、新三板和债券

② 来源: 贝恩公司高净值人群收入——财富分布模型

中国知网 https://www.cnki.net

1.1.1.2 技术背景

近年来，随着大数据、互联网、云计算等信息技术的发展，人工智能逐步应用于各个领域，其中就包括金融领域。2022 年 1 月，中国人民银行印发《金融科技发展规划（2022-2025 年）》，指出以加快金融机构数字化转型、强化金融科技审慎监管为主线，将数字元素注入金融服务全流程，其中，智能量化投资是金融科技创新的重要形式之一。智能量化投资通过以云计算和大数据存储技术为支撑，采用机器学习、神经网络算法等先进算法作为核心驱动，为投资者和资产管理人员提供智能化的投资建议与服务，其技术框架如图 1-2所示。此举旨在为传统金融服务的改革与升级注入持续的技术创新动力，推动实体经济服务模式的发展和转型[1]，在考虑用户的风险收益目标等偏好的情况下，利用计算机技术替代传统投资业务中的交易员或投资经理。例如，平安资产管理公司自 2018 年起，一直在向科技赋能的资产管理公司转型，他们认为智能量化投资将在未来成为大众所认可、业界极力推崇的一种主流投资方式[2–5]。此外，我国许多企业，如招商银行、蚂蚁金服、京东金融等，纷纷将大数据和人工智能技术应用于金融投资业务中，为客户提供了更准确的咨询建议及更好的服务[6–9]。

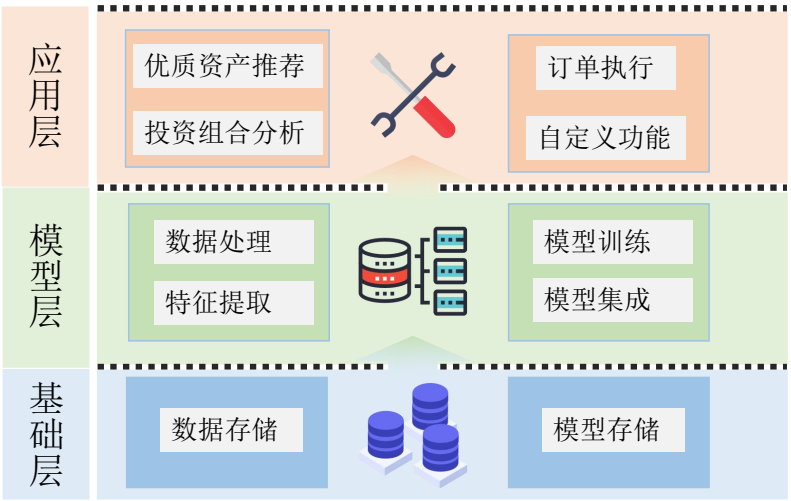


图 1-2 智能资产管理主要技术框架。

1.1.1.3 问题背景

资产管理的策略主要分为主动投资和被动投资两大类别。每类策略都包括两个关键任务：资产选择和资产配置，如图 1-3所示。资产选择的目标是挑选合适的资产进行投资，而资产配置任务则着眼于在选定资产间实现最优的资本分配。主动投资策略的目标是寻求超越市场平均水平的回报，因此在资产选择和配置时，

中国知网 https://www.cnki.net

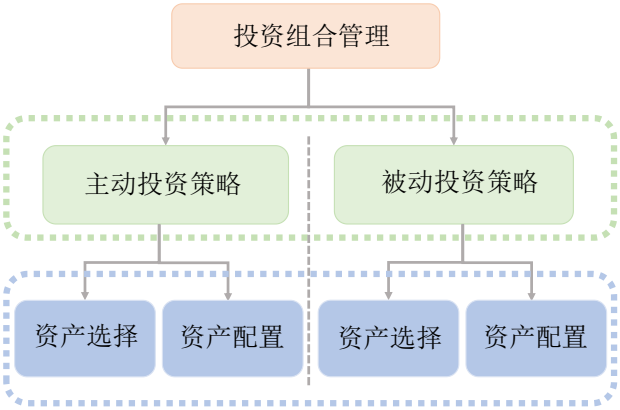


图 1-3 资产管理类别概览。

重点是通过精准的市场分析和资产评估来实现投资收益的最大化。而被动投资策略的目标是实现与整体市场同步的稳定增长。因此，在被动投资中，资产选择和配置的目标便转向了精确地追踪市场趋势和表现。例如，追踪市场指数 S&P 500 的表现（回报率）。这两种投资策略在目标、管理方式、风险与适用对象等特性上都有着明显区别，如表格 1-1所示。

表 1-1 主动投资策略与被动投资策略的对比。

特性	主动投资策略	被动投资策略
投资目标	超越市场平均回报	追踪市场指数表现
管理方式	积极管理，频繁交易	跟踪指数，定期交易
成本	高管理费用	低管理费用
风险与回报	潜在高风险和高回报	较低风险和稳定回报
适用人群	寻求高回报的激进投资者	偏好稳定回报的稳健投资者

随着计算机技术与人工智能算法的发展，资产管理与智能技术相结合的应用案例[10,11]越来越丰富。越来越多的智能算法被应用于主动投资与被动投资以优化资产管理过程。然而现有的基于资产选择与配置的算法在资产管理中的应用仍然存在一定缺陷，主要体现为：

（1）**资产缺关联性**：在主动投资中的资产选择任务上，以往的解决方案只专注于分析资产的时间模式，将每支资产视作独立个体，却忽略了资产之间的相互关系。例如状态频率记忆网络（State Frequency Memory, SFM）[12]、双阶段基于注意力的循环神经网络（Dual-Stage Attention-Based Recurrent Neural Network, DA-RNN）[13]和多尺度双向深度神经网络（Multi-scale Two-way Deep Neural Network, MTDNN）[14]在建模时都忽略了资产之间的关系。虽然也有一些模型考虑到资产间的联系，如关系股票排序（Relation Stock Ranking, RSR）[15]和空间时序超图注意力网络用于股票排序（Spatio Temporal Hypergraph Attention Network for Stock

中国知网 https://www.cnki.net

Ranking, STHAN-SR）[16]，但采用循环神经网络及其变体来提取时序特征，这种方法复杂且效率低下。不仅如此，在投资目标优化方面，这些方法通常将资产预测建模为回归（预测价格）或分类任务（预测趋势，如升或降），而不是针对投资的核心目标进行优化，即从交易市场中选出预期收益最高的最佳资产。

（2）**市场非稳定性**：在主动投资中的资产配置任务上，深度强化学习（Deep-Reinforcement Learning, DRL）因其在序列决策中的突出表现而成为资产配置领域的一个有前景的方法。例如 DeepTrader[17]、EIIE (Ensemble of Identical Independent Evaluations)[18] 和 MetaTrader[19] 等研究结合了深度学习来提取鲁棒特征，以便更好地表示投资组合并用强化学习优化投资回报。然而，因为金融市场的非平稳性，将深度强化学习技术应用于此领域仍面临挑战。金融市场非平稳性的主要原因是具有不同投资视野和策略的多样化交易者的参与。现有基于深度强化学习的解决方案未能充分考虑到这种非平稳性背后的原因。

（3）**组合的复杂性**：在被动投资中，本文主要考虑部分指数追踪这个任务。部分指数追踪的主要复杂性在于资产选择需要满足“基数约束”（Cardinality Constraint），即从较大的资产池中选出有限数量的资产，实现最小追踪误差。基数约束使得资产选择变成一个组合优化问题，需要探索巨大的资产组合空间。除了考虑基数约束外，实际的应用中也会考虑交易成本，过分追求最小化追踪误差而忽略了交易成本也不是一个好的交易策略。现有的方法通常基于启发式算法（Heuristic Algorithms），如遗传算法[20,21]和模拟退火算法[22]，这些方法在处理基数约束和交易成本约束时效率低下，尤其是在大规模指数复制方面，无法在有效时间内找到可行解。

（4）**优化无协同性**：在被动投资的部分指数追踪问题上，资产选择任务受到“基数约束”，而资产配置则受到“非负约束”（Non-negative Constraint）与“和为一约束”(Sum-to-one Constraint)。“非负约束”是指资产配置中分配给被选资产的权重必须是非负的。“和为一约束”要求所有资产的权重之和必须等于 1。这意味着投资者的全部资本都被完全利用，分配到不同的被选资产中。现在的解决方案，例如固定噪声的深度神经网络模型（Deep Neural Network with Fixed Noise, NNF）以及其他在商业实践中广泛使用的前向选择和后向选择[23]方法等，这些模型往往将资产选择和资产配置分开独立优化来解决不同任务中的约束问题。然而这些方法却没有提供一个有效的机制来保证这两个任务之间的协同优化，影响了整体的优化效果。

中国知网 https://www.cnki.net

1.1.2 研究意义

1.1.2.1 理论意义

（1）本文采用图深度学习和强化学习系列技术，基于中国、美国与香港等多个地区的市场数据进行实证分析，针对主动投资与被动投资现有的问题分别提出了解决方案，丰富了现代投资组合理论与金融实证研究，为人工智能技术在经济学和管理学中的深入研究提供了参考。

（2）本文深入探讨了资产管理领域内主动与被动投资策略所面临的一系列问题，并针对这些策略中存在的特定挑战，提出了相应的解决方案。通过这些针对性的方案，本研究不仅为理解投资策略的复杂性提供了新的视角，同时也为现代资产管理的实践与理论发展做出了创新性的贡献，推动了资产管理效率与效果的提升。

1.1.2.2 实践意义

（1）本文提出的智能投资组合管理方法和交易模型，为资产管理公司或基金公司提供了新的思路，使其能够为投资者提供更加多元化和智能化的投资顾问服务。这些工具利用智能算法和金融工具，为投资者定制出合理且高效的投资组合管理方案，能够在市场环境发生变化时，及时对投资组合进行动态优化和再平衡。

（2）本文对于指导个人投资者在面对不确定市场环境时，如何选择优质资产和合理配置资产具有重要的实践意义。它帮助个人投资者提升了理性投资决策能力和风险管理意识，从而在控制投资风险的同时实现更高的投资回报。

1.2 国内外研究现状

智能量化投资代表了一个综合学科，它涉及计算机科学、金融学、数学和管理学等多个学科的研究内容，是一个多学科融合的研究方向。其核心在于运用先进的数学模型和大数据技术，通过自动化算法来指导投资决策，从而达到超越传统投资策略的表现。Jiang 等人[24] 分析了智能技术在多个国家的资产管理应用情况。智能量化投资策略最早在美国起步，可以追溯到 20 世纪 70 年代，起初以在线资产管理公司 Wealthfront、Betterment、Personal Capital 为主，通过 ETF 基金对个人投资者提供自动化服务。其他国家的智能资产管理技术也发展迅速，例如英国公司 Nutmeg 的智能投顾服务提供在线咨询和资产管理等服务，相比传统理财顾问服务费更低。新加坡 Dragon Wealth 公司利用手机移动应用程序提供在线资产管理服务，通过投资动向预测算法结合其他数据分析技术向客户提供投资组合建议。韩国各大券商和银行也陆续推出智能投资服务，通过大数据、计算机算法和量化

中国知网 https://www.cnki.net

分析为客户进行资产管理。我国智能量化技术虽然起步慢，但也呈现百花齐放的景象。例如，招商银行“摩羯智投”，在借鉴国外先进经验的基础上，将人为经验与机器算法相结合，利用计算机智能算法的自我学习功能，构建以公募基金及股票为主要投资目标的智能投顾服务[25]。工商银行的“工银 AI 投”对客户进行画像塑造并将用户的特性考虑进投资策略内，这在一定程度备受追捧和欢迎[26]。

早期的量化策略主要侧重于基于历史数据的统计分析，以识别资产价格的潜在趋势和模式。例如基于技术分析等方法[27,28]来选择预测特征，基于均值回归的方法[29,30]来选择投资组合，基于协整关系[31]来选择资产用于指数追踪等。上述所有方法要么严重依赖于手工制作的特征，例如随机技术指标或移动平均，在实际操作中需要投入大量的精力去选择有效的预测指标，要么因泛化能力差在非平稳市场中无法实现稳定的收益。随着计算机技术的发展和金融市场的复杂性增加，传统的基于统计分析的投资方法已无法满足投资者的需求。自 21 世纪初以来，智能量化策略经历了快速的发展，金融科技的创新使得大数据分析、自然语言处理和深度学习等技术在量化投资中得到广泛应用。例如，Bao 等人[32]利用小波变换[33]和堆叠自编码器提取高级去噪特征，并将这些特征用来预测次日的收盘价。Liu 等人提出多尺度双向神经网络[14]模型自动从基于小波和下采样的信息中学习多尺度模式。Deng 等人[34]和 Li 等人[35]利用外部文本知识或基于文本提取事件来预测股票趋势。这些技术的应用极大地扩展了量化策略的能力，使其不仅能分析结构化数据，还能处理非结构化数据，如新闻报道、社交媒体等内容，为投资决策提供了更全面的信息。

在国内外，随着人工智能技术在金融领域的深入应用，智能量化策略正逐渐从高端市场向中小投资者普及。众多创新型金融科技企业和传统金融机构纷纷布局智能量化领域，研发更加智能化、个性化的投资策略，以满足不同投资者的需求。同时，国内的金融市场和监管环境也在不断优化，为智能量化投资的发展提供了良好的外部条件。然而，无论是在国内还是国际市场，智能量化策略都面临着数据质量参数不齐、模型复杂度高等挑战。特别是在极端市场条件下，模型的鲁棒性和适应性成为了检验智能量化策略成败的关键。在未来，随着技术的不断突破和金融市场结构的演化，智能量化投资策略将进一步向着高效率、透明度和智能化的趋势发展。深度学习和强化学习等尖端技术的应用，将赋予智能量化策略更强的能力，在捕捉市场变化、优化投资决策方面达到更高的准确性。同时，随着区块链和大数据等技术在金融界的广泛应用，智能量化投资将有能力高效地处理与分析大规模数据集，向投资者提供定制化、专业化的服务，推进金融市场朝着更高效率、更公正的目标前进。本研究正是基于这样的发展背景启动，旨在探索

中国知网 https://www.cnki.net

如何将智能技术与量化投资策略更紧密地结合，在这一新兴领域中作出独特贡献。

1.3 本文的主要研究内容和创新

本论文将以资产管理为基础，具体关注资产管理中的主动投资和被动投资两种不同投资策略，并细致讨论资产选择与资产配置这两个任务在两种不同投资策略中的优化方法，如图 1-4所示。具体来说，本论文的主要创新点和贡献有以下四个方面：

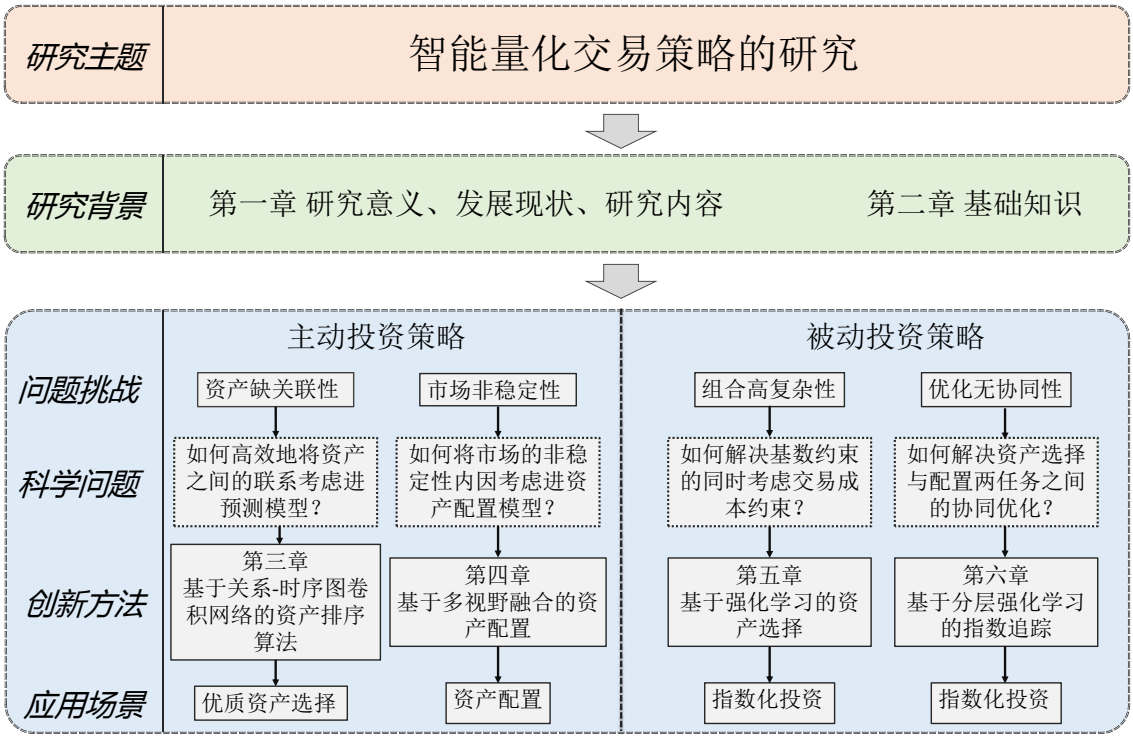


图 1-4 本文的研究框架。

（1）主动投资策略：基于关系-时序图卷积网络的资产选择

在主动投资中的资产选择任务上，以往的解决方案忽略了资产间的关系，只专注于分析资产的时间模式，将每个资产视作独立个体，却忽略了资产之间的相互关系。不仅如此，在投资目标优化方面，这些方法通常将资产预测建模为回归（预测价格）或分类任务（预测趋势，如升或降），因此无法针对投资的核心目标进行优化，即从交易市场中选出预期收益最高的最佳资产。

针对以上问题，本文设计了一种关系-时序图卷积网络（Relation Temporal Graph Convolutional Network, RT-GCN）来充分利用资产之间的关系并实现投资收益的排序。首先，资产的关系及其时序特征会被建模为一个关系-时序图。关系-时序图建成之后，RT-GCN 可以在建好的关系-时序图上进行操作，同步捕获时序维

中国知网 https://www.cnki.net

度上的特征演化和空间维度上的关系依赖性。此外，纯卷积结构使得模型更加高效。最后，这些关系-时序特征被用于得分计算并根据分数高低进行排序，得分越高的资产代表未来投资收益越高。

（2）主动投资策略：基于多视野融合的资产配置

在主动投资中的资产配置任务上，深度强化学（Deep Reinforcement Learning, DRL）因其在序列决策中的突出表现而成为资产配置领域备受关注的方法。然而，因为金融市场的非平稳性，将深度强化学习技术应用于此领域仍面临挑战。金融市场非平稳性的主要原因是具有多样化的投资视野和投资策略的交易者参与。现有基于深度强化学习的解决方案未能充分考虑到这种非平稳性背后的原因。

针对以上问题，本文设计了“多视野融合的交易策略”用以解决资产配置过程中由于金融市场非平稳性所带来的挑战，通过整合多种不同投资视野的交易策略来适应市场的变化。首先，通过为每个策略提供特定视野的信息来学习多个具有不同投资风格的策略。这使得每个策略能够识别其各自视野内的模式，并做出预决策。然后，学习一个跨视野策略，它通过综合第一步中各个特定视野策略所作出的预决策来做出最终的交易决定。通过在决策过程中融合来自不同投资视野的决策意见（预决策），该方法提高了其适应市场变化的能力。

（3）被动投资策略：基于强化学习的资产选择

在被动投资中，本文重点针对部分指数追踪这个任务展开研究。部分指数追踪的主要复杂性在于资产选择需要满足“基数约束”（Cardinality Constraint），即从较大的资产池中选出有限数量的资产，实现最小追踪误差。基数约束使得资产选择变成一个组合优化问题，需要探索巨大的资产组合空间。除了考虑基数约束外，实际的应用中也会考虑交易成本。现有的解决部分指数追踪的方法通常基于启发式算法（Heuristic Algorithms），然而这些方法在处理基数约束和交易成本约束时效率低下，尤其是在大规模指数复制方面，无法在有效时间内找到可行解。

针对此问题，本研究将资产选择问题建模为马尔可夫决策过程（Markov Decision Process, MDP），并使用深度强化学习（DRL）进行优化，将交易成本约束内嵌到奖励函数中，以解决传统方法中优化难的问题。本研究引入关系感知的深度 Q 网络（Relevance-aware Deep Q-Network，RA-DQN）和一个成本感知的奖励函数。RA-DQN 利用领域知识智能地选择指定数量的资产，而成本感知奖励函数评估选择结果，并指导 RA-DQN 选择资产达到最小化追踪误差的同时控制交易成本。该方法能够在考虑交易成本的前提下作出有效决策，从而准确追踪指数。

（4）被动投资策略：基于分层强化学习的指数追踪

在被动投资的部分指数追踪问题上，资产选择任务受到“基数约束”，而资产

中国知网 https://www.cnki.net

配置则受到“非负约束”与“和为一约束”。现有的解决方案往往将资产选择和资产配置分开独立优化来解决不同任务中的约束问题。然而这些方法却没有提供一个有效的机制来保证这两个任务之间的协同优化，影响了整体的优化效率和效果。

针对此问题，本文设计了分层指数追踪模型（Hierarchical Model for Partial Index Tracking, HIT），通过奖励函数来协同优化资产选择和配置这两个任务。在本文中，将部分指数追踪构建为一个分层马尔可夫决策过程，并通过分层强化学习（Hierarchical Reinforcement Learning, HRL）进行优化。模型包含一个高层策略和一个低层策略，前者用来学习从成分股中选择资产，后者则在所选资产中执行资产配置任务。该模型还有一个成本敏感奖励函数，作为高层策略和低层策略之间的桥梁，实现资产选择和资产配置任务的协同优化。

1.4 本论文的结构安排

本论文利用图神经网络的强大图数据处理能力与深度强化学习的决策能力，面向资产管理中的主动投资与被动投资策略，针对这两个策略中的资产选择与资产配置两个核心任务展开了研究，并提出了针对性的方法应对关键挑战。具体章节结构安排如下：

第一章是绪论，总领全文，首先介绍了本文研究工作的背景、意义和研究现状，然后介绍了本研究的内容与贡献。

第二章首先介绍与本文研究相关的基础技术，包括时序神经网络、图神经网络和强化学习。

第三章从主动投资策略的资产选择出发，考虑了资产之间的关系，从资产中选出预期收益高的资产。为此，本文将资产的特征与关系建模成关系-时序图，并设计了一种关系-时序图卷积网络来提取资产的关系-时序特征，从而更高效地考虑资产之间的关系。

第四章从主动投资策略的资产配置出发，通过学习多个具有不同投资视野的策略以及一个跨视野策略，通过跨视野策略来融合多个不同投资视野策略的投资预决策，以解决市场非平稳性问题。

第五章从被动投资策略的资产选择出发，提出一种基于强化学习的关系感知 Q 网络和成本感知奖励函数，通过解决资产选择中的“基数约束”问题，来解决指数追踪任务。

第六章从被动投资策略的资产选择与配置问题出发，学习一个高层策略来和一个低层策略，前者用来选择资产，后者则用来解决资产配置任务。同时，本章还提出一个成本敏感奖励函数来协同两个策略的优化，从而解决两个任务独立优

中国知网 https://www.cnki.net

化的问题。

第七章对本文的主要成果进行简要总结，并对今后可继续拓展和深入的研究方向进行展望。

中国知网 https://www.cnki.net

第二章 基础知识

2.1 时序神经网络

本节介绍两种经典的时序神经网络，它们在时间序列预测、语音识别、自然语言处理等任务中展现出卓越的成效。本文第三、四章应用了时序卷积网络，第五、六章采用了循环神经网络。因此，本节详细探讨了循环神经网络和时序卷积网络的架构与区别，旨在全面理解其机制和实用性。

2.1.1 循环神经网络

循环神经网络（Recurrent Neural Network, RNN）是处理序列数据的一种强有力的神经网络架构，它在文本处理[36]、机器翻译[37]、以及对话系统搭建[38]领域取得了显著的创新成就，逐步发展成为自然语言处理领域的关键技术之一。在本研究的中间章节，特别是第五、六章采纳了 RNN 的一种变体，即门控循环单元（Gated Recurrent Unit, GRU），来编译金融时间序列。考虑到循环神经网络的概念已被广泛接受，本文将不对其底层原理进行深入讨论，仅对 RNN 及其两个主要的改进形式——长短期记忆网络（Long-Short Term Memory, LSTM）和门控循环单元（Gated Recurrent Unit, GRU）进行简要的概述。

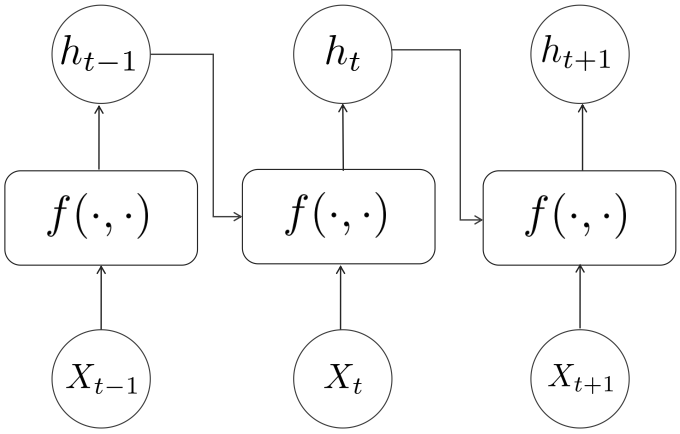


图 2-1 循环神经网络结构。

RNN 定义为包含内部循环链接的网络结构，使得网络能够将之前的输出作为一部分信息再次输入到网络中，如图 2-1所示，其中 $f(\cdot,\cdot)$ 表示单元核，它可以接收上一个时刻 $t-1$ 单元核的输出 h_{t-1} 以及当前时刻 t 的输入 X_t。这种结构在展开后表现为一个周期性重复的模式，并且通过参数共享的模式大大地降低了模型的参数量。RNN 的核心目标是实现从一系列输入到一系列输出的转换，而不必要求

中国知网 https://www.cnki.net

输出序列与输入序列的长度一致，从而适应于不同长度的序列处理。RNN 的数学表达式为：

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t), \tag{2-1}$$

其中，x_t 表示输入，h_t 表示当前时刻的隐藏状态，h_{t-1} 表示上一时刻的隐藏状态，x_t 表示当前时刻的输入，W_{hh} 和 W_{xh} 分别表示隐藏状态和输入的权重，σ 表示激活函数。然而，研究指出标准 RNN 在捕获长距离数据依赖方面存在困难，随着序列的延展，早期的信息可能会逐渐消失，这可能导致梯度消失或梯度爆炸的问题[39]。为应对这一挑战，提出了若干种优化后的 RNN 变种。

LSTM 是对 RNN 的重要改进，它通过引入门控机制来解决 RNN 在处理长期数据依赖问题上的不足，并有效地缓解了梯度消失或爆炸的问题。LSTM 网络由一系列的循环单元构成，其中每个单元包括三种门控机制：输入门、遗忘门、以及输出门，它们协同工作以实现对序列信息的有效追踪和更新。遗忘门在确定保留哪部分历史信息上发挥着至关重要的作用。研究表明，在 LSTM 体系中，遗忘门至关重要，其次是输入门，而输出门的影响则相对较小。LSTM 的数学表达式为：

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i),\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f),\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o),\\
\tilde{c}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c),\\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t,\\
h_t &= o_t \odot \tanh(c_t),
\end{aligned} \tag{2-2}$$

其中，i_t、f_t、o_t 分别表示输入门、遗忘门和输出门的输出，$\tilde{c}_t$ 表示当前时刻的候选单元状态，c_t 表示当前时刻的单元状态，h_t 表示当前时刻的隐藏状态，W_{xi}、W_{hi}、W_{xf}、W_{hf}、W_{xo}、W_{ho}、W_{xc}、W_{hc} 分别表示输入、遗忘、输出和候选单元状态的权重，b_i、b_f、b_o、b_c 分别表示输入门、遗忘门、输出门和候选单元状态的偏置，σ 表示 sigmoid 激活函数，$\odot$ 表示逐元素相乘。

GRU 是 RNN 的另一种广受欢迎的改进版本，它同样采用了门控机制以调控信息的流动性。与 LSTM 相比，GRU 单元仅涵盖两种门控——重置门和更新门，这种设置实质上是将 LSTM 中的输入门和遗忘门简化为单一的更新门。GRU 在众多领域的任务中展示了与 LSTM 相媲美的效果，但由于其结构更加精简且参数更少，计算效率更高，因此在许多研究中被广泛采用。深入研究发现，在大部分非

中国知网 https://www.cnki.net

语言模型任务中，GRU 的性能超越了 LSTM。GRU 的数学表达式为：

$$\begin{aligned} r_t &= \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r), \\ z_t &= \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z), \\ \tilde{h}_t &= \tanh(W_{xh}x_t + r_t \odot (W_{hh}h_{t-1}) + b_h), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \tag{2-3}$$

其中，r_t、z_t 分别表示重置门和更新门的输出，$\tilde{h}_t$ 表示当前时刻的候选隐藏状态，h_t 表示当前时刻的隐藏状态，W_{xr}、W_{hr}、W_{xz}、W_{hz}、W_{xh}、W_{hh} 分别表示输入、重置、更新和候选隐藏状态的权重，b_r、b_z、b_h 分别表示重置门、更新门和候选隐藏状态的偏置，σ 表示 sigmoid 函数，$\odot$ 表示逐元素相乘。

2.1.2 时序卷积网络

相比于传统的序列模型如循环神经网络（RNN）和长短期记忆网络（LSTM），时序卷积网络（Temporal Convolutional Network, TCN）在高效捕捉数据中的长期依赖性方面具有明显优势。TCN 的结构主要包含因果卷积、空洞卷积和残差连接三个模块[40]。

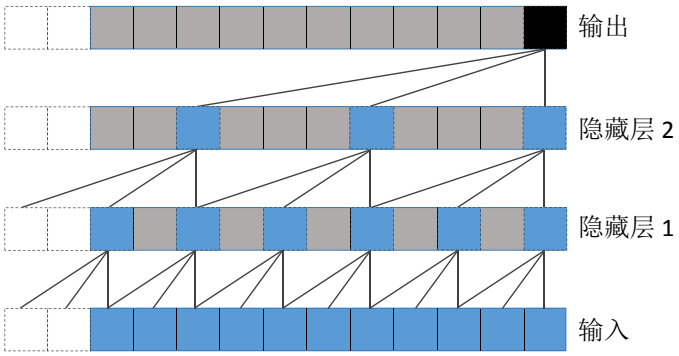


图 2-2 时序卷积网络中的扩张因果卷积结构

1. 因果卷积

因果卷积的设计原理在于确保时间序列分析的真实性和可靠性，使得在任何给定的时间点，模型预测仅依赖于历史和当前观测数据，如图 2-2所示。这种策略排除了对未来信息的依赖，符合时间序列分析的基本原则，即在实时预测场景中，未来的信息不应影响当前的预测结果。这种因果关系的明确界定，为时间序列的顺序分析提供了坚实的基础，确保了模型的预测不会被尚未发生事件的信息所干扰。

2. 空洞卷积

中国知网 https://www.cnki.net

空洞卷积，或称为扩张卷积，通过在传统卷积核内引入固定间隔的空洞，实现了对输入数据更广泛区域的覆盖，而不增加额外的计算成本或参数数量。这一策略使得模型能够有效地整合更长范围内的信息，极大地增强了对长期数据依赖关系的捕捉能力。在处理具有复杂时间依赖特征的序列数据时，空洞卷积的应用显著提高了模型的性能，使其能够识别和利用时间序列中的关键模式，即使这些模式在时间轴上相隔甚远。

3. 残差连接

残差连接技术通过为深层网络结构引入直接的前向连接，有效地解决了梯度消失问题，使得梯度能够在网络中更加顺畅地流动。这种机制不仅提高了深层模型训练的稳定性和效率，还保证了输入信号在经过多层处理后仍能保持其原始信息的完整性。残差连接的应用，特别是在深层时间卷积网络中，允许模型学习到复杂的数据表示，同时维护对输入数据的高度敏感性，从而在各种预测任务中实现更高的准确度。

除了以上三个模块，在 TCN 中，采用的特殊填充策略旨在保持模型输出序列的长度与输入序列相匹配，确保了模型在处理不同长度序列数据时的灵活性和准确性。这种策略通过在序列的适当位置引入零填充，与因果和空洞卷积技术相结合，实现了模型输出的精确控制。此外，这一设计还支持了模型对于序列数据中时间位置信息的精确解读，对于需要保持时间信息一致性的应用场景尤为关键。

2.1.2.1 相对循环神经网络的改进优势

并行处理能力：TCN 通过其独特的结构设计，实现了数据处理的并行化，突破了循环神经网络（RNN）顺序处理数据的限制。这一改进大幅度提升了模型在大规模数据集上的训练速度和效率，使得 TCN 在处理大数据时表现出更高的性能。

内存效率的显著提升：与依赖复杂门控机制的 RNN 相比，TCN 通过简化的网络结构减少了对内存资源的需求，从而提高了模型的内存效率。这一改进使得 TCN 不仅在训练过程中占用更少的计算资源，同时也简化了模型的部署过程，使其更适合在资源受限的环境中应用。

对时间依赖性捕获的优化：TCN 通过结合因果卷积和空洞卷积技术，为模型提供了更大的灵活性和更强的能力，以学习和理解时间序列数据中的复杂依赖关系。这种设计使 TCN 能够识别并利用长期的时间依赖，提升了模型在各种时间序列预测任务上的表现，尤其是在需要处理长距离历史信息的场景中，TCN 的性能优于传统的循环神经网络。

综上所述，通过对因果卷积、空洞卷积、残差连接以及特殊的填充策略等关

中国知网 https://www.cnki.net

键技术的深入分析可以了解到 TCN 在处理时间序列数据方面的显著优势。TCN 不仅在理论上优于传统的循环神经网络，其在实际应用中的表现也证明了其作为时间序列分析工具的高效性和可靠性。随着深度学习技术的不断进步，TCN 及其变体有望在更广泛的领域内发挥重要作用，为解决复杂的序列预测问题提供更为强大和灵活的解决方案。

2.2 图神经网络

在本文的第三章将探讨如何利用图神经网络（Graph Neural Network, GNN）来捕捉不同资产之间的相互影响。因此，本部分将着重介绍图神经网络的核心概念。不同于常规神经网络模型通常处理的是具有固定空间结构的欧几里得数据，图神经网络则旨在处理展现为图结构的非欧几里得数据，例如化合物的分子结构、社交网络的人际连接，以及交通网络中各交叉点之间的时空关系等。作为一种专为图形数据设计的深度学习架构，图神经网络能够深入挖掘实体间的复杂相互作用，因而在众多领域比传统的卷积神经网络和循环神经网络展现出更加出色的性能，吸引了学术和产业界的广泛注意。研究人员经常将非图结构数据转换成图形式，以便更好地在各种任务中应用 GNN。在面对非结构化数据时，图结构往往不是预先定义好的，而是需要根据特定任务的需求和数据内在的联系来确定节点和边的关系。以股票市场预测为例，常见的方法包括使用新闻、企业股东信息等额外数据源 [34,35,41] 来揭示不同公司之间的潜在联系，注意力机制 [42] 的引入也能够自动识别股票间的潜在联系。

图神经网络通过将图结构数据转换为计算图来进行建模，利用边作为连接，实现了图中节点信息的传递与整合。这一过程使得模型能够识别并学习节点间、节点与边之间，以及不同子图间的相互关系。本部分将通过介绍图神经网络中的几种典型模型，详细说明这些模型的构建方法及其工作原理。以一个无向图 $G = \{V, E\}$ 为例，其中 V 代表节点集，E 代表边集合。另设 A 为图 G 的邻接矩阵，表示节点间的连接关系，N 代表图 G 中节点的总数量，而 D 是图 G 的度矩阵，一个描述节点连接数的对角矩阵。

图卷积神经网络主要可分为基于谱域和基于空域两种方法。谱域方法借助图信号处理理论，在谱域内定义图卷积；空域方法则利用图的拓扑结构，直接在图上定义聚合函数以执行卷积操作。对于基于谱域的方法，其核心思路在于首先将图信号通过图傅里叶变换转入谱域，执行卷积运算后，再利用逆图傅立叶变换将信号转换回原始域。谱域中的不同方法对于卷积算子和计算流程各有不同的界定。作为谱域方法的典型代表，图卷积网络（Graph Convolutional Network, GCN）[43]

中国知网 https://www.cnki.net

采用了分层传播规则。

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}),\ H^{(0)} = X, \tag{2-4}$$

在上面的公式中，$\sigma(\cdot)$ 是激活函数，X 代表节点特征的输入矩阵，而 $\tilde{A} = A + I_N$ 则是加入自连接的邻接矩阵，其中 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ 定义了每个节点的度，I_N 表示单位矩阵。$H^{(l)}$ 为第 l 层的的节点表征。$\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 作为对称标准化的拉普拉斯算子，其目的是对邻接矩阵执行标准化处理。从节点的视角看，上述公式可以重新表述为：

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i|}\sqrt{|\mathcal{N}_j|}} h_j^{(l)} W^{(l)}\right), \tag{2-5}$$

这里 $\mathcal{N}_x$ 代表节点 x 的邻域，$h_i^{(l)}$ 为第 l 层中节点 i 的隐藏层向量。

2.3 强化学习

强化学习是机器学习三大类型算法之一（即监督学习、无监督学习与强化学习）。强化学习涉及到智能体在与环境不断互动的过程中，通过反复的尝试和错误来学习。在这一过程中，智能体会不断观察环境的状态，并根据观察结果做出相应的决策。环境会对智能体的每个决策给出反馈（即回报），用以评估在特定状态下采取该行动的效果。强化学习旨在使智能体通过持续与环境的互动，学习到一套最优策略，从而实现对回报的最大化累积。

强化学习建立在马尔可夫决策过程（MDP）的理论基础之上，采用四元组 (S, P, A, R) 来定义状态空间、状态转移概率、动作空间以及奖励函数。其中，S 为状态的有限集合，$s \in S$ 表示集合中的某个状态；A 为动作的有限集合，$a \in A$ 表示集合中的某个动作；P 为状态转移方程，$P(s'|s,a)$ 表示在状态 s 下执行动作 a 后跳转到 $s\prime$ 的概率；R 为奖励函数，当收到智能体在状态 s 下执行动作 a 后，智能体会得到反馈的奖励 $r = R(s,a)$。图 2-3对状态 S、动作 A、状态转移概率 P 以及奖励 R 进行了概括性的描述。强化学习通过智能体与环境的不断交互、试错，旨在学习最优策略 π^* 以获得最大的累积奖励。策略 π 作为一种映射机制，将状态 S 映射至动作 A: $\pi \times S \to A$，指导智能体的动作选择。根据策略的概率分布，策略可分为确定性策略 $a = \pi(s)$ 及随机性策略 $a \sim \pi(a|s)$。在给定策略 π 的情况下，智能体将遵循该策略与环境互动，产生一系列的状态-动作序列 $(s_0, a_0, s_1, a_1, ..., s_T)$。MDP 通过 G_t=$r_{t+1} + r_{t+2} + ... + r_T$ 定义从时间 t 的状态 s_t 开始到终止状态 s_T 所获得的累计奖励。在确定性环境中，累积奖励 G_t 可直接作为评估策略好坏的标准。然而，实

中国知网 https://www.cnki.net

际环境由于状态转移概率 p 的随机性，可能导致多条路径从当前状态达到终止状态，即存在多种不同的回报 G_t。在这种情况下，采用期望回报值作为策略评价标准成为 MDP 的核心。智能体在遵循策略 π 时，在状态 s 获得的期望回报为状态值函数 $V_\pi(s)=\mathbb{E}_\pi(G_t|s_t=s)$，以及在执行动作 a 后遵循策略 π 所得到的期望回报为状态-动作值函数 $Q_\pi(s,a)=\mathbb{E}_\pi(G_t|s_t=s,a_t=a)$。这两种值函数满足以下贝尔曼方程：

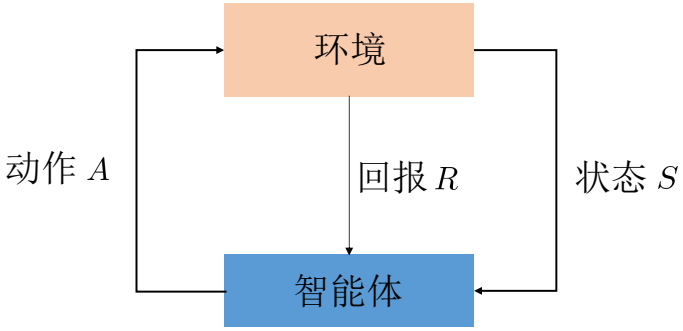


图 2-3 强化学习的基本框架

$$\begin{aligned} V_\pi(s) &= \sum_a \pi(a \mid s)\left[r+\gamma \sum_{s'} p(s' \mid s,a) V_\pi(s')\right], \\ Q_\pi(s,a) &= r+\gamma \sum_{s'} p(s' \mid s,a) V_\pi(s'), \end{aligned} \tag{2-6}$$

其中 s' 代表下一个状态，γ 是折扣率，表示未来奖励的当前价值。在有限 MDP 环境中，状态和动作的空间是有限的，可以准确求得至少一个最优策略 π^*，该策略相比其他所有策略更优或等价。在优化过程中，对于任何一个更优的策略 π'，智能体采用此策略下的所有状态的期望回报应大于或等于采用策略 π 的期望回报，即对所有 $s\in S$，如果 π' 优于 π，则有 $V_{\pi'}(s)\geq V_\pi(s)$。

2.3.1 集成强化学习

集成强化学习（Ensemble Reinforcement Learning，ERL）是一种融合强化学习训练方法于集成学习框架中的新型人工智能技术，它通过采用更为复杂的强化学习（RL）方法替代传统的模型训练手段[44]。ERL 的训练过程涉及模型与环境之间的双向数据交换，其采用的策略与在各种问题中超越其他解决方。根据集成方式的不同，目前集成强化学习的方法被划分为以下几类：基于 Q 函数的集成、基于奖励函数的集成、基于损失函数的集成、基于训练方法组合的集成以及模型（策略）集成等[44]。鉴于本研究的第四章主要基于集成多个策略的强化学习，因此本文特别关注模型组合（Model Combinations）的方法。

中国知网 https://www.cnki.net

在 ERL 中，集成不同类型的模型是一种常见且简单的策略，通过模型组合能够实现结构多样性。这些模型既可以是机器学习模型，也可以是人工神经网络模型[45]。模型组合的结构根据具体问题和解决目标来确定，采用单一类型模型或多种类型模型组合都较为流行。对于仅使用一种类型的模型，可以考虑具有不同深度的网络结构。而其他研究则采用了不同的随机初始化策略[46]或在不同训练阶段的 ANN[47]。

模型选择过程通常涉及从专为解决此类任务而设计的经典或新近提出的方法中进行选择。例如，在预测任务中，大多数研究者会选择卷积神经网络、门控循环单元、人工神经网络等领域内的其他最新技术方法[48,49]。除了确定 ERL 的潜在组成外，确定基础学习器的数量也至关重要。在大多数研究中，通常采用两到三个单一模型。此外，也有一些研究在 ERL 框架内集成了三个以上的模型。Saadallah 等人[50]、Li 等人[49]和 Sharma 等人[51]的研究便是这方面的例证。这些研究广泛证明了，在 ERL 中采用多种策略的组合不仅能超越基线方法，也能胜过最新的集成学习技术。

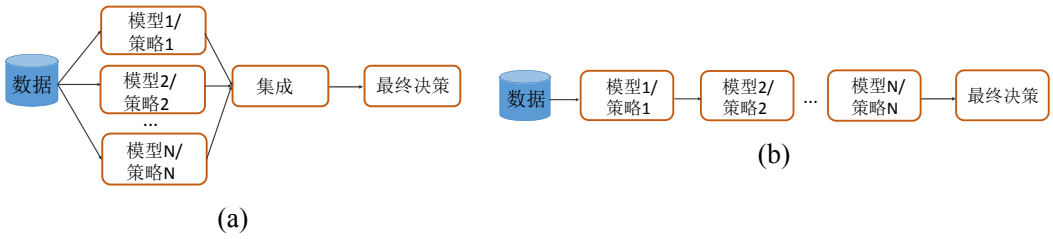


图 2-4 集成强化学习中两种常见的模型/策略集成方式。(a) 为并行式，(b) 为串行式。

集成强化学习（ERL）可以根据 ERL 中模型/策略之间的关系被划分为并行 ERL 和串行 ERL。图 2-4提供了这两种 ERL 方法的示意图。在大多数 ERL 研究中，如 Schubert 等人[52]和 Shen 等人[53]所展示的，模型/策略是在一个并行框架内构建的。这些机器学习模型/策略负责同一任务。每个模型处理后，最终的预测结果将通过特定策略生成。也有一些研究，如 Qin 等人和 Ferreira 等人试图在序列框架内构建 ERL 方法[54,55]。在此框架中，模型/策略按照某种顺序逐步完成最终预测步骤。

模型或策略的集成不仅实现起来相对容易，而且集成之后能够实现 $1+1\geq 2$ 的效果。通过集成多个经典或先进模型，可以在利用每个模型的优势的同时，保持 ERL 方法的多样性[51]。然而，并不是增加 ERL 中模型的数量和种类就一定能够导致性能的提升[56]。单一模型和决策机制的设计对最终输出有着显著的影响。

中国知网 https://www.cnki.net

此外，由于涉及到众多参数，训练这些模型需要大量的数据和时间以及如何衡量每个模型/策略对最终结果的贡献等都需要考虑。因此，在采用模型或策略的集成时，实现所使用模型数量、性能提升与训练成本之间的平衡成为关键考量。

2.3.2 分层强化学习

分层强化学习是将复杂的强化学习问题分解成一些容易解决的子问题（Sub-problem），通过分别解决这些子问题，最终解决原本的强化学习问题[57]。常见的分层强化学习方法可以大致分为四大类，分别为基于选项的强化学习、基于分层抽象机的分层强化学习以及基于 MaxQ 函数分解（Max Q-value Function Decompostion）的分层强化学习。由于第六章主要是基于 MaxQ 函数分解的分层强化学习，因此本节着重介绍 MaxQ 函数分解的分层强化学习。

表 2-1 分层强化学习方法总结。

类型	提出者	基本概念	关键特征
基于选项	Sutton 等人[58]	以三元组 $\langle I, \pi, \beta \rangle$ 表示，其中 I 为初始状态集合，$\pi : S \times A \to [0,1]$ 为策略，$\beta : S \to [0,1]$ 为终止条件	可以将单个动作视为特殊的选项
分层抽象机	Parr 和 Russell[59]	建立在 semi-MDP 基础上，通过有限状态机的状态和当前状态选择策略	将决策过程中的状态与有限状态机状态结合，进行策略选择
MaxQ 值函数分解	Dietterich[60]	将 MDP 分解为多个子任务 $\{M_0, M_1, \ldots, M_n\}$，每个子任务有终止条件和动作集合	学习高层次策略 $\pi = \{\pi_0, \ldots, \pi_n\}$，通过子任务策略解决整个问题

Dietterich 在 2000 年提出了一种层次化强化学习框架，名为 MaxQ 值函数分解[60]，旨在提升层次化强化学习的效率。该方法通过将较大的任务拆分为若干个较小的子任务来实现，即将任务 M 分解成子任务集合 $\{M_0, M_1, M_2, \cdots, M_n\}$。每个子任务都被赋予了特定的学习目标，并且整个学习过程构建了一个以 M_0 为根的任务层次结构。这一结构从下往上逐级解决每个子任务，直至完成根任务。在此过程中，需要形成一套高级策略 $\pi = \{\pi_0, \cdots, \pi_n\}$，其中 π_i 是子任务 M_i 的策略，以有效地指导子任务的解决过程。不同于传统的线性任务模型，层次模型中的子任务是基于上层任务的结果来选择下一步执行哪个子任务的。通过逐一解决任务中的所有子任务，MaxQ 方法能够有效完成整个任务。由于每个子任务的状态-动作空间较小，因此可以更快地得到解决。

中国知网 https://www.cnki.net

2.4 本章小结

本章节旨在为智能量化资产管理策略的研究提供基础性介绍，详细探讨了几种关键的模型结构，包括时序神经网络、图神经网络以及强化学习等。这些模型结构在理解和预测复杂的金融市场动态中发挥着至关重要的作用。通过深入分析这些先进技术，本章为后续深入探讨本文的研究成果和策略实现奠定了坚实的理论基础。本文将展示如何利用这些模型结构来捕捉金融时间序列的特点，挖掘资产之间的关联性，以及如何通过强化学习方法优化投资决策过程，从而在复杂多变的市场环境中实现资产的有效管理和价值增长。下一章将正式开始介绍本文的研究内容。

中国知网 https://www.cnki.net

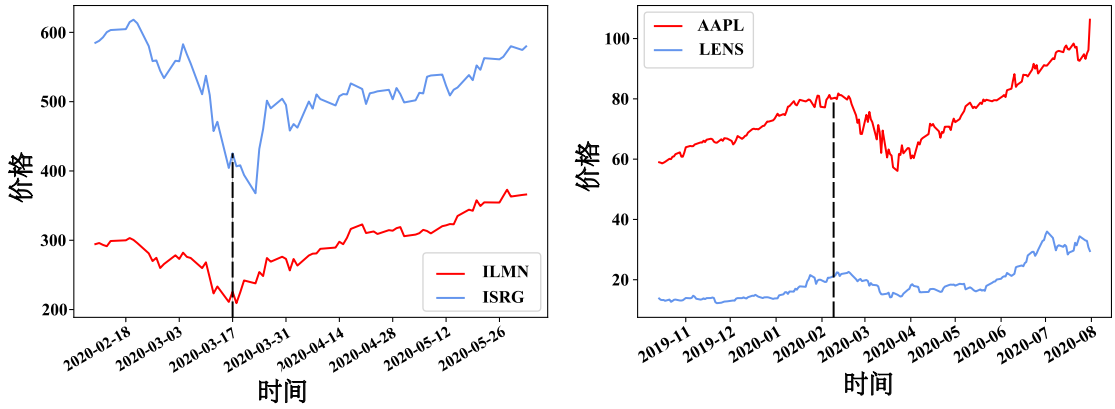
第三章 主动投资策略：基于关系-时序图卷积网络的资产排序算法

3.1 引言

鉴于资产种类繁多，而股票数据易于获取且便于理解，本章将特别聚焦于股票作为研究对象。近年来，股票交易吸引了众多投资者的注意。投资者能否在股市中获利，很大程度上取决于投资者能否在适当的时间买入或交易股票。因此，准确预测股票趋势是做出正确投资决策的关键。在计算机技术的协助下，人们已经投入了大量努力，开发人工智能技术（例如，深度学习方法）来预测股票走势。例如，状态频率记忆（State Frequency Memory, SFM）[12] 是一个优秀的模型，用于捕捉短长期时间频率，并在股票预测上取得了令人印象深刻的表现。多尺度双向深度神经网络（Multi-scale Two-way Deep Neural Network, MTDNN）[14] 通过使用极端梯度提升 [61] 和循环神经网络 [62] 分别自动学习基于小波和降采样信息的多尺度模式。双阶段注意力网络 (Dual-Stage Attention-Based Recurrent Neural Network, DA-RNN) [13] 通过双重注意力机制适当捕捉长期时间依赖性。然而，上述工作只将预测的股票视为个体，忽略了股票之间的联系。众所周知，随着金融市场的发展，公司（股票）通过各种关系广泛相连。目标股票的价格会受到相关公司的影响。例如，Illumina（ILMN）股票与 Intuitive Surgical（ISRG）股票相关，因为它们都属于医药行业。如图 3-1 (a) 所示，它们的股价在 COVID-19 爆发后迅速上涨。另一个例子是苹果（AAPL）股票和蓝思科技（LENS）股票。它们之间的供应商-客户关系对 LENS 的股票价格影响较大，尤其是在发布新版 iPhone 的时期，如图 3-1 (b) 所示。因此，本工作提出在预测股票走势时考虑相关股票之间的关系，而不是将它们视为个体。

在最近的相关研究中，许多文献已经关注到了结合股票间关系的重要性。具体来说，一些研究工作 [34,35,41] 提出了在预测股票趋势时考虑相关股票的额外信息，如事件或新闻，并探讨了这些事件/新闻如何影响股票价格的动向。这些方法通常将股票预测视作一个分类或回归任务，而不是优选预期收益最高的顶尖股票，学术界认为这种预测方式在指导股票选择方面不够理想。相较于上述方法，其他研究 [15,16,63] 则提出了基于排名分数对股票进行排序，并选出得分最高的前 N 支股票作为投资的首选，旨在最大化投资收益。这些工作采取了两步法来提取股票特征，首先识别时序模式，接着分析关系依赖性。例如，有相关文献利用长短期记忆网络（LSTM）捕获时间序列特征，并通过图神经网络分析股票间的相互依赖 [15]；也

中国知网 https://www.cnki.net



(a) (b)

图 3-1 公司关系影响股票价格的例子。(a) 互为同个行业的两只股票走势图，(b) 互为供需关系的两只股票走势图。

有相关工作 [16] 运用超图模型表达复杂的股票关系，并采用 Hawkes 注意力机制 [42] 来捕捉时序特征；文献 [63] 则通过门控循环单元（GRU）提取时序模式，随后使用图注意力网络来实现股票之间的交互。尽管这些模型展示了卓越的性能，但这些工作都采用循环神经网络来捕捉时空特征。由于循环神经网络的“链式结构”，导致以上这些工作的复杂度较高，运行效率较低。

本项研究提出将股票及其日常特征关系建模为关系-时序图，并提出关系-时序图卷积网络（Relation Temporal Graph Convolutional Network, RT-GCN）来提取每只股票的关系-时序特征，结合三种图传播策略，最终根据特征为每只股票打分。得分较高的股票预示着较高的投资回报。在将股票及其日常特征建模为关系-时序图之后，图神经网络可以在其上操作，并同步捕获时序维度上的特征演化和空间维度上的关系依赖性，这是上述其他方法无法达到的 [15,16,63]。此外，纯卷积结构使得模型更高效。在纳斯达克、纽约证券交易所和中国证券交易所三个真实世界股票市场上的实验表明，本研究所提出的模型不仅在投资回报上效果显著，而且在训练速度上也比其他模型有着更优秀的表现。本章的技术贡献有三个方面：

- 为了充分利用每只股票的关系和时序信息，本研究将股票关系和每只股票的时序特征建模为关系-时序图，这是应用 RT-GCN 的基础，也是模型高效的前提。
- 本研究提出了 RT-GCN，针对基于排名的股票预测任务。该模型采用纯卷积结构来同时捕获时序动态和关系依赖性，实现更快的训练速度。
- 在 RT-GCN 中设计了三种关系感知传播策略，以满足股票预测任务的需求。
- 在纳斯达克、纽约证券交易所和中国证券交易所数据集上进行的广泛实验表

中国知网 https://www.cnki.net

明，RT-GCN 可以在有效性和效率方面超越最先进的方法。

3.2 相关工作

本研究涉及的相关工作主要聚焦于三大核心领域。首先是股票预测，这是金融科技领域中长期以来一直备受关注的研究方向，涉及到利用历史数据来预测股票价格或市场趋势的各种方法和技术。接下来是图神经网络，这是近年来快速发展的一种深度学习框架，它通过图结构来处理数据，以此捕捉数据间复杂的关系和模式，广泛应用于包括社交网络分析、推荐系统、以及生物信息学等多个领域。最后，本研究关注的是基于图的股票预测，这是一个新兴的研究领域，它结合了股票预测和图神经网络的优势，通过构建股票之间的关系图来揭示和利用股票间的相互影响，从而提高预测的准确性和效率。这一领域的研究不仅推动了股票预测技术的进步，也为理解和分析复杂的金融市场提供了新的视角和工具。

3.2.1 股票预测

传统的股票预测方法，如技术分析[27,28]，需要投入大量努力来选择有用的指标，然后根据这些选定的指标来预测股票的走势。通常，基本变量如开盘价、收盘价、交易量和最高价等被选作预测股票趋势的指标。然而，这些方法需要经济和金融的背景知识来分析哪些指标适用于股票。此外，这些手工制定的指标经常因其较差的泛化能力而受到批评[64]。与传统模型相比，基于深度学习的模型在发现时序依赖性方面更为强大，且不需要大量的人工特征工程。已经有一些基于深度学习的股票预测方法。Bao 等人[32] 提出从小波变换[33] 以及堆叠的自编码器中获得高级去噪特征，并将这些高级特征输入到 LSTM 以获得次日的收盘价。MTDNN[14] 通过使用极端梯度提升[61] 和 RNN[62] 分别自动学习基于小波和下采样的信息的多尺度模式。Zhang 等人[12] 提出捕捉历史市场数据的多频特征，使用状态频率记忆（SFM）递归网络进行短期和长期预测。Qin 等人[13] 提出双阶段基于注意力的循环神经网络（DA-RNN），通过注意力机制适当捕捉长期时序依赖性。与上述基于深度学习的方法不同，它们通过最小化预测值与真实值之间的差距来训练模型，基于强化学习的方法通过根据与环境交互的反馈调整行为来动态学习，以最大化奖励。Carta 等人[47] 利用集成 Q 学习策略来最小化过拟合问题并最大化回报利润。Liu 等人[65] 提出一个自适应模型，引入模仿学习来利用传统交易策略自动改进交易代理。这些基于深度学习和基于强化学习的模型接受时间序列序列输入，并取得了优异的性能，但它们忽略了股票之间的关系信息，将每只股票视为孤立的个体。

中国知网 https://www.cnki.net

3.2.2 图神经网络

图学习模型的目标是学习每个节点和边的最优表示，以实现特定任务的最佳性能。与基于序列的深度模型相比，基于图结构的神经网络具有更强的表达能力[66]，因为它用一个固定维度的向量表示每个节点，该向量包含了从节点的邻域聚合的信息。由于其性能卓越，图神经网络（Graph Neural Network, GNN）已被应用于各种研究领域。例如，Rhee 等人[67] 利用图卷积和关系网络进行乳腺癌亚型分类。Schlichtkrull 等人[68] 采用图神经网络在知识图谱表示学习中取得了突出的性能，成功处理了现实知识库高度多关系数据的特性。Yan 等人[69] 提出了一个稳健的模型——空间-时序图卷积网络，自动学习空间和时序模式，并在基于骨架的动作识别中取得了令人印象深刻的性能。关于 GNN 与股票预测结合的介绍将在下一小节描述。

3.2.3 基于图神经网络的股票预测

Chen 等人[70] 证明，将股票之间的关系纳入预测模型有助于捕捉不同股票之间流数据的演变，可以使预测更加准确。近年来，已经有大量工作尝试将知识图用于股票预测。通过利用外部知识，Deng 等人[34] 和 Li 等人[35] 分别提出了基于知识驱动的时序卷积网络（Knowledge-driven Temporal Convolutional Network, KDTCN）和事件驱动的 LSTM 模型来预测股票趋势。他们通过结合价格值和从财经新闻中提取的结构化事件来改进股票预测。Li 等人[41] 提出了一种 LSTM 关系图卷积网络，利用股票之间的联系，并同时探索夜间新闻对股票走势的影响。值得注意的是，上述提到的工作[34,35,41] 主要关注外部信息如事件或新闻对股票走势的影响。与[34,41] 不同，本研究的数据集不包含任何诸如新闻或事件的边际信息。Hsu 等人[63] 设计了一个多任务模型，推荐有利可图的股票，同时通过利用全连接的行业图来预测股票走势。Feng 等人[15] 提出关系股票排序（Relation Stock Ranking, RSR）将股票预测视为一个排序任务，并展示了学习排序方法在预测股票中的潜力。RSR 通过应用 LSTM 将股票的时序模式编码成序列嵌入，然后通过考虑股票关系来修正这些嵌入。这些修正后的嵌入输入到一个全连接网络中计算排序得分。Sawhney 等人[16] 提出空间时序超图注意力网络用于股票排序（Spatio Temporal Hypergraph Attention Network for Stock Ranking, STHAN-SR），将复杂的股票关系建模成一个超图，并使用 Hawkes 注意力机制[42] 捕捉时序特征。尽管[15,16,63] 取得了令人惊讶的表现，但这些工作[15,63] 中采用的 LSTM 结构严重限制了框架的训练速度。

与其他研究[15,16,63] 相比，本研究更加注重简洁性和效率。通过将股票之间的

中国知网 https://www.cnki.net

关系及其时间特性建模成一个关系-时序图，本研究使用简单的卷积核直接在该图上提取股票预测所需的关系和时间特征，避免了依赖其他模型进行特征提取的必要。整个模型完全基于卷积结构，从而实现了更快的训练速度，对于需要快速响应的高频股票交易系统来说，这是一个显著的改进。

3.3 预备知识

3.3.1 基于排名的股票预测

传统的股票预测是学习一个预测函数 $y = f(X_t)$，该函数将股票特征 $X_t = [x_{(t-T+1)}, ..., x_t] \in \mathbb{R}^{T\times D}$ 映射到时间步 t 的目标标签空间中，其中 T（窗口大小）代表预测时间步 $t+1$ 之前的时间序列长度，D 是每个时间步的特征维度。与将不同股票视为孤立对象的传统预测不同，基于排名的股票预测旨在学习一个排名函数 $r_{t+1} = R(X_t, \mathcal{A})$，同时考虑股票关系 $\mathcal{A}$，将一组股票的特征 X_t 映射到排名列表中。这里 $X_t \in \mathbb{R}^{T\times N\times D}$，$N$ 是股票数量，而 T 和 D 的定义与传统股票预测相同。$\mathcal{A} \in \mathbb{R}^{N\times N\times K}$，$K$ 表示关系类型，两个股票之间的成对关系被编码为多热二进制向量（Multi-hot Binary Vector）。例如，假设有三种关系，分别是“供应商-客户”、“资金提供者”、“同一行业”，以及两只股票（公司）i，j。考虑到股票（公司）j 作为股票（公司）i 的供应商和资金提供者，可以通过多热向量将它们之间的关系进行编码：$a_{ij} = a_{ji} = [1, 1, 0]$。在基于排名的股票预测机制中，股票在排名列表中的得分越高，在时间步 $t+1$ 中获得的投资回报就越高。

3.3.2 关系-时序图的定义

本节将描述关系-时序图的定义，将关系-时序图表示为 G_{RT}。本章工作利用 G_{RT} 瘵股票及其每日特征进行层次化表示。关系-时序图可以在图 3-2中体现。G_{RT} 基于一系列关系图 G_R 来构建，$G_R \in G_{RT}$，其中关系图中的每个节点对应一个股票。这个图中存在两种类型的边，即关系边和时序边。关系边基于从维基百科和纳斯达克网站获得的先验知识，连接股票，表示为蓝色实线。时序边链接连续时间步中的相同股票，表示为黑线实线。节点和边的数量是固定的，即在训练和测试期间不会动态添加节点或边。为了更好地理解关系-时序图，可以将图 G_{RT} 视为一个圆柱体。每个关系图 G_R 可以看作是圆柱体的一个平面，而这些平面通过时间边连接形成圆柱体。每个平面通过关系边由多个点（股票）连接。

中国知网 https://www.cnki.net

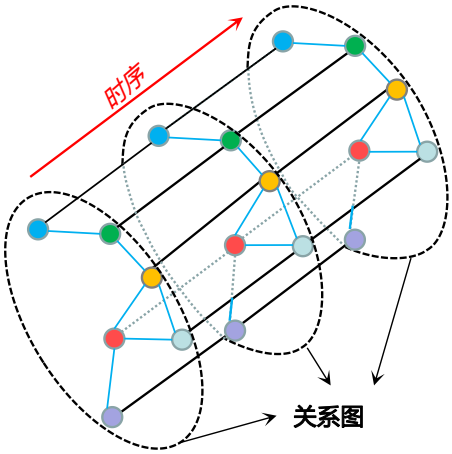


图 3-2 关系-时序图的示意图，RT-GCN 基于此图上运行。这个关系-时序图由三个关系图组成，这些关系图通过时序边（用实心黑线表示）连接。在每个关系图中，股票通过关系边（用实心蓝线表示）。

3.3.3 图卷积网络

由于图的非欧几里得结构，传统的卷积操作无法直接应用于普通图结构上[71]。目前，已经探索出两种实用的方法，成功地将卷积操作应用于图上。一种方法是扩展空间图卷积的空间定义[72]，即空间图卷积，它将节点重新排列成某种网格形式，使得普通的卷积操作可以处理。另一种方法是在谱域上操作，使用图傅立叶变换[73]。与前一种方法相比，后者由于有可靠的理论基础支持，并且在降低计算复杂度从 $O(n^2)$ 到 $O(n)$ 方面取得了有希望的进展[43,74]，因此被更广泛地应用。

典型的图卷积可以被描述为[43]所提出的公式 (3-1)，因为它通过仅考虑一阶切比雪夫多项式，简化了旧版图卷积操作[74]的参数：

$$g * x \approx \theta(I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}})x, \tag{3-1}$$

其中 g 是图卷积操作符。通过堆叠多个卷积层，如公式 (3-1)所示，可以构建一个深度图神经网络。然而，太多的卷积层可能会导致梯度爆炸/消失[43]。为了缓解这个问题，Kipf 和 Welling[43] 引入了以下规范化技巧：$I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \rightarrow \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$，其中 $\tilde{A} = A + I_N$ 且 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$。值得注意的是，上述定义是针对单个特征 x 和单个滤波器 θ 的。这个定义可以推广到 F 个滤波器和每个节点有 C 维特征的图，如公式 (3-2)所描述：

$$Z = f(X, A) = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X\Theta, \tag{3-2}$$

其中 $X \in \mathbb{R}^{N\times C}$，且 $\Theta \in \mathbb{R}^{C\times F}$。公式 (3-2)中的卷积定义是本工作中最终采用的。

中国知网 https://www.cnki.net

3.4 模型

本部分将介绍基于排名的股票预测方法、关系-时序图理论以及图卷积的公式化表达。为了便于理解和参考，表 3-1总结了常用的术语和符号。这些定义和符号是理解后续内容的基础，通过对这些基本概念的阐述，可以帮助读者更好地掌握基于图的股票市场分析方法。

表 3-1 常用符号总结。

符号	符号描述
X_t	股票在 t 时刻的特征
N	股票的数量
T	用于预测股票的时间窗口长度
D	每只股票的特征维度
$\mathcal{A}$	股票之间的关系矩阵
K	股票之间关系的数量
G_{RT}	关系时序图
G_R	关系图
Θ	卷积算子
F	关系卷积算子的数量
H	时序卷积算子的数量

3.4.1 关系-时序图卷积网络

本节将详细阐述了本章提出的模型架构。模型的示意图见于图3-3。在给定一系列股票及其历史顺序数据的情况下，构建了一个以股票为图节点、外部抽取的关系和连续时间连接为图边的 G_{RT}。RT-GCN 模型的输入是每只股票的历史数据。对输入数据将应用多层关系时序图卷积操作，以在 G_{RT} 的每个节点上生成更高层次的特征图。每个关系-时序图卷积层由一个关系卷积结构和一个时序卷积结构组成，输入数据将首先通过关系-时序卷积。关系卷积结构旨在通过聚合其邻居信息丰富每个节点的表示，而后时序卷积结构通过提取连续时间步骤中的关键信息来压缩时间维度。在 RT-GCN 层之后添加了平均池化层，以生成每只股票的表示。这个表示将被输入到一个全连接层以计算排名分数。整个模型以端到端的方式通过反向传播进行训练。接下来将详细讨论每个模块。

1. 关系-时序图构建

本研究利用 G_{RT} 表示一个层次化的股票及其日常特征的表示。更具体地说，本研究构建了一个无向的时空关系图 $G_{RT} = (V, E)$，包含 N 支股票和 T 个时间步，既包括股票内部的连接，也包括时间之间的连接。G_{RT} 的节点集可以描述为

中国知网 https://www.cnki.net

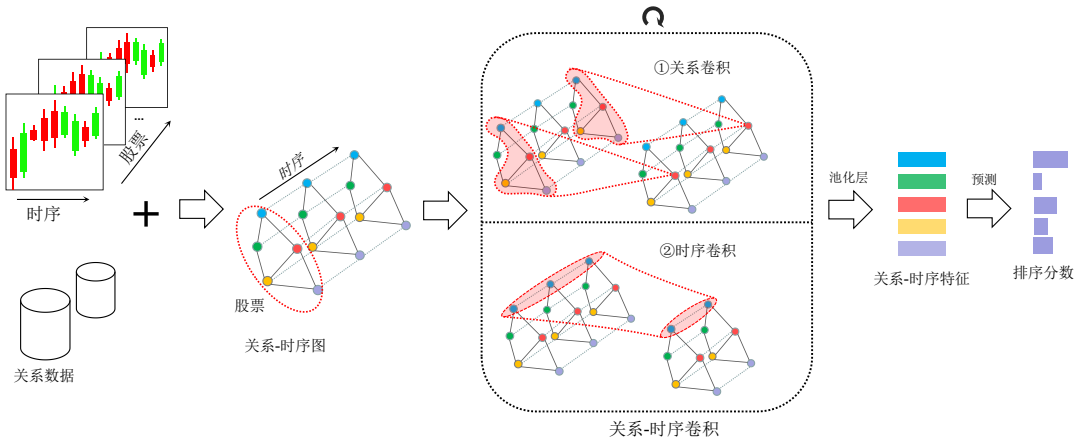


图 3-3 框架示意图。RT-GCN 模块上方的循环标志表示 RT-GCN 网络可以堆叠，以构建更深层次的网络结构。

$V = \{v_{ti} | t = 1, ..., T, i = 1, ..., N\}$。边集 E 由两个子集组成。第一个子集描述了每个时间步内部股票之间的连接，记为 $E_S = \{v_{ti}v_{tj} | (i,j) \in G_R\}$。第二个子集包含了时间之间的边，它们将连续时间步中的相同股票连接起来，记为 $E_T = \{v_{ti}v_{(t+1)i}\}$。通过构建这样一个层次化的表示能够捕捉到股票之间以及股票随时间变化的复杂关系。这种时空关系图不仅能够揭示股票内部的动态特征，也能够捕获股票之间的相互影响，为分析和预测股票市场的行为提供了一个全面而深入的视角。通过对这种图结构的研究可以更好地理解市场的整体结构和个股的行为模式，为投资决策提供科学的数据支持。

G_{RT} 的构建分为两步。首先，根据从外部资源（例如，维基百科和纳斯达克网站上的行业关系①）提取的先验知识，将一个时间步内的股票通过关系边连接起来。维基百科是一个自由且开放的知识库，其中包含了大量关于公司（股票）间关系的事实信息，这些信息以三元组格式存在。本研究根据维基百科中的事实来连接股票。研究 [15,16] 表明，与股票相关性的合作可以显著提升效果，且属于同一行业的股票通常会经历同步的价格运动趋势。遵循前人的工作 [15,16]，本研究同样从纳斯达克网站收集行业信息，并连接属于同一行业的股票。接下来，为了模拟每只股票的时间特征演化，不同时间步的相同股票将被连接起来，正如图3-2中所示的实线黑线。这样，不同时间步的相同股票就可以建立连接，其特征演化可以通过神经网络捕捉。作为 RT-GCN 的输入，节点上的特征向量是灵活的，可以是开盘价、收盘价、最高价或 n 日移动平均等其他有用特征的组合。通过将股票的相关信息以及它们之间的行业关系结合起来，G_{RT} 能够更准确地捕获股票间的相互作用和时间上的特征演化，从而为股票市场的分析和预测提供了一个更加全面

① https://www.nasdaq.com/screening/industries.aspx

中国知网 https://www.cnki.net

和深入的视角。这种方法不仅利用了传统的价格和交易量信息，还融合了来自维基百科等开放知识库的丰富背景知识，以及行业关系等结构化信息，使得模型能够在更宽广的数据基础上进行学习和推断。此外，通过连接不同时间步的相同股票，该模型能够自然地模拟和捕捉股票特征随时间的演化过程，为理解和预测股票的长期行为提供了可能。

2. 关系卷积

本研究在第 3.3.3节中定义了主要的图卷积。本节将详细介绍如何将 GCN 应用于关系-时序图。

如前所述，两只股票之间的关系被编码成一个多热向量。如何有效地将股票对之间的向量融合进 GCN 是 RT-GCN 模型的关键。本研究提出了三种不同的策略来充分利用股票之间的这些关系。这三种策略通过不同的关系感知函数 $\mathcal{R}$ 实现，该函数可以返回不同的加权邻接矩阵 A，以关系矩阵 $\mathcal{A}$ 作为输入。下面将详细说明如何设计这三种策略。

(1) 统一策略：这个策略是最简单的，可以用公式 (3-3) 来表达。它忽略了股票之间不同的关系，而是将所有股票对之间的不同关系视为相同，直接将这些不同的关系在图中传播信息的作用认为是相似的。

$$A_{ij} = \mathcal{R}(\mathcal{A}) = \begin{cases} 1 & if \quad sum(\mathcal{A}_{ij}) > 0 \\ 0 & if \quad sum(\mathcal{A}_{ij}) = 0 \end{cases}. \tag{3-3}$$

$\mathcal{A} \in \mathbb{R}^{N\times N\times K}$ 是在第 3.3.1节中定义的关系矩阵，$\mathcal{A}_{ij} \in \mathbb{R}^{K}$，$A \in \mathbb{R}^{N\times N}$。有了定义的邻接矩阵 A，GCN 模型就可以在每个 G_R 上操作。注意，在卷积过程中，不同时间步的所有 $G_R \in G_{RT}$ 都共享相同的 A。

(2) 加权策略：然而，第一种策略较为粗糙，因为它忽略了不同关系可能导致两只股票之间影响力的差异。为了克服这个缺陷，本研究在图信息传播时应用了一个非均匀系数，如公式 (3-4)所示：

$$A_{ij} = \mathcal{R}(\mathcal{A}) = \mathcal{A}_{ij}^{T} w + b, \tag{3-4}$$

其中，$w \in \mathbb{R}^{K}$ 和 b 是需要学习的参数。与策略 1 相同，所有在不同时间步的 $G_R \in G_{RT}$ 也使用相同的邻接矩阵 A。这种策略更关注关系，因为它能根据不同的关系向量输入返回不同的权重。

(3) 时间敏感策略: 上述两种策略的一个共同局限是，不同时间步的所有关系图共享相同的权重。然而，股票市场的动态性极高，股票的状态和关系的强度可能会在不同的时间步持续演变。对于不同时间步的所有关系图采用固定权重无疑会

中国知网 https://www.cnki.net

限制 RT-GCN 的能力。例如，在之前的图3-1(b) 的例子中，苹果公司（Apple Inc.）股价在新产品发布期间对 Lens 股票的影响比平时更加显著。为了解决这一局限性，本研究定义了考虑时间信息的时间敏感传播过程，如等式(3-5)所示：

$$A(t)_{ij} = \mathcal{R}(X, \mathcal{A}) = \underbrace{\frac{X(t)_i^T X(t)_j}{\sqrt{n}}}_{\text{时间相关性}} \times \underbrace{\mathcal{A}_{ij}^T w + b}_{\text{关系重要性}}. \tag{3-5}$$

在这里，$A = \left[A(1), A(2) \ldots A(T)\right] \in \mathbb{R}^{T \times N \times N}$，$X(t)_i, X(t)_j$ 分别是时间步 t 时股票 i 和股票 j 的特征，而 n 是 $X(t)_i$ 的维度。$w \in \mathbb{R}^K$ 和 b 是可训练的参数。这一策略中的关系感知函数由两部分决定：时间相关性和关系重要性。具体来说，第一部分通过使用当前时间步的股票特征来衡量两只股票之间的相关性。本研究采用了缩放的点积方法[75] 来计算这一项，该方法受文献 [76] 启发，用于计算图中两个节点的权重。第二部分与公式 (3-4)相同，用于返回关系向量的关系权重。与上述两种策略不同，这种策略可以为 G_{RT} 中的每个 G_R 分配一个独特的加权邻接矩阵。它是一种更加高级的策略，因为它同时考虑了两只股票之间的时间相关性和关系重要性，因此比其他两种策略更适合于股票预测。

3. 时序卷积

本研究参考了在其他工作[69,77] 中提出的时序卷积架构，构建了一个时序卷积网络（Temporal Convolutional Network, TCN）模块。TCN 通过使用因果卷积[78] 来捕捉时间依赖性，正如图3-4所示，其中时间点 t 的输出仅与前一层的时间点 t 及之前的元素进行卷积，确保未来信息不会泄露到过去。例如，在第二层中的 x_t 仅与第一层中的 x_{t-2}、x_{t-1} 和 x_t 进行卷积，不依赖于第一层中时间步 $t+1$ 和 $t+2$ 的任何未来信息。通过这种方式，可以确保未来信息不会泄露到过去。通过改变滤波器移动步长并结合零填充[79] 来扩大感受野，可以捕捉输入特征的更长距离依赖性。公式(3-6)给出了 TCN 在第 t 个元素 X_t 上，由滤波器 $\mathcal{F} \in \mathbb{R}^k$ 操作的数学表达式：

$$\mathcal{O}(X_t) = (\mathcal{F} * X) = \sum_{i=0}^{k-1} \mathcal{F}(i) * x_{t-k+i}. \tag{3-6}$$

在 TCN 层之后还增加了残差连接[80]，以简化网络训练过程，并对卷积滤波器应用了权重归一化[81]。此外，为了正则化处理，在每个 TCN 层之后还添加了空间随机失活设置[82]。

4. 池化层与预测层

在 RT-GCN 处理之后，可以获得关系-时序图的关系-时序特征 $Z \in \mathbb{R}^{H \times N \times F}$，其中 H 是经过时序卷积组件处理后的最终时序维度，N 是股票的数量，F 是经过关

中国知网 https://www.cnki.net

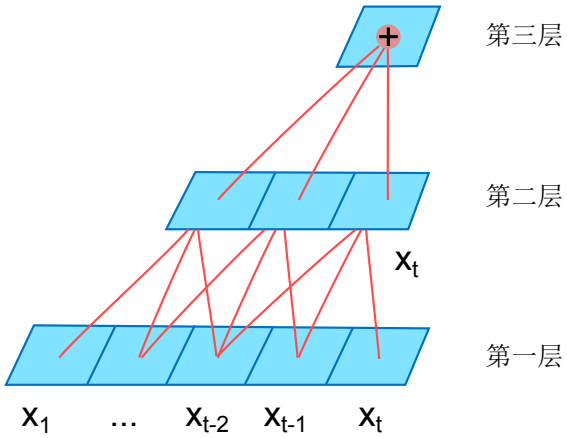


图 3-4 时序卷积示意图。

系卷积组件处理后股票的特征维度。接下来，使用步长为 H 的平均池化层为每只股票生成表示 $Z' \in \mathbb{R}^{N \times F}$。最终，每只股票的表示将被送入一个全连接层来计算排名得分 $\hat{r}_{t+1}$。本研究使用均方误差损失 τ_{reg} 来优化 RT-GCN 模型，以最小化预测与实际回报率之间的差异，并配合成对的排名感知损失 τ_{rank} 来维护排名股票的相对顺序。

$$\tau_{reg} = ||\hat{r}_{t+1} - r_{t+1}||^2, \tag{3-7}$$

$$\tau_{rank} = \sum_{i=0}^{N} \sum_{j=0}^{N} ReLU(-(\hat{r}_i^{t+1} - \hat{r}_j^{t+1})(r_i^{t+1} - r_j^{t+1})), \tag{3-8}$$

其中 r_{t+1} 是真实值，$\hat{r}_{t+1}$ 是预测的排名分数，而 $ReLU$ 是激活函数。因此，RT-GCN 模型的最终损失函数可以表示为公式 (4-6)：

$$\tau = \tau_{reg} + \alpha \tau_{rank} + \lambda ||\beta||^2, \tag{3-9}$$

其中，β 是所有可学习参数的集合，λ 是 $L2$ 正则化的权重，而 α 是回归损失与排名损失之间的平衡参数。

3.5 实验结果与分析

3.5.1 数据采集与处理

本实验在三个主要股票市场进行：纳斯达克证券交易所（NASDAQ）、纽约证券交易所（NYSE）以及中国 A 股市场。对于纳斯达克和纽约证券交易所，采用了 RSR 研究 [15] 中的股票列表。中国 A 股市场的数据来源于沪深 300 指数（CSI）成分股。由于一些股票已退市，因此只选取了 2015 年 2 月 1 日至 2020 年 12 月 31

中国知网 https://www.cnki.net

表 3-2 实验数据情况统计表。

数据集	股票数量	训练时间	训练天数	测试时间	测试天数
NASDAQ	854	2015/01-02 – 2020/02/28	1295	2020/03/02 – 2020/12/31	207
NYSE	1405	2015/01-02 – 2020/02/28	1295	2020/03/02 – 2020/12/31	207
CSI	242	2015/01-02 – 2020/04/30	1295	2020/05-01 – 2020/12/31	139

日期间有交易记录的股票，最终获得纳斯达克、纽约证券交易所和沪深 300 指数的 854 只、1405 只和 242 只股票，详细数据见表 3-2。数据收集包括两部分：（1）历史收盘价数据，与先前研究 [14,15] 保持一致，本研究只使用收盘价及其 n 日移动平均作为特征；（2）关系数据，涵盖行业关系（例如生物技术、计算机软件、护理服务）和公司间的维基关系，如供应商-消费者关系和资助者关系。

1. 股票历史价格数据

股票的历史收盘价数据通过公开的 API——雅虎财经①进行收集。随后，对这些收集的价格序列数据进行了以下四个步骤的进一步处理。首先进行归一化收盘价操作，遵循先前研究 [83,84] 的方法，通过将每期的收盘价除以输入时间序列最后一期的收盘价来实现，防止未来信息泄露，即通过 $p^t = \frac{p^t}{p^T}$ 进行归一化处理，其中 p^T 为输入时间序列最后一期的收盘价。接着，除了收盘价外，还计算了收盘价的 5、10、20 日移动平均值，以表示每只股票的短期和半月趋势。之后，为了根据前 T 个交易日的特征预测 $T+1$ 个交易日股票的排名列表，如第 3.3.1节中基于排名的股票预测定义所示，需计算股票的回报率，这一指标用于评估股票收益。遵照 [15] 的方法，设置股票的回报率为真实值，因为它反映了股票的预期收益，较高的回报率意味着与前一交易日相比有更高的收益。第 i 只股票在交易日 t 的回报率依据等式 (3-10)计算：

$$r_i^{t+1} = (p_i^{t+1} - p_i^t)/p_i^t, \tag{3-10}$$

其中 p_i^t 为第 t 天的收盘价。第四步：分离序列数据。对于所有数据集，按时间顺序将序列化的价格数据划分为训练和测试两个时间段。数据统计结果如表 3-2所示。

2. 股票关系数据

如图 3-1(a) 所示，同一行业下的股票往往表现出相似的趋势。例如，自新冠疫情爆发以来，医疗行业的股票迅速上涨。为了捕捉这种隐性模式，本研究根据 [15] 提供的数据集，筛选出可用股票之间的行业关系，并将其作为 G_{RT} 的关系边。如果两只股票属于同一行业，就将这个行业视为这两只股票之间的一种关系。例如，

① https://github.com/ranaroussi/yfinance

中国知网 https://www.cnki.net

Facebook Inc. 和 Twitter Inc. 都属于“技术服务-互联网软件/服务”行业，这可以表示为（Facebook；技术服务：互联网软件，服务；Twitter）。

本研究通过设置另一种关系信息——基于维基百科的公司关系，来丰富 G_{RT} 的关系边。维基百科 [85] 是最大且最活跃的开放领域知识库之一，拥有丰富的实体关系资源。它包含了大量的公司实体和公司间关系，这些关系可能反映了股票之间的影响。维基关系可以通过公共 API wikimedia[①]查询获得。两支股票之间可能存在多种关系，例如，谷歌公司（GOOGLE Inc.）和字母表公司（ALPHABET Inc.）就有两种关系：“ALPHABET; Technology Services:Internet Software; GOOGLE”和“ALPHABET; own by GOOGLE”。两支股票之间的多重关系将会被编码为一个多热向量。表 3-3总结了关系类型的数量和股票对之间的关系比例。与其他两个数据集相比，由于大多数中国公司没有在维基百科上记录，且这些中国公司之间的关系不完整，CSI 数据集没有维基关系。需要注意的是，为了防止未来信息的泄露，这些关系是在测试期之前收集的。

表 3-3 维基关系数据跟产业关系数据情况统计。

数据集	维基关系		产业关系	
	数量	比例	数量	比例
NASDAQ	41	0.3%	97	5.4%
NYSE	28	0.4%	108	6.9%
CSI	-	-	24	6.7%

3.5.2 实验设置

1. 评测程序

参照文献 [15, 86] 的做法，本研究采用了日买卖交易策略来评估模型在股票收益排名方面的表现。本实验遵循了以往的研究 [15,16] 中提出的所有交易假设，即在交易日 t 购买前 N 名股票，并在 $t+1$ 日卖出这些股票，以确保公平性。本研究的工作重点是准确地选择回报率最高的股票，而不考虑购买股票的数量或投资的资本金额。

2. 对比模型

对比模型可以分为基于分类（CLF）、基于回归（REG）、基于强化学习（RL）以及基于排名（RAN）的方法。为了确保实验结果的显著性，所有方法都使用相同的特征来预测具有相同窗口大小 T 的日级别回报率。另外本研究还通过将关系

① https://www.mediawiki.org/wiki/Wikibase/DataModel/JSON

中国知网 https://www.cnki.net

图卷积替换为图注意力网络 [76] 实现另一个变体 RT-GAT，并通过连接至少具有一种类型关系的节点对（股票）来为 RT-GAT 构建图。在这些对比模型中，ARIMA [28]、A-LSTM [87]、SFM [12]、LSTM [32]、DQN [47]、iRDPG [65] 以及 Rank_LSTM [32] 无法利用关系数据。而作为基于图的方法，RSR_I [15]、RSR_E [15] 和 RT-GAT [76] 能够像本研究的模型一样利用关系数据。

3. 评测指标

股票排名预测的主要目标是对股票的相对顺序进行精准排名。因此，本研究采用了平均倒数排名（Mean Reciprocal Rank, MRR）和累计投资回报率（Investment Return Ratio, IRR）作为评估模型性能的两项关键指标。MRR 是一个在排名任务中广泛采用的度量标准，用于计算测试日期中排名首位的股票的 MRR 结果。IRR 则作为主要评估指标，直接映射了股票投资的回报效果，通过汇总每个测试日期中排名列表前 N 位股票的回报率得出。分别计算排名列表中前 1 位、前 5 位和前 10 位股票的 IRR 结果，标记为 IRR-1、IRR-5 和 IRR-10。MRR 和 IRR 指标的增大表明模型性能的提升。

4. 训练设置

所有实验均在配备 32GB RAM 和 2 个 NVIDIA TITAN GPU 的计算机上完成。为确保结果的统计显著性并降低随机变异影响，各模型均执行 15 次，取其平均性能值进行分析。遵循各基线模型在相关文献中报道的调参策略，以达到最佳性能。RT-GCN 模型采用 PyTorch 框架实现，并选用学习率为 0.001 的 Adam 优化器进行训练。基于图方法的超参数优化采用统一的策略和网格搜索方法，以选定最佳的超参数组合。具体而言，基于图模型的两个关键超参数——窗口大小 T 和公式 (4-6)中的 α 值，分别从集合 {5, 10, 15, 20} 和 {0.01, 0.1, 0.2} 中确定。公式 (4-6)中的 λ 值设置为 0.01。实验发现，使用 RT-GCN 模型的单层结构最为合适，过多层次可能引发过拟合问题。

3.5.3 实验结果

1. 对比模型实验结果对比

性能比较的结果在表 3-4展示，表现最优的模型以粗体标出，而次优模型则用下划线高亮。分类方法只提供三种结果（上涨、中立、下跌）而不能对股票进行排名，因此随机选择前 N 支股票计算 IRR。该模型的三种策略——统一（RT-GCN (U)）、加权（RT-GCN (W)）和时间敏感（RT-GCN (T)）——均展示出超过所有对比模型的回报率。通常，基于强化学习和排名的方法相较于分类和回归方法能够获得更高的投资利润，这说明了将股票预测问题视作排序问题并选取最有盈利

中国知网 https://www.cnki.net

表 3-4 所有对比模型的表现。‘-’ 表示排序指标无法计算出来。

	模型	NASDAQ				NYSE				CSI			
		MRR	IRR-1	IRR-5	IRR-10	MRR	IRR-1	IRR-5	IRR-10	MRR	IRR-1	IRR-5	IRR-10
CLF	ARIMA [28]	-	0.12	0.20	0.23	-	0.10	0.17	0.28	-	-0.13	-0.09	0.07
	A-LSTM [87]	-	0.23	0.40	0.59	-	0.43	0.51	0.69	-	-0.07	0.15	0.11
REG	SFM [12]	0.032	0.11	0.28	0.36	0.029	0.13	0.23	0.31	0.007	-0.21	0.10	0.17
	LSTM [32]	0.027	0.13	0.22	0.28	0.013	0.09	0.15	0.19	0.011	-0.33	0.07	0.12
RL	DQN [47]	0.041	0.20	0.34	0.58	0.033	0.12	0.33	0.60	0.009	0.11	0.13	0.15
	iRDPG [65]	0.049	0.29	0.37	0.46	0.040	0.20	0.28	0.57	0.010	0.12	0.14	0.15
RAN	Rank_LSTM [32]	0.041	0.35	0.42	0.82	0.029	0.30	0.64	0.63	0.016	-0.24	0.14	0.19
	RSR_I [15]	0.043	0.81	0.93	0.88	0.032	0.66	0.98	1.01	0.015	-0.04	0.29	0.33
	RSR_E [15]	0.055	0.89	0.83	0.84	0.048	0.88	0.92	0.95	0.021	0.29	0.26	0.28
	RT-GAT [76]	0.049	0.11	0.67	0.87	0.044	0.48	0.87	0.97	0.027	0.27	0.30	0.25
Ours	RT-GCN (U)	0.055	0.45	0.71	0.82	0.040	0.65	0.88	0.96	0.017	0.13	0.15	0.19
	RT-GCN (W)	0.059	0.63	0.86	0.98	0.047	0.80	0.95	0.99	0.022	0.29	0.26	0.35
	RT-GCN (T)	**0.061**	**1.25**	**0.97**	**1.03**	**0.056**	**0.92**	**1.10**	**1.13**	**0.031**	**0.35**	**0.35**	**0.38**
	改进	10.9%	40.4%	4.3%	17.0%	16.7%	4.5%	12.2%	11.9%	14.8%	20.7%	16.7%	15.2%
	p 值	3.05e-4	9.16e-5	6.10e-5	3.05e-4	1.52e-4	4.27e-4	7.63e-4	3.05e-5	3.05e-5	1.53e-4	6.10e-5	4.82e-4

潜力的股票进行交易的方法是正确的。虽然 RSR_E 和 RSR_I 在大部分指标上优于 RT-GCN (U) 和 RT-GCN (W)，但这并不代表该模型效用不足。与 RT-GCN (U) 和 RT-GCN (W) 相比，RSR 模型更为精细，因为它们包括了时间序列的关系动态。与 RT-GCN (T) 比较则更为公平，后者也包含了时间维度的关系变化。表 3-4显示，RT-GCN (T) 的性能超过了 RSR_E 和 RSR_I，得益于采用缩放点积 [75] 来更高效地计算股票间的相关性。RT-GAT 虽与多数基准具有竞争力，但未能全面超越 RT-GCN (U) 和 RT-GCN (W)，原因在于 GAT 仅通过时间特性来计算股票间权重，未能充分考虑多重关系，限制了其性能表现。此外，观察到各策略相比于前一种都有所改进，特别是 RT-GCN (W) 在四项指标上优于 RT-GCN (U)，因为前者考虑了图中信息传播中不同关系的影响力，即对不同关系赋予了不同权重。而 RT-GCN (T) 作为三种策略中的佼佼者，因其考虑了不同状态下的权重演变。

此外，本研究采用配对 Wilcoxon 符号秩检验 [88] 来验证方法超越其他模型的统计性意义，并在表 3-4中报告了显著性水平的 p 值。统计性意义分析基于 RT-GCN (T) 与最强对比模型之间的 15 对实验结果进行。所有结果均由模型在相同参数设置下运行 15 次获得。许多研究 [89,90] 以及本研究采纳了 0.05 的置信水平作为“经验法则”。换言之，当 p 值小于 0.05 时，认为模型性能超越对比模型具有统计学意义。从表 3-4可见，所有 p 值均小于 0.05，因此可以得出实验结果具有统计学显著性的结论。

为了深入验证模型的性能优势，本研究采用了与文献 [15] 相同的数据集进行评估，该数据集包括仅含行业关系的关系数据和序列价格数据（分别标记为 NASDAQ-II 和 NYSE-II 以便区分），并采用相同的参数配置，即窗口大小 T 和学习率。表 3-5中展示的结果直接引用自 [15,16]。最优结果以粗体显示，而次优模型则

中国知网 https://www.cnki.net

表 3-5 与 STHAN-SR 和 RSR 模型在 NASDAQ 和 NYSE 数据集上的表现对比。‘-’ 表示因无开源代码或数据原因无法复现。

模型	NASDAQ-II			NYSE-II		
	MRR	IRR-5	IRR-10	MRR	IRR-5	IRR-10
RSR_I [15]	0.032	0.13	0.22	0.045	0.10	0.12
RSR_E [15]	0.032	-	-	0.043	-	-
STHAN-SR [16]	-	0.44	-	-	0.33	-
RT-GCN (T)	**0.040**	**0.48**	**0.50**	**0.053**	**0.37**	**0.48**
p 值	0.002	0.008	0.001	0.003	0.001	0.0

通过下划线高亮。本研究对公开数据集进行了 15 次运行，得到 15 个结果，并通过单样本 Wilcoxon 符号秩检验[88] 评估结果的统计性意义，该检验可用于判断模型性能相对于对比模型的统计学优势。表 3-5中的结果表明，从统计学角度看，本研究的模型性能显著优于 RSRs 和 STHAN-SR（所有 p 值均小于 0.05），这进一步证明了基于关系-时序图建模的方法相较于 RSRs[15] 和 STHAN-SR[16] 采用的两步处理格式更具前瞻性。

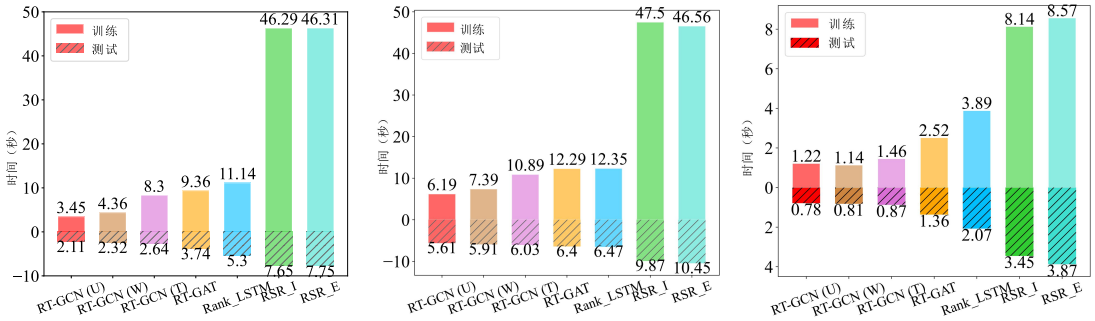
2. 效率对比实验

除了在上述评价指标上展现出的卓越性能，本模型在效率方面也表现优异。如图 3-5所示，以下是观察到的情况（为保证比较公平，仅与需对股票进行排名且考虑股票间关系的基于排序的方法进行比较）：无论在哪个数据集上，本模型的训练和测试时间均显著低于其他对比方法。例如，在纳斯达克数据集上，本模型的训练速度分别是 Rank_LSTM 和 RSR 的 3.2 倍和 13.4 倍。在测试速度方面，本模型比 Rank_LSTM 和 RSR 分别快 2.5 倍和 3.6 倍。RT-GAT 在训练和测试速度上都与 RT-GCN（T）接近，同时超过了 Rank_LSTM 和 RSR，进一步证明了将股票的时间特征和关系融入到一个关系-时序图中的效率。在其他两个数据集上，本模型的训练和测试速度也有显著降低。由于篇幅限制，不对其进行详细分析。本模型之所以比 Rank_LSTM 更高效，是因为通过一个简单的卷积操作即可实现，无需借助其他模块来提取特征。相较于 RSR，本模型的效率更高，原因在于 RSR 模型需要等待 LSTM 完成时序特征的提取后才能进行关系依赖的计算，这一过程大幅降低了其效率。

3. 回报率对比实验

图 3-6 展示了测试期间回报率变化的过程。可以看出，IRR-1 相比于 IRR-5 和 IRR-10 出现了严重的波动（例如，在图 3-6(b) 中，IRR-1 在第 30 个交易日迅速上升，并在第 160 个交易日急剧下降）。IRR-1 的结果表明，从数百只股票中仅选择一只股票是一个高风险的操作。一旦排名顺序准确，买卖股票将实现更高的回报率，反之亦然。与 IRR-1 相比，IRR-5 和 IRR-10 上升得更平稳。原因是 IRR-5 和

中国知网 https://www.cnki.net



(a) (b) (c)

图 3-5 不同模型的训练和测试速度对比。阴影部分为测试速度。(a) 为 NASDAQ 数据集结果，(b) 为 NYSE 数据集结果，(c) 为 CSI 数据集结果。

IRR-10 考虑了多只股票，并且投资得到了良好的多样化以降低风险。

为展示模型的有效性，该研究将其与道琼斯指数（Dow Jones Index, DJI）、标准普尔 500 指数（Standard and Poor's 500, S&P 500）与沪深 300 指数（China Securities Index 300, CSI 300）进行比较，这些均是投资者比较关注的市场指数。通过从公开金融网站收集的 S&P 500①、DJI② 和 CSI 300③ 在测试期间的数据值，结果表明该模型在三种策略下获得的回报率均高于这些主流市场指数，如图 3-6所展示。一般情况下，模型回报率超过市场指数表明其实用性和竞争力[15]，验证了模型的有效性。

3.5.4 模型消融实验

本节主要研究两个问题：（1）不同的关系图是否会影响模型的性能，这对于未来构建更有效的关系图具有积极意义；（2）关系卷积模块和时序卷积模块对 RT-GCN 的影响。下文中将会详细介绍这两项消融研究。

1. 关系数据对模型的影响

在纳斯达克（NASDAQ）和纽约证券交易所（NYSE）数据集上，维基关系与行业关系的性能比较见表 3-6。从表中可以有以下几点发现：（1）提出的 RT-GCN 模型采用三种策略，在两种关系数据上的表现均优于 Rank_LSTM 模型。这一结果验证了与仅考虑时序特征的模型相比，建模关系数据的有效性。此外，（2）在两个数据集上，采用行业关系的 RT-GCN 比采用维基关系的 RT-GCN 表现更好。这可能是因为行业关系的关系比例（例如，在纳斯达克为 5.4%，在纽约证券交易所为 6.9%）大于维基关系的关系比例（例如，在纳斯达克为 0.3%，在纽约证券交易

① https://www.macrotrends.net/2526/sp-500-historical-annual-returns
② https://dqydj.com/dow-jones-return-calculator
③ https://hk.investing.com/indices/csi300-historical-data

中国知网 https://www.cnki.net

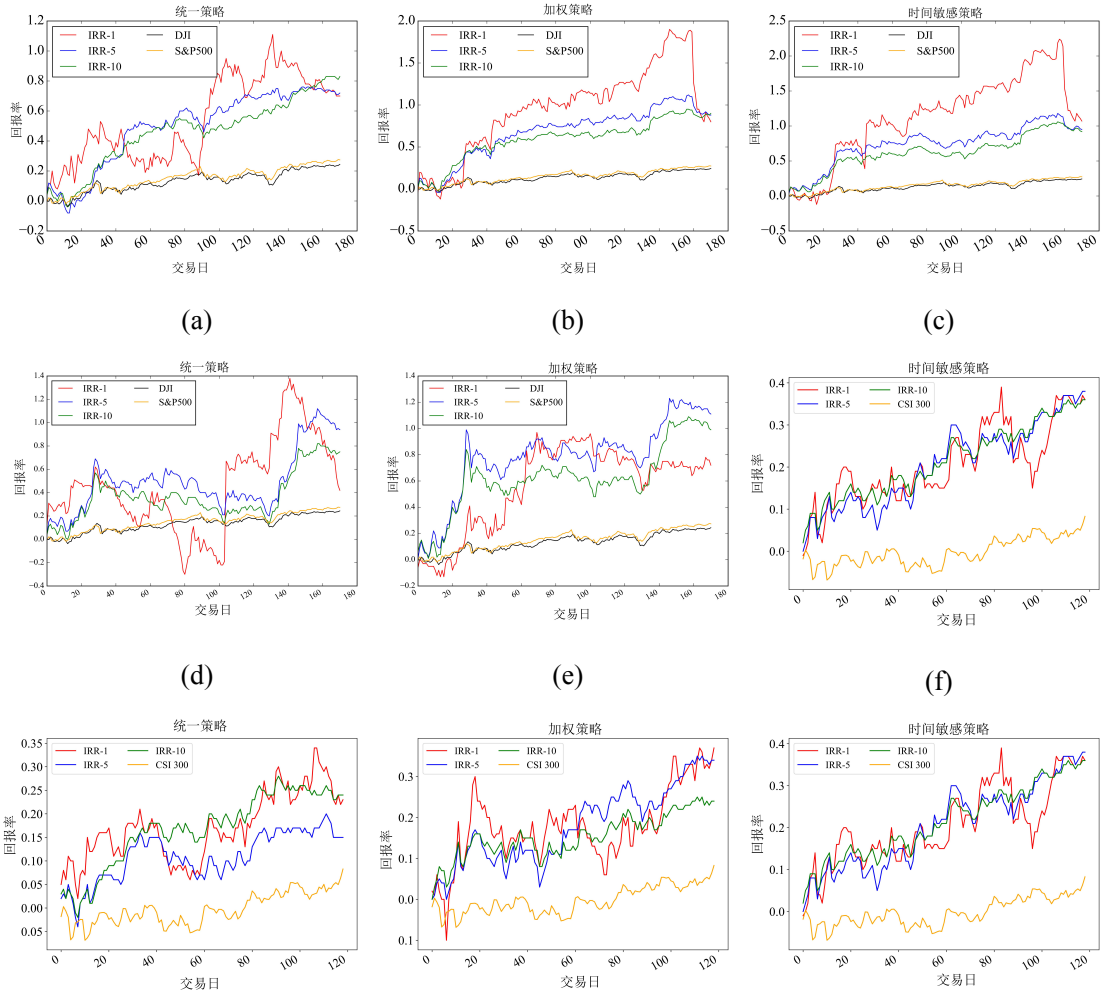


图 3-6 三种策略在三个指标（IRR-1、IRR-5、IRR-10）和三个市场指数（DJI、S&P 500 和 CSI 300 指数）方面的绩效比较。(a)-(c) 为 NASDAQ 数据集的结果，(d)-(e) 为 NYSE 数据集的结果，(g)-(i) 为 CSI 数据集的结果。

所为 0.4%）。关系比例越大，信息在图上的传播范围就越广。因此，这也表明完善股票间的关系、使股票之间的连接更贴近现实世界的价值。（3）在纳斯达克数据集上，采用维基关系的 RT-GCN（T）在 IRR-1 和 IRR-10 指标上表现优于采用行业关系的 RT-GCN（T），这与观察（2）相反。IRR-1 表现较低表明，从数百只股票中仅选择一只股票是一种高风险操作。一旦排名顺序不准确，投资回报率将遭受严重波动，而 IRR-10 表现较低的原因是，纳斯达克是以科技股为主，其中大多数股票通过行业关系相连。这些股票价格实际上由于 2020 年 3 月至 2020 年 4 月测试期间的急剧下跌而下降，导致在前 10 名中选择的科技股数量减少，因此 IRR-10 较低。

中国知网 https://www.cnki.net

表 3-6 在维基关系与产业关系两个关系数据集上的实验结果表现。

模型	维基关系				产业关系			
	MRR	IRR-1	IRR-5	IRR-10	MRR	IRR-1	IRR-5	IRR-10
	NASDAQ							
Rank_LSTM	0.040	0.36	0.44	0.83	0.040	0.36	0.44	0.83
RT-GCN (U)	0.037	0.21	0.58	0.62	0.051	0.33	0.77	0.79
RT-GCN (W)	0.036	0.43	0.66	0.71	0.053	0.35	0.79	0.87
RT-GCN (T)	**0.049**	**0.70**	**0.90**	**0.97**	**0.055**	**0.41**	**0.91**	**0.88**
	NYSE							
Rank_LSTM	0.040	0.36	0.44	0.83	0.040	0.36	0.44	0.83
RT-GCN (U)	0.037	0.21	0.58	0.62	0.051	0.33	0.77	0.79
RT-GCN (W)	0.036	0.43	0.66	0.71	0.053	0.35	0.79	0.87
RT-GCN (T)	**0.049**	**0.70**	**0.90**	**0.97**	**0.055**	**0.41**	**0.91**	**0.88**

2. 关系卷积和时序卷积对模型的影响

在进行的一项额外实验中，旨在探讨关系卷积模块和时序卷积模块对模型性能的贡献。该实验通过分别去除时序卷积模块和关系卷积模块，构造了两种变体模型：R-Conv 和 T-Conv。R-Conv 模型中移除了时序卷积模块，只留下了关系卷积模块，用以评估模型在缺乏时序处理能力时的表现。相反，T-Conv 模型中剔除了关系卷积模块，仅保留了时序处理功能，用于评估模型在没有考虑股票间关系依赖时的性能。性能比较结果见表 3-7。从结果中可以看出，R-Conv 的表现最弱，而 T-Conv 在性能上超过了 R-Conv。尽管如此，两种变体模型的表现均不及完整模型 RT-GCN (U)。R-Conv 模型虽然考虑了股票间的关系依赖，但由于股票预测任务高度依赖时间序列特征，其表现受限。相比之下，RT-GCN (U) 不仅利用时序卷积捕获了时间依赖性，还通过整合周边信息来丰富了各股票的特征表示。

表 3-7 模型在关系卷积和时序卷积两个模块的消融实验结果。

数据集	NASDAQ				NYSE				CSI			
	MRR	IRR-1	IRR-5	IRR-10	MRR	IRR-1	IRR-5	IRR-10	MRR	IRR-1	IRR-5	IRR-10
RT-GCN (U)	**0.055**	**0.45**	**0.71**	**0.82**	**0.040**	**0.65**	**0.88**	**0.96**	**0.017**	**0.13**	**0.15**	**0.19**
R-Conv	0.024	0.35	0.37	0.58	0.016	0.20	0.53	0.52	0.009	0.06	0.08	0.10
T-Conv	0.033	0.54	0.49	0.70	0.021	0.31	0.71	0.68	0.013	0.10	0.12	0.15

3.5.5 模型参数分析

本实验的目标是通过改变不同超参数的值来理解 RT-GCN 的性能表现，这包括训练窗口大小、模型中使用的特征数量，以及决定两种不同损失贡献度的平衡

中国知网 https://www.cnki.net

参数。所有接下来的比较实验都遵循相同的规则进行：在变化这些超参数中的任何一个时，保持其他超参数不变。由于时间敏感策略的出色表现，接下来的实验仅在该策略上进行。

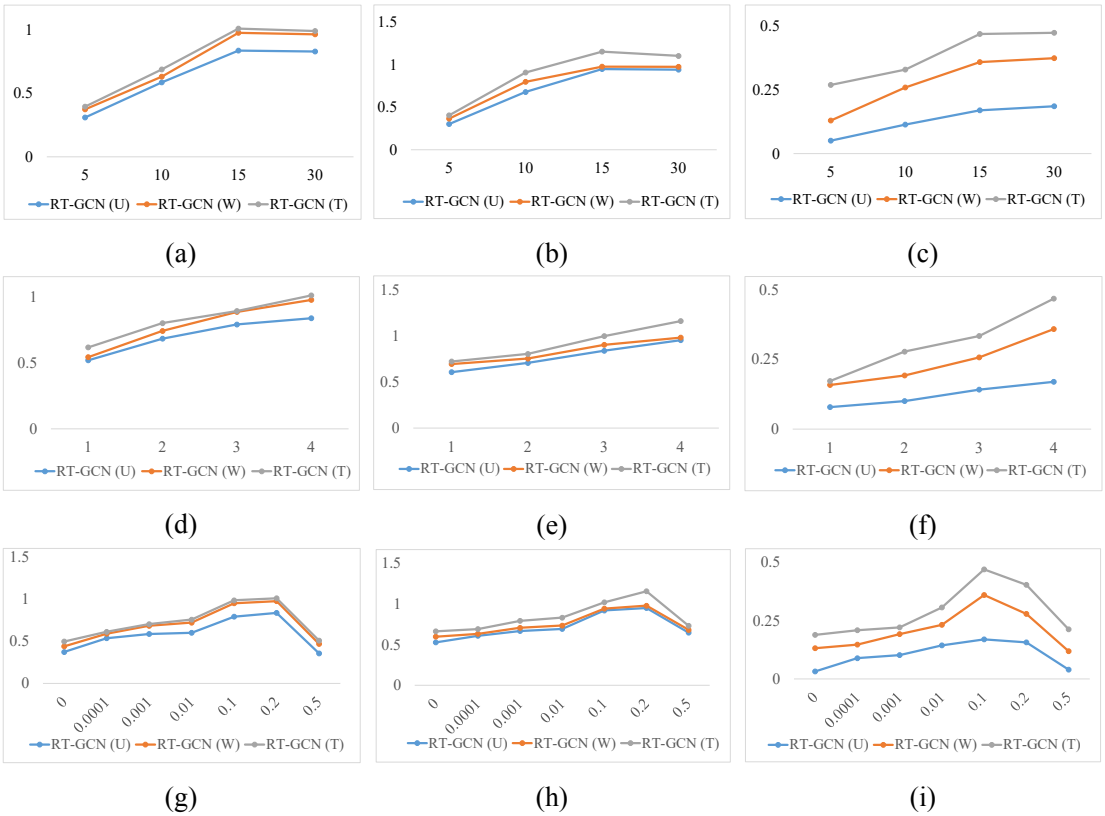


图 3-7 模型在训练时使用不同超参数时的表现。其中 (a)-(c) 为 NASDAQ、NYSE 和 CSI 三个数据集上采用不同窗口大小参数的模型表现；(d)-(f) 为 NASDAQ、NYSE 和 CSI 三个数据集上采用不同特征数量的模型表现；(g)-(i) 为 NASDAQ、NYSE 和 CSI 三个数据集上采用不同的 α 值的模型表现。x 轴表示超参数的值，y 轴代表 IRR 值。

1. 时间窗口分析

图 3-7(a)-(c) 展示了通过改变窗口大小（即用于预测的时间序列长度，亦称为时间跨度）对性能的影响。从图中可以发现，使用大约 15 天的序列长度进行训练可以获得更高的收益率（IRR）。训练时使用较少的天数（例如，5 天或 10 天）可能会导致性能下降。当窗口大小超过 15 天时，性能没有显著变化。

2. 特征数量分析

图 3-7(d)-(f) 中的结果显示，特征数量为 4 时性能较好。本实验使用收盘价、5 天移动平均线、10 天移动平均线和 20 天移动平均线作为候选特征。特征组合在表 3-8中展示。随着特征数量的增加，模型性能得到提升。可能的原因是更多的特

中国知网 https://www.cnki.net

表 3-8 特征组合的类型。

编号	特征组合
1	收盘价
2	收盘价, 5 日移动平均线
3	收盘价, 5 日, 10 日移动平均线
4	收盘价, 5 日, 10 日, 20 日移动平均线

征可以实现更好的拟合。因此，建议在 RT-GCN 中使用 4 个特征数量。未来将探索更有价值的特征，并将其应用于 RT-GCN。

3. α 值大小分析

图 3-7(g)-(i) 中展示了 $\alpha \in \{0, 0.0001, 0.001, 0.01, 0.1, 0.2, 0.5\}$ 的性能。可以发现，$\alpha = 0.2$ 对于纳斯达克（NASDAQ）和纽约证券交易所（NYSE）数据集而言表现最佳，而 $\alpha = 0.1$ 对于沪深指数（CSI）数据集表现最佳。较小的 α 表示模型在优化时更加强调回归损失而非排名损失。另外，当仅考虑回归损失（即 $\alpha = 0$）或对排名损失赋予过多权重（即 $\alpha = 0.5$）时，模型的表现较差。结果表明，适当融合回归损失和排名损失可以获得更好的性能。因此，建议根据不同数据集将 α 设置为 0.1 或 0.2。

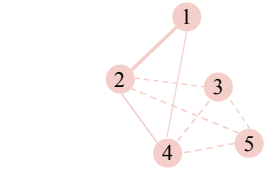
3.5.6 案例分析

本节中通过可视化 3 月 3 日至 4 月 3 日纳斯达克测试集中五只股票的预测结果，定性分析了 RT-GCN (T) 的性能。从图 3-8中可以观察到 RT-GCN (T) 能够在时间维度上出色地预测股票的回报率。具体来说，在 3 月 4 日，模型预测股票的回报率将会上升，实际上它们确实如此（3 月 4 日颜色变浅），3 月 16 日的预测同样准确。同时，本实验还观察到模型能够捕捉到空间维度上的关系依赖。例如，CDNS、CXDM 和 ICUI 的预测表现相似（从热图的颜色更好地观察），因为这三只股票更紧密地相连（从 (a) 中边的宽度更好地观察）。

3.6 本章小结

本章提出了一种基于关系-时序图卷积网络（RT-GCN）的资产排名预测模型（具体以股票为研究对象），旨在充分利用股票间的关系，实现股票投资的最高收益。首先将股票间的关系及其每日特征建模为一个关系-时序图。然后，应用图卷积网络和三种提出的关系感知策略，提取每个节点（股票）的关系-时序特征。最后，将这些特征用于排序的方式计算排名得分，得分最高的股票代表未来投资收

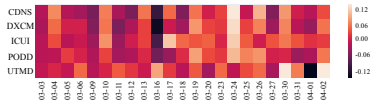
中国知网 https://www.cnki.net



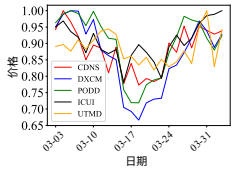
(a)

ID	股票标识符	股票代码
1	CDNS	Q608776
2	DXCM	Q15109865
3	PODD	-
4	ICUI	Q5969799
5	UTMD	-

(b)



(c)



(d)

图 3-8 以上是本章模型对日收益率预测的一个示例。(a) 展示了五只股票之间关系图的一部分。边的宽度表示两只股票之间的权重（由模型学习得到），权重越大，边越宽。实线边表示维基关系，虚线边表示行业关系。(b) 详细介绍了 (a) 中的股票信息。ID 是股票在 Wikipedia 数据库中的实体标识。(c) 是 (a) 中五只股票预测的收益率热图。(d) 是 (a) 中股票的真实价格（标准化后）曲线。

益最高。在纳斯达克、纽约证券交易所和中国证券指数上的实验结果展示了本章所提模型的有效性和效率。

本项工作聚焦于构建股票间的关系，建模其依赖性，并研究一个股票如何影响其他股票。一旦模型能够捕捉到股票间的依赖关系，外部信息如新闻和推文就可以丰富特征，并更准确地预测股票趋势，这可能是未来的研究方向。

中国知网 https://www.cnki.net

第四章 主动投资策略：基于多视野融合的资产配置

4.1 引言

资产配置，也叫资产优化（Portfolio Optimization, PO），旨在动态地在一系列资产之间分配财富以最大化投资回报，是金融、统计学和人工智能等多个研究领域中的一个基本研究问题[91]。传统方法往往使用手工制定的特征，如移动平均[92]和随机技术指标[93]来优化资产间的财富分配。也有研究尝试采用深度强化学习技术来解决资产配置问题[17,19,94,95]。与这些传统方法相比，强化学习（Reinforcement Learning, RL）被认为是一种更有前景的资产配置技术，因为它可以通过与金融环境的试错互动直接输出交易决策，以最大化投资回报。由于金融市场的非平稳性质难以被普通的 RL 策略所模拟，尽管现有研究已经做出相当大的努力，将 RL 应用于资产配置仍然面临巨大的挑战。

金融市场的非平稳性质对于资产配置构成了挑战，通过经济学理论——分形市场假说（Fractal Market Hypothesis, FMH）[96] 可以更好地理解金融市场的非平稳性质，从而为解决非平稳性问题提供一个思路。FMH 将混沌理论的概念应用于投资和经济学，为理解市场的动态本质和不同市场参与者之间的复杂相互作用[97]提供了一个框架。根据 FMH，金融市场由具有不同投资周期的各种投资者组成，这些投资者关注不同的信息，例如，长期投资者关注长期视野的信息，而短期投资者关注短期视野的信息。这些不同投资者之间的交易活动共同影响价格趋势，导致价格序列呈现出混沌、复杂和非平稳的特性。

有许多研究致力于解决这一非稳定性问题。一些预处理技术，比如移动平均[92,98]和信号转换技术[99,100]，被用于平滑价格序列[101,102]。在资产组合管理中，现有的基于强化学习的方法通过集成多个策略来增强模型的鲁棒性，在金融行业得到了广泛应用[18,19,100]。然而，这些方法并未考虑金融市场中存在不同投资周期信息，忽略了金融市场中普遍存在的多样化投资信息，限制了现有策略的表现。由于每个投资视野都具备独特的见解和模式，为了做出有效的交易决策，交易员需要全面考虑不同投资视野所提供的信息，正如例子 4.1所述。

例子 4.1 交易员可能会结合短期波动性分析和长期基本面评估来识别股票的投资价值。例如，一只股票尽管由于市场情绪的原因在短期内价格波动，但基于其产品创新性强，可能展现出强大的长期增长潜力。单一视野的投资策略可能忽略这只股票的未来潜力。

受到上述 FMH 理念启发，本文提出了多视野融合的交易策略。该策略共包含

中国知网 https://www.cnki.net

两阶段，目的是在交易决策过程中同时融合不同视野的投资预决策。策略的第一阶段将原始价格序列分解成若干子序列，每个子序列携带了与特定投资视野（如长期投资视野，短期投资视野）相对应的信息。随后，这些子序列被分别输入到各自的视野专属策略中（称为子策略），促使子策略学习对应的子序列特征与模式，并做出资产配置的预决策。第二阶段，本章介绍了一个跨视野策略，其目的是综合各个子策略所做出的预决策做出最终的交易决策，通过直接与金融环境进行交互获取奖励。然而，这种从环境获得的奖励反馈是基于跨视野策略作出的最终交易决策而得到的，这意味着奖励并不能直接反映每个子策略所做出的预决策的具体贡献，从而无法有效地指导每个子策略的优化。为了克服这一局限并确保两类策略（子策略与跨视野策略）的有效学习，本文提出了一个集中式评论家来评估所有策略的决策。具体来说，跨视野策略直接根据评论家估算的 Q 值进行优化，而每个子策略则依据集中式评论家内部的反事实机制推断出的优势值进行调整优化。

总结来说，本章的贡献有三个方面：

- 为了应对金融市场的非稳定性特性，本章提出了一个多视野融合策略来管理资产，该策略在做出交易决策时考虑了不同的投资视野的决策。据调研，本章所提出的方法在以往的研究中还未曾进行过探索。
- 基于理论推导，在本章的框架中设计了一种反事实机制，用以推导各个不同视野的贡献并引导不同子策略的优化。
- 在三个真实世界股票市场数据集上的广泛实验结果表明，与最先进的方法相比，本章所提的框架实现了显著的改进。

4.2 相关工作

本节将现有的资产配置方法分为两大类：传统方法和基于强化学习的方法。

4.2.1 传统的资产配置方法

为实现高利润投资，传统研究提出指数梯度（Exponential Gradient, EG）[103,104]、在线牛顿步骤（Online Newton Step, ONS）[105]、M0[106]、恒定再平衡投资组合（Constant Rebalanced Portfolios, CRP）[107]、反相关（Anticor）[108] 以及通用投资组合（Universal Portfolios, UP）[109] 等方法。Helmbold 等人[103] 提出了 EG 策略，该策略将相对熵作为正则项纳入。EG 的优化目标是关注上一期表现最好的股票，同时保持新投资组合接近前一个投资组合，以减少交易成本。Agarwal 等人[105] 提出了 ONS，它与 EG 类似，但用 ℓ_2 范数正则项替换了相对熵正则项，并通过在线凸优化技术优化模型。CRP[107] 是一种更直接的策略，每个周期都将投资

中国知网 https://www.cnki.net

组合重新平衡到一个固定的组合。UP[109]最初将一定比例的财富投资给每个执行CRP策略的资产配置者，并让管理者运行它。最后，UP策略将这些投资组合策略中个别管理者的财富汇总。与UP中的无分布假设不同，反相关策略[108]假设市场遵循均值回归原则。为了利用均值回归，它从统计的角度出发，利用了正滞后交叉相关性和负自相关性的一致性进行资产配置。

还有一些其他方法也利用均值回归来选择投资组合，例如被动攻击型均值回归（Passive Aggressive Mean Regression, PAMR）[29]、置信加权均值回归（Confidence Weighted Mean Reversion, CWMR）[30]、在线移动平均回归（Online Moving Average Reversion, OLMAR）[92]以及鲁棒中位数回归（Robust Median Reversion, RMR）[110]。PAMR[29]的目标是设计一个损失函数来反映均值回归属性，即如果基于最后相对价格的预期收益大于阈值，损失将线性增加；否则，损失将为零。CWMR[30]进一步利用二阶投资组合信息，即投资组合权重的方差，遵循均值回归原则，并通过置信加权在线学习。与PAMR和CWMR隐含地假设单周期均值回归不同，OLMAR[92]利用多周期均值回归来避免在真实数据集上的一种失败情况。观察到PAMR、CWMR和OLMAR因为没有考虑数据中的噪声和异常值而受到估计错误的影响，Huang等人[110]提出了RMR，通过鲁棒的ℓ_1中位数估计器来探索均值回归。然而，上述所有方法都严重依赖于手工特征，例如随机技术指标或移动平均，导致这些模型在非稳定市场中表现出较差的泛化能力，并且无法实现稳定的利润。

4.2.2 基于强化学习的资产配置

近年来强化学习作为一个新兴的方法广泛应用于资产配置并取得亮眼的表现。基于强化学习的方法运用强化学习来优化特定的交易策略并实现奖励最大化[111,112]。一些研究[18,64,83,95]结合了深度学习来提取鲁棒特征，以便更好地表示投资组合并用强化学习优化投资回报。其中一个流行的方法是相同独立评估器的集成（Ensemble of Identical Independent Evaluations, EIIE）[18]。EIIE通过使用三个相同的深度学习神经网络实现特征提取，这三个网络分别是卷积神经网络（CNN）、循环神经网络（RNN）和长短期记忆网络（LSTM）。为了改进EIIE，Zhang等人[113]结合了一个特征注意力模块和一个资产注意力模块，以实现对资产的更好信息提取。Xu等人[83]和Zhang等人[95]都更加强调在使用强化学习算法进行资产配置时，捕捉序列模式和资产间的相关性。一些研究如DeepTrader[17]和SARL[94]在做出投资决策时也考虑了包括金融新闻和市场条件在内的额外因素。Lien等人[114]提出了一个深度资产配置模型，该模型使用对比奖励和奖励平滑来

中国知网 https://www.cnki.net

帮助智能体学习鲁棒的表示。一些研究[10,19,115]提议将集成策略与强化学习结合起来，以处理金融市场的非稳定性。Ma 等人[115]提议训练两个不同的智能体，在两种不同的趋势下学习最佳交易策略，并动态切换它们以找到最佳的投资组合策略。Lee 等人[10]介绍了协作型多智能体资产管理系统 (Cooperative Multi-agent Portfolio Management System, MAPS)，一个协作的多智能体资产配置系统，其中每个智能体通过不同的损失函数指导以实现自己的最大奖励。类似地，Niu 等人[19]提出了 MetaTrader，它将模仿目标整合到强化学习框架中，以提高各种策略的表现多样性。然而，与本章所提出的框架不同，之前的研究如 MAPS 和 MetaTrader 忽视了市场内不同投资视野所包含的信息。与前面所提的研究相比，本章工作整合了多种投资视野，通过利用这些视野信息对市场的独特见解，本章方法为资产管理提供了更全面、更深刻的交易决策意见。

4.3 预备知识

资产配置的目标是重新分配财富在 m 个资产（金融产品）上的比例，以便在一系列 (T 个) 交易时间步内最大化投资回报。每个资产具有 d 种特征，例如当 $d = 4$ 时资产特征为开盘价、最高价、最低价和收盘价。设 $p_{t,i} \in \mathbb{R}^d$ 表示在时间步 t 资产 i 的价格，其中 $t \in \{1, 2, ..., T\}$ 且 $i \in \{1, 2, ..., m\}$。本章还定义 $P_t = \{p_{t,1}, p_{t,2}, ..., p_{t,m}\} \in \mathbb{R}^{m \times d}$，以及 $\mathcal{P}_t = \{P_{t-z}, ..., P_{t-1}\} \in \mathbb{R}^{z \times m \times d}$ 为在时间步 t 的所有资产的价格，以及关于时间步 t 的前 z 个时刻价格序列。本章工作只考虑收盘价，因此 $d = 1$。

本章将资产优化（PO）任务形式化为一个广义的马尔可夫决策过程（Markov Decision Process, MDP），正如许多其他研究所做的那样[83,95,116]。MDP 可以定义为一个四元组 (S, A, Z, R)，其中 $S = \{s_1, ..., s_T\}$ 是当前状态的有限集合，$A = \{a_1, ..., a_T\}$ 是动作的集合，Z 是一个随机状态转移函数，用于确定在给定动作 $a_t \in A$ 的情况下，从当前状态 s_t 转移到下一个状态 s_{t+1}，表示为 $s_{t+1} \sim Z(s_t, a_t)$，而 $R = \{r_1, ..., r_T\}$ 是根据动作从环境中获得的一系列奖励。请注意，本工作假设下一个状态由市场决定，而本章所提框架中的交易不会影响每只股票的状态转移，即 $s_{t+1} \sim Z(s_t)$。在资产配置的背景下，MDP 可以定义如下：

- 状态 s：一组描述 m 项资产当前状态的特征集。通常，各种类型的信息，如交易量、历史价格、情绪得分和财务报表等，都可以用作资产的当前状态。本章的研究中定义了两种不同类型的策略：针对特定视野的子策略和跨视野策略。针对这两种策略，本章设计了两种相应的状态作为输入，这些状态包含的信息不仅仅是价格序列 $\mathcal{P}_t$，正如第 4.4.2节所描述的。

中国知网 https://www.cnki.net

- 动作 a：通过一个投资组合向量指定动作 $a_t = [a_{t,1}, a_{t,2}, ..., a_{t,m}]^\top \in \mathbb{R}^m$，其中 $a_{t,i} > 0$ 表示在时间步 t 关于资产 i 的财富比例，且 $\sum_{i=0}^{m} a_{t,i} = 1$。
- 奖励 r：环境（即金融市场）基于智能体在当前状态的动作给出的奖励。一个常用的奖励是投资组合价值的对数回报率：$r_t = \log(\frac{a_t^\top x_t}{a_{t-1}^\top x_{t-1}})$，其中 x_t 定义为收盘价增长比例 $\frac{p_t}{p_{t-1}} - 1$，而 $a_t^\top x_t$ 表示在时间步 t 的累积回报率。
- 策略 π：策略本质上是一个神经网络，它接收当前状态 s 作为输入，输出动作的概率分布。第 4.4.2节介绍了两种类型的策略。

例子 4.2进一步解释了在 MDP 定义下资产配置的详细过程。

例子 4.2 假设投资组合包含三种资产（股票 A、B 和 C）。当前状态 s_t 包括过去一周每只股票的历史价格等特征。基于当前状态，操作 a_t 可能是一个投资组合向量，例如 [0.5, 0.3, 0.2]，表示在时间 t，本章提出的策略将 50% 的财富分配给股票 A，30% 分配给股票 B，20% 分配给股票 C。这一操作反映了策略对当前对市场的理解并进行了多样化的投资。执行操作 a_t 后，市场发生变化，环境根据投资组合的对数回报率计算奖励 r_t。策略 π 是一个神经网络，它根据奖励自我更新，学习如何最大化累积奖励。

4.4 模型：多视野交易策略

根据分形市场假说（Fractal Market Hypothesis, FMH）[96]，由于不同投资视野的交易者参与交易导致股票市场价格呈现非稳定性，这使得资产配置任务变得难以解决。受到这一理论的启发，本章提出了一种基于强化学习（RL）的两步骤方法来进行资产配置，该方法通过融合其他具有不同投资视野的决策意见，以缓解金融市场的非稳定性所带来的挑战。该框架在图 4-1中进行了说明。接下来，本章将深入探讨如何分割不同策略的视野信息以确保它们专注于特定的投资视野，以及两种不同策略的学习过程。本章在表 4-1中总结了经常使用的符号。

4.4.1 视野信息分解

本节介绍如何为不同视野的策略生成专属的视野信息。众所周知，金融价格序列可以被视为一个具有多个频率的信号。低频成分描绘了时间序列的粗糙结构和长期趋势，而高频成分反映了短期波动。为了分解出不同视野的信息，本章采用了一种分解技术，将原始价格序列基于频率带分解成不同的子序列。例如，将原始序列分解成包含与长期投资者相关的低频成分的子序列，以及包含与短期投资者相关的高频成分的子序列。这种分解可以在不同的粒度级别上进行。通过将原始序列分解为不同子序列，每个策略能够针对其各自交易视野的信息，做出该

中国知网 https://www.cnki.net

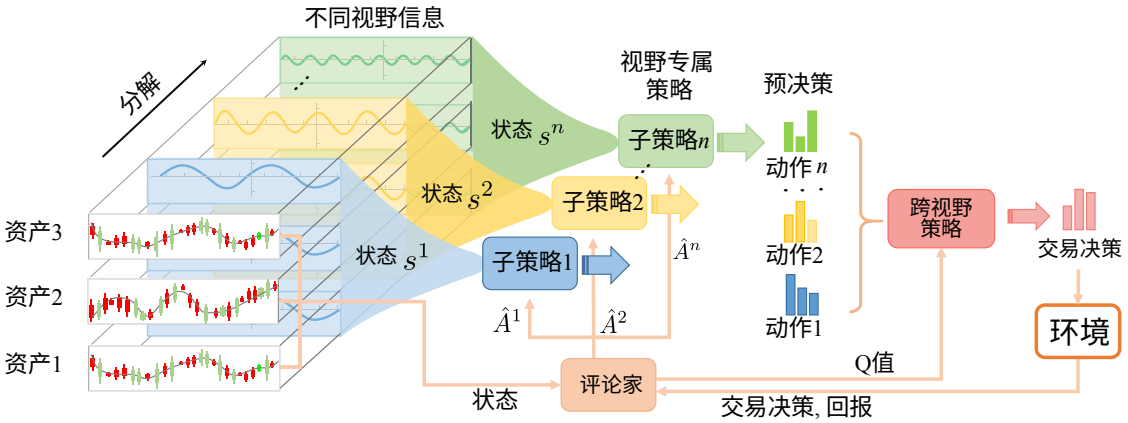


图 4-1 本章提出框架的示意图。$\hat{A}^k$ 表示通过反事实机制推理出的策略 k 的优势值。*Env* 代表金融环境。

表 4-1 常用符号总结。

常用符号	符号描述
m	资产的数量
n	子策略的数量
$\mathcal{P}, \mathcal{P}^k$	所有资产的价格信息；分解后的价格信息作为子策略 k 的输入
$s^k, \tilde{s}$	输入给第 k 个子策略和跨视野策略的状态
$a^k, \tilde{a}$	第 k 个子策略和跨视野策略所采取的动作
π^k, θ^k	第 k 个子策略及其参数
$\tilde{\pi}, \tilde{\theta}$	跨视野策略及其参数
$\hat{A}^k, \mathfrak{B}^k$	第 k 个子策略的优势与该子策略的反事实基线

投资视野下相应的决策。

有几种方法可以用来分解价格序列，例如短时傅里叶变换[117,118]和小波变换[119]。本章的工作采用了离散小波变换（Discrete Wavelet Transform, DWT）方法，因为它在分析非稳定时间序列方面起到了至关重要的作用，这一点在交通流量[120]和金融时间序列[96]等多个领域都有所体现。DWT 像带通滤波器一样，将给定信号 $x(t)$ 分解成多级频率带。具体来说，在第一级分解中，信号 $x(t)$ 通过低通滤波器（Low-pass Filter, LPF）和高通滤波器（High-pass Filter, HPF）被分解成近似系数 α^1 和细节系数 d^1。近似系数 α^1 保留了长期视野信息，而细节系数 d^1 则

中国知网 https://www.cnki.net

保留了短期视野信息。分解可以被表述为：

$$\begin{cases} \alpha^1 = \int x(t)\varphi(t)dt \\ d^1 = \int x(t)\psi(t)dt \end{cases}, \tag{4-1}$$

其中，α 和 d 的上标表示离散小波变换（DWT）的层级，而 $\varphi(\cdot)$ 和 $\psi(\cdot)$ 分别代表低通滤波器（LPF）和高通滤波器（HPF）[121]。前一层级的近似系数输出将在下一次迭代中被分解，直到达到指定的粒度层级。在本章的框架中采用了 Haar 小波变换，因为它在金融时间序列处理中得到广泛应用[14]。

在获取了系数之后，将要移除的频带的系数进行掩码处理，然后将剩余的系数逆变换回时间信号 $x'(t)$。例如，考虑到分解层级为 1，当使用 DWT 处理 $\mathcal{P}_t = \{P_{t-z}, ..., P_{t-1}\} \in \mathbb{R}^{z\times m\times d}$ 时，可以获得近似系数 $\alpha^1 \in \mathbb{R}^{z/2\times m\times d}$ 和细节系数 $d^1 \in \mathbb{R}^{z/2\times m\times d}$。通过掩码 α^1 并与 d^1 进行逆向变换，可以获得高频尺度信息；反之，可以获得低频尺度信息。分解的粒度 $\mathcal{P}_t^k$ 等于策略的数量，表示为 $\mathcal{P}_t^k, k \in \{1, 2, ..., n\}$。分解结果如图 4-2所示。

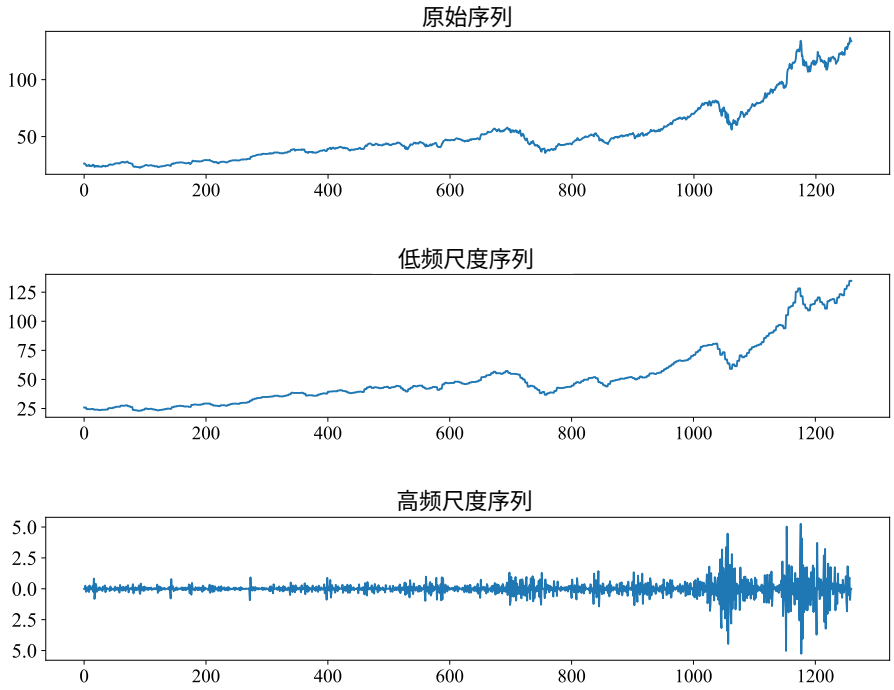


图 4-2 将分解的视野信息进行可视化，粒度为 2，代表长期和短期视野。x 轴代表时间步，y 轴代表分解的值。低频尺度对应于长期视野信息，而高频尺度代表短期视野信息。

中国知网 https://www.cnki.net

4.4.2 策略学习

本节将详细阐述 (i) 如何处理两种类型的策略学习以及 (ii) 在提取稳健特征时，如何有效地模拟时间动态和资产关系。

（1）子策略学习

假设到有一组由 $\theta = \{\theta^1, ...\theta^n\}$ 参数化的 n 个子策略，表示为 $\pi = \{\pi^1, ...\pi^n\}$。每个策略 $k \in \{1, ..., n\}$ 根据状态 s^k 做出预先决策。子策略的输入（状态）、输出（动作）和优化如下：

(i) **状态**: 状态 s^k 包含了从 $\mathcal{P}^k$ 通过神经网络（详见第 4.4.3节）提取出的投资视野特征 H^k。除了视野特征外，状态 s^k 还包含了子策略的 ID 和上一时间步执行的动作，这使得策略表现出多样性，并防止连续两个时间步之间的动作变化幅度过大，如图 4-3(a) 所示。

（ii）**动作（预决策）**：动作 $a^k_{t,i} > 0$ 代表第 k 个子策略在时间步 t 对资产 i 分配的财富比例，且 $\sum_{i=0}^{m} a^k_{t,i} = 1$。

（iii）**优化**：子策略 k 的预期回报梯度为：

$$\nabla_{\theta^k} J(\theta^k) = \mathbb{E}_{a^k \sim \pi^k}[\nabla_{\theta^k} \log \pi^k(a^k|s^k)\Psi_k], \tag{4-2}$$

其中 Ψ_k 表示评论家对策略 k 估算的价值。实际上，Ψ_k 有许多选择方案，例如，奖励的期望 [122]、状态价值 [123]、动作优势 [124] 以及时序误差（Temporal Difference Error, TD-error）[125]。然而，这些方案是次优的，因为它们不能显式地反映每个子策略的预决策对最终交易决策的贡献。本节通过引入了一种反事实机制来计算每个子策略的动作优势，使得多个子策略能够以更有效的方式进行训练。

（2）跨视野策略学习

与子策略不同，跨视野策略 $\tilde{\pi}$ 由 $\tilde{\theta}$ 进行参数化，基于子策略所做出的预决策来做出最终的交易决策 $\tilde{a}$。在输入（状态）、输出（动作）和优化方面，与子策略相比有细微的差异：

（i）**状态**：跨视野策略的状态 $\tilde{s}$ 包含了所有子策略所做出的预决策 $\{a^1, ..., a^n\}$。此外，状态 $\tilde{s}$ 还融合从所有资产的原始价格序列 $\mathcal{P}$ 中提取的特征，这特征提供了关于整体市场条件的信息。

（ii）**动作**：动作 $\tilde{a}_{t,i} > 0$ 表示跨视野策略在时间步 t 对资产 i 分配的财富比例，且 $\sum_{i=0}^{m} \tilde{a}_{t,i} = 1$。

（iii）**优化**：跨视野策略的梯度写为：

$$\nabla_{\tilde{\theta}} J(\tilde{\theta}) = \mathbb{E}_{\tilde{a} \sim \tilde{\pi}}[\nabla_{\tilde{\theta}} \log \tilde{\pi}(\tilde{a}|\tilde{s})Q], \tag{4-3}$$

中国知网 https://www.cnki.net

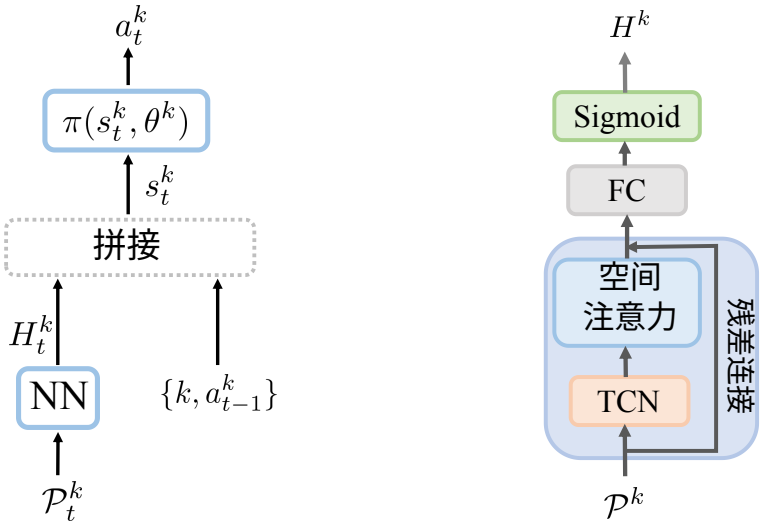


图 4-3 策略的输入示意图与其网络结构。

其中 Q 是由评论家网络估计的 Q 值。与专属视野策略的优化不同，跨视野策略的优化是基于评论家网络直接估计的 Q 值进行的。这是因为评论家是基于从跨视野策略与环境的交互直接获得的奖励进行优化的。将预决策和所有资产的原始价格序列包含在状态 $\tilde{s}$ 中，使跨视野策略能够有效利用多种投资视野，捕捉市场的动态变化。子策略和跨视野策略的学习过程如算法 4-1所示。

4.4.3 网络结构

为了应对挑战（ii），本文提出的策略采用了一个精心设计的神经网络结构，用于提取资产之间的时间动态和空间关系。这一结构如图 4-3(b) 所示。

时间卷积块：如图 4-3(b) 所示，首先利用时序卷积网络（Temporal Convolution Network, TCN）块 [126] 来提取资产的时间动态。与循环神经网络相比，TCN 具有几个吸引人的特性，包括促进并行计算、缓解梯度爆炸/消失问题，并展现更长的有效记忆 [17,126]。在对 $\mathcal{P}^k$ 进行 TCN 操作后，可以获得了一个输出张量，记为 $\hat{H}^k \in \mathbb{R}^{m \times f \times z}$，其中 f 是隐藏特征的维度。

空间注意力层：随后，通过采用了注意力机制 [75] 来处理不同资产之间的空间关系。给定 TCN 的输出向量，空间注意力权重的计算可以表示为如下：

$$\mathcal{S} = V_s \cdot sigmoid((\hat{H}^k W_1) W_2 (W_3 \hat{H}^{k\top})^\top + b_s), \tag{4-4}$$

其中 $\top$ 表示转置操作，$W_1 \in \mathbb{R}^z$, $W_2 \in \mathbb{R}^{f \times z}$, $W_3 \in \mathbb{R}^f$ 和 $V_s \in \mathbb{R}^{m \times m}$ 是需要学习的参数，$b_s \in \mathbb{R}^{m \times m}$ 是偏置向量。矩阵 $\mathcal{S} \in \mathbb{R}^{m \times m}$ 通过行标准化，以表示资产间的相关

中国知网 https://www.cnki.net

算法 4-1 多视野融合的交易算法

Input: 原始价格序列 $\mathcal{P}$
分解 $\mathcal{P} \longrightarrow \{\mathcal{P}^1, ..., \mathcal{P}^n\}$;
初始化 $\{\theta^1, ...\theta^n\}$, $\tilde{\theta}$ 和 θ^c;
while 未到训练结束 **do**
  foreach $\{\mathcal{P}^1, ..., \mathcal{P}^n\}$ **do**
    $s^k = (\mathcal{P}^k, k, a^{k-1})$;
    $a^k = \pi^k(s^k; \theta^k)$;
  end
  $\tilde{s} = (a^1, ..., a^n, \mathcal{P})$;
  $\tilde{a} = \tilde{\pi}(\tilde{s}; \tilde{\theta})$;
  $\mathbf{x} = (\tilde{s}, \tilde{a})$;
  $Q = f(\mathbf{x}; \theta^c)$;
  foreach $\{\pi^1, ..., \pi^n\}$ **do**
    使用集中式评论家中的反事实机制计算优势 $\hat{A}^k$，根据公式 (4-8);
    根据其优势值 $\hat{A}^k$ 优化 π^k，根据公式 (4-2);
  end
  根据价值 Q 优化 $\tilde{\pi}$，根据公式 (4-3);
end
Result: 训练好的子策略 $\{\pi^1, \ldots, \pi^n\}$ 和跨视野策略 $\tilde{\pi}$

性。

$$\mathcal{S}_{i,j} = \frac{\exp(S_{i,j})}{\sum_{u=1}^{m} \exp(S_{i,u})}, \forall 1 \leq i \leq m. \tag{4-5}$$

本工作采用残差连接 [80] 来缓解深度学习中的梯度消失问题。具体来说，从 $\mathcal{P}^k$ 提取的最终表示由 $H^k = \mathcal{S} \times \hat{H}^k + \mathcal{P}^k$ 表示，并且将其与智能体 ID k 和上一时间步骤执行的动作 a_{t-1}^k 连接作为最终状态 s_t^k。状态 s_t^k 被送入策略网络，转换为一个维度为 m 的向量 $\hat{a}^k$。最后，向量 $\hat{a}^k$ 被标准化为一个满足约束 $\sum_{i=0}^{m} a_i^k = 1$ 的动作 a^k。

4.4.4 中心化评论家

评论家学习：训练评论家网络 $f_t(\theta^c)$ 是强化学习的另一个关键组成部分。评论家网络的参数 θ^c 更新以最小化如下的损失函数：

$$\mathcal{L}(\theta^c) = (y^{(\lambda)} - f_t(\mathbf{x}; \theta^c))^2, y^{(\lambda)} = (1-\lambda) \sum_{n^*=1}^{\infty} \lambda^{n^*-1} G_t^{(n^*)}, \tag{4-6}$$

其中 $y^{(\lambda)}$ 是 n 步返回的混合，$\mathbf{x}$ 是评论家网络的输入。注意，为了区分代表策略数量的符号 n，本文对 n 步返回使用 n^* 符号表示。$G_t^{(n^*)}$ 的计算可以定义为公式 (4-7)：

$$G_t^{(n^*)} = \sum_{l=1}^{n^*} \gamma^{l-1} r_{t+l} + \gamma^{n^*} f_{t+n^*}(\mathbf{x}; \theta^c), \tag{4-7}$$

其中，γ 是折扣因子。评论家网络由一个双层全连接网络实现。输入向量 $\mathbf{x}$ 包含了多个组成部分，包括代表整个市场状态的所有资产的原始价格序列 $\mathcal{P}_t$，所有策略接收到的视野信息 $\{\mathcal{P}_t^1, \mathcal{P}_t^2, \ldots, \mathcal{P}_t^n\}$，由跨视野策略采取的交易动作 $\tilde{a}$，以及所有

中国知网 https://www.cnki.net

策略的编号 $\{1,2,\ldots,n\}$。将这些特征作为评论家网络的输入可以获得更准确的动作评估，为所有子策略的训练提供更好的指导。

然而，仅基于评论家网络估算的相同价值来优化所有策略，忽略了每个子策略对累积回报的贡献。为了解决这个问题，本章在评论网络中引入了一个反事实机制来评估每个子策略的优势。这一机制能够评估每个子策略的独特优势，并提高策略训练的有效性。

反事实机制：在衡量子策略对利润的贡献时，我们可能会问：如果没有这个策略的动作（或者该策略的操作被默认动作替代），同时其他策略的动作保持不变，会发生什么？这种事后推断出的可能结果被称为反事实反应，其中反事实的概念描述了通过推理来评判某个动作的后验结果。

本章工作通过采用反事实机制来评估每个子策略的动作。在本章的场景中，评估子策略贡献的最直接方法是计算去除该子策略动作后的优势值（即去除子策略动作后的累积回报期望与未去除该子策略动作累积回报期望之间的差）。在本章所提出的框架中，累积回报由状态-动作值表示，记为 $Q(\cdot)$，该值由评论网络估计。考虑到最终的交易决策是基于所有子策略的合作结果，直接移除某一子策略的动作并不是最佳方法。因此，本章选择了另一种替代方案，即用默认动作替换子策略的动作，然后计算预估的累积回报，也称为反事实基线，记为 $\mathfrak{B}^k$。因此，反事实机制可以描述如下：（1）集中式评论者根据输入 $\mathbf{x}$ 估计交易决策 $\tilde{a}_t$ 的 $Q(\mathbf{x},\tilde{a}_t)$；（2）通过将 $Q(\mathbf{x},\tilde{a}_t)$ 与子策略 k 当前动作的反事实基线 $\mathfrak{B}^k$（即用默认动作替换其当前动作）进行比较，可以为子策略 k 计算一个优势值 $\hat{A}^k$。换句话说，$\hat{A}^k$ 代表了策略当前动作（资产上的分配比例）相对于默认动作的优势。$\hat{A}^k > 0$ 意味着当前动作比策略 k 的默认动作更有利。

然而，如何计算 $\mathfrak{B}^k$ 是一个难题并吸引了很多研究者的关注。目前已经提出了许多方法来计算基线 $\mathfrak{B}^k$，例如，使用基线模拟器[127]或计算所有可能动作值的期望[128]。然而，上述解决方案主要关注于离散动作空间，而资产配置任务却是一个连续动作空间的任务。通常，对于连续动作空间的任务，动作是从高斯分布中以一定概率抽取的样本，其中均值和标准差是基于状态/观测来学习的。考虑到在连续空间中无法列举所有动作，本章采取了一种替代方法，通过将子策略 k 的动作替换为高斯均值 μ^k（由策略网络学习）来计算基线 $\mathfrak{B}^k$。选择 μ^k 作为计算 $\mathfrak{B}^k$ 的默认动作，可以优化高斯分布，以更高的概率采样有优势的动作。如果 $\hat{A}^k > 0$，则当前选择的动作被鼓励，高斯分布的 μ^k 将在下一次更新时移向靠近所选动作的区域（因为 μ^k 所在的区域有更高的被抽样概率）。如果 $\hat{A}^k < 0$，表明当前动作不被鼓励，高斯分布的 μ^k 将在下一次更新时远离所选动作附近的区域，最小化这一动

中国知网 https://www.cnki.net

作被抽样的概率。优势值 $\hat{A}^k$ 的计算过程数学表达示如下公式：

$$\begin{aligned}\hat{A}^k &= Q(\mathbf{x}, \tilde{a}_t) - \mathfrak{B}^k \\ &= Q(\mathbf{x}, \tilde{a}_t) - Q(\mathbf{x}, (a_t^{-k}, \mu_t^k)),\end{aligned} \tag{4-8}$$

其中，a^{-k} 表示除了子策略 k 动作以外的其他策略的交易动作。通过这种方式，本章通过利用集中式评价者来分析仅策略 k 的动作发生变化的场景，从而为每个子策略计算出一个独特的反事实基线。下方提供了一个反事实计算的例子以及展示反事实机制不会影响到最优收敛。

例子 4.3 本章的算法集成了多个不同视野的子策略。具体来说，子策略 k 在时间 t 将资产的 60% 分配给股票 A，40% 分配给股票 B。为了评估子策略 k 的动作影响，本章首先将子策略 k 做的动作替换为一个默认动作（例如，将股票 A 和 B 的分配均等化为 50%-50%），而其他策略的动作保持不变。然后，计算在执行这个默认动作后的预期累积回报（使用状态-动作价值 Q）并与子策略原始动作的预期回报进行比较。$\hat{A}^k > 0$ 表示原始动作（60%-40% 分配）优于默认动作（50%-50% 分配），标志着该子策略的动作在时间 t 的影响是正面的。相反，$\hat{A}^k < 0$ 表明原始动作（60%-40% 分配）相对于均等化分配（50%-50%）可能不具有优势。

以下证明反事实机制推理不会影响多策略强化学习的最优收敛：

证明: 反事实基线的梯度由以下公式给出：

$$g = \mathbb{E}_\pi \left[\sum_{k=1}^{n} \nabla_{\theta^k} \log \pi^k \left(a^k \mid s^k \right) \hat{A}^k \right], \tag{4-9}$$

其中，$\hat{A}^k$ 表示为

$$\hat{A}^k = Q(\mathbf{x}, \tilde{a}) - \mathfrak{B}^k, \tag{4-10}$$

$\mathfrak{B}^k$ 是在 Eq. (4-8)中定义的反事实基线。因此，Eq. (4-9) 可以表示为：

$$\begin{aligned} g &= g_1 + g_2 \\ &= \mathbb{E}_\pi \left[\sum_{k=1}^{n} \nabla_{\theta^k} \log \pi^k \left(a^k \mid s^k \right) Q(\mathbf{x}, \tilde{a}) \right] - \mathbb{E}_\pi \left[\sum_{k=1}^{n} \nabla_{\theta^k} \log \pi^k \left(a^k \mid s^k \right) \mathfrak{B}^k \right]. \end{aligned} \tag{4-11}$$

首先，考虑 g_2:

$$g_2 = -\mathbb{E}_\pi \left[\sum_{k=1}^{n} \nabla_{\theta^k} \log \pi^k \left(a^k \mid s^k \right) \mathfrak{B}^k \right]. \tag{4-12}$$

中国知网 https://www.cnki.net

设 d_π 是由 [129] 定义的状态分布：

$$\begin{aligned} g_2 &= -\int_s d_\pi(s) \sum_{k=1}^{n} \int_{a^{-k}} \pi^{-k}\left(a^{-k}|s^{-k}\right) \cdot \int_{a^k} \pi^k\left(a^k|s^k\right) \nabla_{\theta^k} \log \pi^k\left(a^k|s^k\right) \mathfrak{B}^k \\ &= -\int_s d^\pi(s) \sum_{k=1}^{n} \int_{a^{-k}} \pi^{-k}\left(a^{-k}|s^{-k}\right) \cdot \int_{a^k} \nabla_{\theta^k} \pi^k\left(a^k|s^k\right) \mathfrak{B}^k \\ &= -\int_s d^\pi(s) \sum_{k=1}^{n} \int_{a^{-k}} \pi^{-k}\left(a^{-k}|s^{-k}\right) \mathfrak{B}^k \cdot \int_{a^k} \nabla_{\theta^k} \pi^k\left(a^k|s^k\right) \\ &= -\int_s d^\pi(s) \sum_{k=1}^{n} \int_{a^{-k}} \pi^{-k}\left(a^{-k}|s^{-k}\right) \mathfrak{B}^k \nabla_\theta 1 \\ &= 0. \end{aligned} \tag{4-13}$$

显然，基线 $\mathfrak{B}^k$ 减少了方差但不改变期望的梯度，因此反事实基线不会影响强化学习的收敛性。

期望策略梯度的剩余部分 g_1 如下：

$$g_1 = \mathbb{E}_\pi \left[\sum_{k=1}^{n} \nabla_{\theta^k} \log \pi^k\left(a^k|s^k\right) Q(\mathbf{x}, \tilde{a}) \right]. \tag{4-14}$$

在文献 [130] 中证明了，遵循此梯度的方法会收敛到期望奖励的局部最大值。 ■

4.5 实验结果与分析

首先，本节将提出的框架在三个真实世界股票市场上的盈利能力与其他主流方法进行对比，以评估其有效性。此外，本节进行了一系列实验，以深入探讨提出的模型的性能，对子策略数量与不同网络结构的影响进行了深入的探究。这些实验能够评估不同分解层次（粒度）与不同数量的子策略的有效性，并分析它们如何影响框架的整体性能。进一步地，本节研究了反事实机制的效果来理解它对提高策略训练有效性的贡献。

4.5.1 实验设置

本节将描述数据集统计、对比模型、评估指标以及框架的实现细节。

1. 数据集介绍

本节所采用的数据集源自真实世界市场，对框架的全面评估提供支持。数据收集包括来自美国纳斯达克交易所（简称美国市场）、香港股票交易所（简称香港市场）和上海证券交易所（简称中国市场）的成分股每日收盘价。美国市场的数据集覆盖了 2020 年的熊市期间，这允许在具有挑战性的市场条件下对框架的有效

中国知网 https://www.cnki.net

表 4-2 数据情况统计。

数据集	资产数量	训练时间段	测试时间段
美国市场	80	2009-01 至 2020-06	2020-07 至 2022-12
香港市场	45	2009-01 至 2020-06	2020-07 至 2021-07
中国市场	34	2009-01 至 2020-06	2020-07 至 2021-07

性进行测试。股票数据来自雅虎财经①。数据的统计情况在表 4-2中总结。

2. 对比模型

在基准模型选择上，一个关键条件是评估对比模型的官方或开源库的可用性。通过采用各相关论文中报告的确切超参数，本文确保了比较的公正性。接着，本文的方法将与两种类别的方法进行比较。(1) 传统方法：OLMAR [92]，CRP [107]，ONS [105]，UP [109] 以及 EG [103]。以上模型是基于开源项目②得到的运行结果。(2) 基于深度强化学习的方法：EIIE [18]，A2C [129]，DDPG [131]，PPO [132] 以及 MetaTrader [19]。其中，EIIE 和 MetaTrader 属于集成算法类别。这些模型中的大多数是基于开源库 FinRL [116]。DeepTrader [17] 和 SARL [94] 模型来自一个开源交易库 TraderMaster③。为了确保对比模型的可靠性和最佳性能，本章采用了与官方一致的超参数。

3. 评测指标

本章借鉴其他研究 [83] 通过三个指标来评估模型性能。第一个指标是累计回报率（Accumulative Return, AR）：

$$AR = \prod_{t=1}^{T}\left(\frac{P_{t+1,close} - P_{t,close}}{P_{t,close}}\right) \cdot a_t. \tag{4-15}$$

第二个指标是夏普比率（Sharp Ratio, SR）：$SR = \frac{\mathbb{E}(r_t)}{\sigma(r_t)}$，其中 r_t 是每日回报率，$\mathbb{E}(r_t), \sigma(r_t)$ 分别代表 r_t 的期望和标准差。SR 弥补了 AR 未考虑风险因素的不足。然而，AR 和 SR 忽略了价格暴跌的风险，这在投资过程中也是至关重要的。第三个指标是卡玛比率（Calmar Ratio, CR），用以评估下行偏差的影响。$CR = \frac{S_n}{MDD}$，其中 S_n 是年化回报率，MDD 表示从峰值（在时间 t）到谷底（在时间 ε）的最大损失，并通过 $MDD = max_{t:\varepsilon>t}\frac{S_t - S_\varepsilon}{S_t}$ 计算得出。

4. 参数设置

本文采用 PyTorch 来实现框架。在训练过程中，选用了 Adam 优化器，并使用单个 NVIDIA RTX 3090 GPU 进行训练。对于所有三个数据集，训练步骤均设为

① https://finance.yahoo.com/
② https://github.com/ZhengyaoJiang/PGPortfolio
③ https://github.com/TradeMaster-NTU/TradeMaster

中国知网 https://www.cnki.net

表 4-3 所有模型的实验效果对比。

类型	模型	美国市场			香港市场			中国市场		
		AR↑	SR↑	CR↑	AR↑	SR↑	CR↑	AR↑	SR↑	CR↑
	Market	0.12	0.34	0.17	0.15	1.01	1.56	0.18	1.16	2.27
传统方法	OLMAR [92]	-0.52	-1.17	-0.65	-0.65	-2.19	-0.95	-0.57	-0.15	-0.83
	CRP [107]	0.16	0.05	0.21	0.20	1.16	1.60	0.32	1.35	1.77
	ONS [105]	0.15	0.28	0.11	0.21	1.15	1.44	-0.02	-0.07	-0.17
	UP [109]	0.18	0.25	0.13	0.20	1.17	1.62	0.27	1.30	1.98
	EG [103]	0.17	0.14	0.27	0.19	1.16	1.60	0.27	1.29	1.96
基于强化的方法	A2C [129]	0.18	0.15	0.24	0.17	1.22	1.51	0.28	1.38	2.16
	DDPG [131]	0.19	0.33	0.32	0.20	1.31	2.00	0.21	1.13	1.61
	PPO [132]	0.20	0.38	0.30	0.18	1.22	1.85	0.30	1.43	2.34
	SARL [94]	0.22	0.41	0.38	0.18	1.27	1.87	0.28	1.40	2.35
	DeepTrader [17]	0.26	0.47	0.34	0.20	1.30	1.97	0.31	1.44	2.38
	Ours	**0.31**	**0.60**	**0.45**	**0.22**	**1.43**	**2.33**	**0.33**	**1.56**	**2.70**
	p 值	2.0e-3	3.0e-3	1.0e-3	1.5e-3	1.5e-3	6.0e-3	8.0e-3	6.0e-3	2.0e-3

表 4-4 不同集成模型的效果对比。

模型	美国市场			香港市场			中国市场		
	AR↑	SR↑	CR↑	AR↑	SR↑	CR↑	AR↑	SR↑	CR↑
EIIE [18]	0.19	1.13	0.31	1.05	1.22	1.51	0.38	1.57	1.77
MetaTrader [19]	0.25	1.31	1.29	0.44	1.38	1.60	**0.43**	1.73	1.92
Ours	**0.38**	**1.65**	**1.45**	**0.47**	**1.46**	**1.81**	0.41	**1.86**	**2.01**
p 值	5.0e-3	8.0e-3	2.0e-3	1.0e-3	6.0e-3	8.0e-3	2.6e-3	1.3e-3	3.0e-3

50000 步，其中批量大小为 128。模型的学习率设置为 10^{-4}，并将 ℓ_2 正则化的权重衰减设置为 10^{-7}。时间窗口大小设置为 252，而 n 步回报的长度设置为 10。此外，所有策略的投资组合向量通过平均分配进行初始化。所有的结果均是随机初始化种子运行 5 次的平均值。

4.5.2 实验结果

本章提出的框架与其他对比模型的性能比较在表 4-3中展示，其中性能最佳的模型以粗体表示，最强基线模型则用下划线突出显示。根据表 4-3，可以得出以下观察结果：(1) 总体而言，与传统方法和基于深度强化学习（RL）的方法相比，本章提出的框架在所有数据集上都显示出了优越的性能。具体来说，在美国市场数据集上，与 ONS 和 CRP 相比，本章提出的框架在 SR（夏普比率）和 CR（累积回报）方面分别提高了 51%和 87%。此外，在美国市场的 SR 指标上本章所提出的框架也超过了所有基于强化学习的模型，并比最先进的模型 DeepTrader 提高了 27%。这些结果表明，通过捕捉和利用与不同投资视野相关的独特模式和洞察力，本章提出的框架在资产配置中实现了更好的性能。这凸显了本章所提的解决方案在应

中国知网 https://www.cnki.net

对金融市场的非稳定性和利用多个投资视野方面的优势。(2) 与基于 RL 的方法相比，传统方法在 SR 和 CR 指标的表现不佳。一些模型甚至变成了投资亏损，例如 OLMAR。原因是这些模型严重依赖手工特征，导致在多变和复杂的金融环境中泛化能力较差。相比之下，RL 是更适合这种多变环境的解决方案，因为 RL 可以通过与金融环境的试错交互直接输出交易决策，以最大化投资回报。这一发现强化了在资产配置任务中应用强化学习的有效性和重要性。(3) 本节还将提出的方法与两种集成算法 EIIE 和 MetaTrader 进行了对比。鉴于 MetaTrader 的代码未公开，本研究采用了与 MetaTrader 相同的数据集进行实验，以便实现公平的比较。实验结果在表 4-4中展示。相比之下，其他两种集成策略仅是结合不同的交易模型，而本章的模型考虑了多种投资视野，通过融合各种视野的交易策略，展现出更加显著的优势。

此外，本节实验还进行了配对 Wilcoxon 符号秩测试，以验证所提方法在统计上显著优于最强基线的结果。统计显著性分析基于本章提出的模型和最强基线的 10 对实验结果，每个模型使用相同的参数设置运行 10 次。在许多研究中 [89,90]（包括本章的工作）通常采用 0.05 的置信水平。在表 4-3中可以观察到所有的 p 值都小于 0.05，证实了实验结果具有统计学意义，并且本章提出的模型性能超越了对比方法。

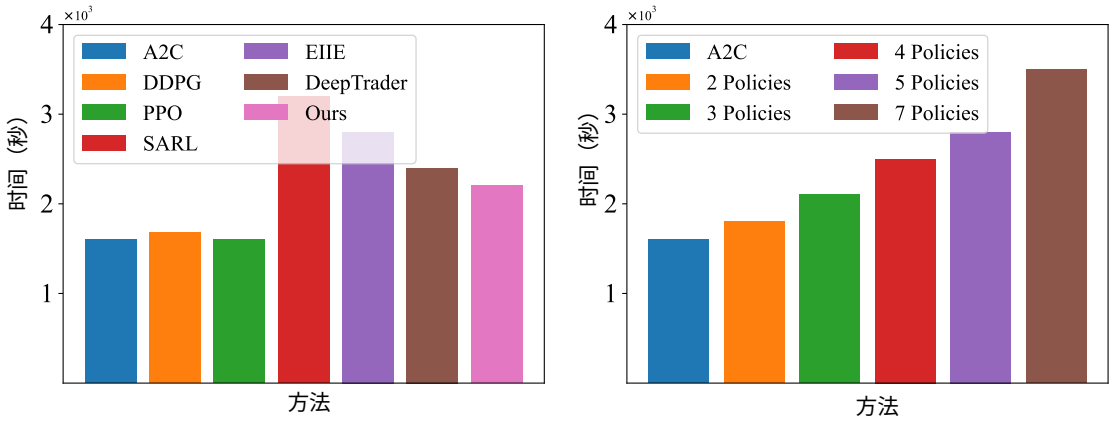
1. 运行时间比较

本节展示了所提出的模型与其他对比模型在时间消耗上的比较。图4-4展示了基于强化学习方法训练 50,000 个样本所需的时间。传统方法相对简单，所需时间较少。因此，本节在此仅将提出的方法与其他基于强化学习的对比模型进行比较。从图4-4(a) 中可以观察到，像 A2C、DDPG 和 PPO 这样的简单方法，虽然复杂度较低，但如表4-3所示，表现并不理想。另一方面，如 EIIE、SARL 和 DeepTrader 这样的最新对比模型在效率和效果上都没有超越本章的模型。这种低效率来源于它们的架构中使用了 LSTM 或 GCN 模块。此外，本节还分析了不同数量子策略的时间消耗情况，如图4-4(b) 所示。显然，随着子策略数量的增加，时间消耗也上升。因此，需要在确定策略数量时找到一个平衡点，避免因过多的策略而导致的不必要的复杂性。

2. 可视化结果

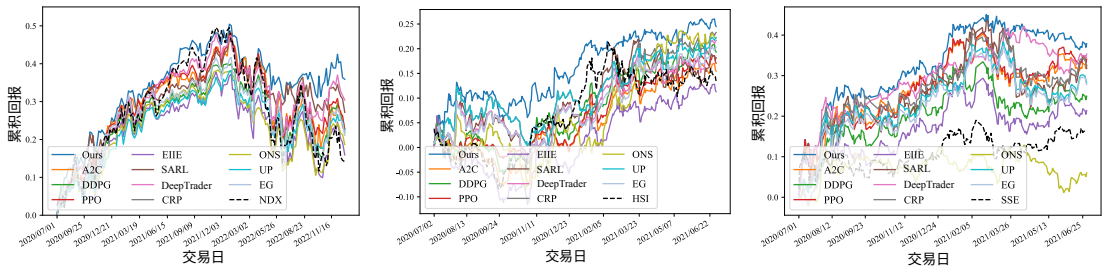
图 4-5 展示了在三个真实世界股票市场中，经过回测的各个模型的累计收益情况（由于 OLMAR 表现不佳被排除在外）。为了验证这些模型的有效性，本节还在图 4-5中绘制了这三个市场的指数，即上海证券交易所的上证综合指数（SSE）、香港证券交易所的恒生指数（HSI）以及美国市场的道琼斯工业平均指数（DJIA）。

中国知网 https://www.cnki.net



(a) (b)

图 4-4 不同方法的时间消耗对比以及不同数量的子策略的时间消耗对比。



(a) (b) (c)

图 4-5 测试阶段三个数据集的累积回报。

从图 4-5中可以明显看出，无论是牛市还是熊市，本章所提的框架始终表现优于市场表现（指数表现）和其他模型。特别值得注意的是，在美国市场，测试期间在 2022 年之后进入了熊市。尽管市场条件具有挑战性，多视野融合的交易策略不仅优于市场表现，而且与其他模型相比也产生了显著的优势。这些结果展示了本章所提方法在不同市场条件下的稳健性。无论市场是处于牛市还是熊市趋势，融合多视野的交易策略始终能够实现卓越的表现。

4.5.3 模型消融实验

1. 子策略的数量

本小节旨在解决以下问题：(1) 使用多种不同子策略是否能够提升资产管理的性能？(2) 要实现最佳性能，多少个子策略才是最优数量？

为了回答问题（1），本节丢弃了子策略，仅保留了以所有资产的价格序列 $\mathcal{P}_t$ 为输入、通过 Q 值优化的跨视野策略，使本章所提出的方法简化为传统的 A2C 算法。A2C 的实验结果在表 4-5中报告。从表 4-5中可以看到 A2C 的性能显著低于本章所提出的结合多个子策略的方法。这表明，多个子策略所做出的预决策确实有

中国知网 https://www.cnki.net

表 4-5 不同数量的子策略下模型的性能表现。

专属策略数量	美国市场			香港市场			中国市场		
	AR	SR	CR	AR	SR	CR	AR	SR	CR
A2C	0.18	0.15	0.24	0.17	1.22	1.51	0.28	1.38	2.16
2	0.21	0.23	0.29	0.20	1.27	1.73	0.29	1.41	2.40
3	0.25	0.33	0.33	0.20	1.28	1.96	0.30	1.45	2.48
4	0.29	0.45	0.40	0.21	1.30	2.05	0.31	1.47	2.53
5	**0.31**	**0.60**	**0.45**	**0.22**	**1.43**	**2.33**	**0.33**	**1.56**	**2.70**
7	0.23	0.37	0.25	0.14	1.18	1.74	0.25	1.34	2.35

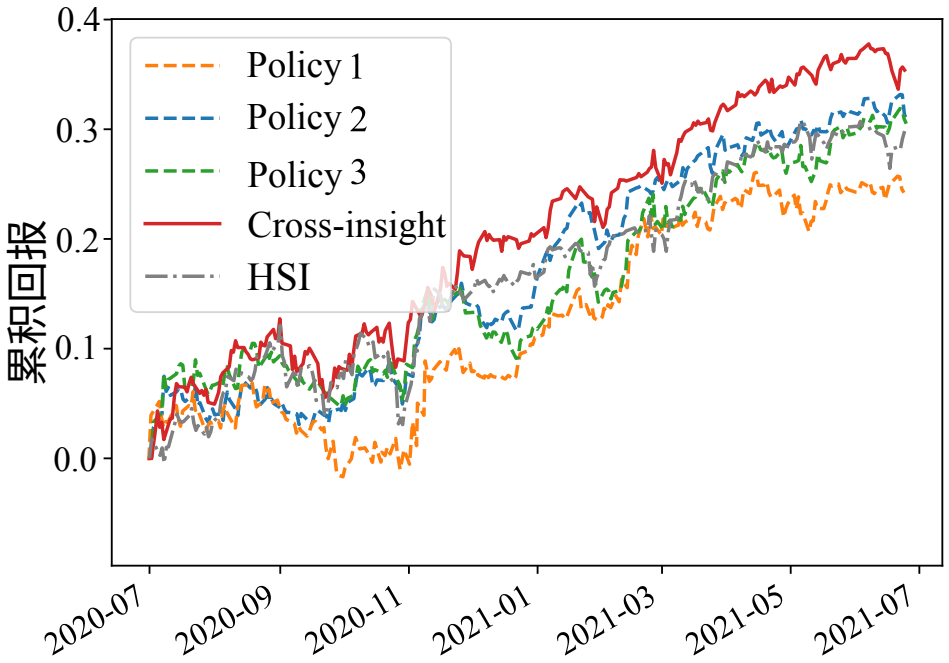


图 4-6 分解粒度为 3 时不同策略的累计收益率。

效地提升了资产组合管理的性能。因此，与仅依赖单一的视野策略相比，结合多个视野的策略能够增强整体表现。此外，图 4-6可视化了每个子策略在香港市场数据集的累计回报，其中策略 1、2 和 3 分别负责短期、中期和长期的投资视野。如图所示，多视野融合策略的表现显著优于市场指数（恒生指数）和所有其他单独策略。然而，专注于短期视野的策略表现不如其他策略。这种差异可以归因于短期预测可用的信息量有限，使得实现投资盈利变得困难。

关于问题（2），本节部署了不同数量的子策略，并评估了它们各自的表现。相应的结果展示在表 4-5中。从表中可以观察到，随着子策略数量增加到 5 个（提高了分解的粒度），框架的性能有所提高，而时间消耗也相应增加，如图 4-4(b) 所示。这表明，尽管以增加时间消耗为代价，多个子策略可以带来更好的资产配置

中国知网 https://www.cnki.net

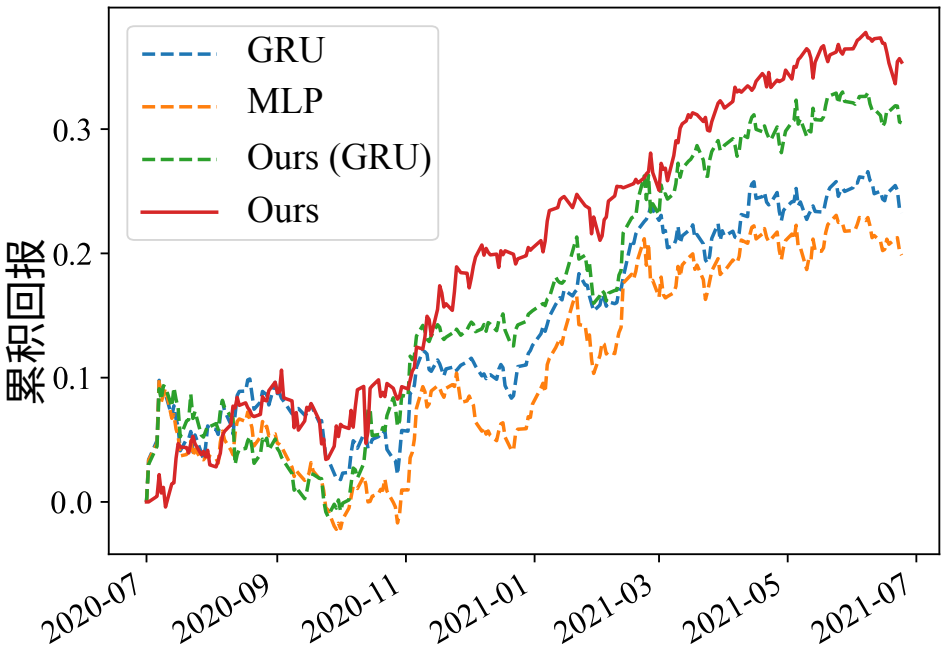


图 4-7 不同神经网络结构的策略所带来的累计收益率。

结果。通过利用具有不同投资视野的多个子策略对市场条件的观察，有效地提升了性能。然而，当策略数量增加到 7 个时，性能反而下降，并且时间消耗继续增加。当存在太多的投资视野时，每个视野中包含的信息变得过于稀薄，基于这样稀薄的信息做出有效的投资决策变得很难。因此，子策略数量不应超过 5 个，因为这个数量在时间消耗和性能之间取得了比较好的权衡。超过 5 个后，模型消耗更多的时间却得到更差的性能。

2. 网络结构分析

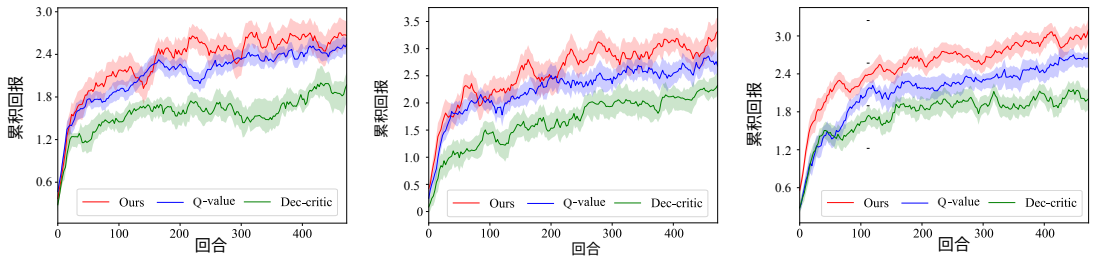
为了验证本章所设计的神经网络的有效性，本节将提出的方法与三种变体进行了比较。前两种变体分别用多层感知机（MLP）和门控循环单元（GRU）替换了精心设计的神经网络，分别称为 MLP 和 GRU。第三种变体，称为 Ours（GRU），保留了与本章方法相同的网络结构，但是将时序卷积网络（TCN）替换为 GRU 单元。结果展示在图 4-7中。

本章提出的策略融入了空间注意力机制，其表现超越了其余三种变体方法。一方面，这一实证发现确认了在提取时间模式时，有效考虑资产之间的相关性具有极高的重要性。此外，使用 TCN 进行时间模式提取比使用 GRU 单元更为有效。另一方面，那些仅专注于时间模式而不考虑资产相关性的方法（即 GRU 和 MLP）表现较差。这一比较说明了精心设计的神经网络在捕捉时间和空间信息方面具备突出能力，从而在资产配置任务中实现了卓越的表现。

3. 反事实机制分析

中国知网 https://www.cnki.net

为了评估本章提出反事实机制的有效性，本节本章提出将其与两个变体进行了比较。第一个变体是对所有策略使用相同的 Q 值进行优化，而本章提出的反事实机制根据每个子策略对累积收益的贡献分配不同的值。第二个变体，称为 Dec-critic，采用了一个分散式评论家结构，其中每个策略都有自己的评论家，每个评论家只接收相应策略在上一时间步执行的状态和动作作为输入。学习曲线如图 4-8所示，清楚地表明本章提出的反事实机制优于其他优化策略及变体。这一成就可归因于本章所提出的反事实机制能够提供不同优势值用于优化不同的子策略。通过考虑每个策略的贡献，指导子策略更有效地训练，从而在资产配置中提高了性能。另一方面，Dec-critic 变体的性能相比其他变体较差。这是因为每个策略都有自己的评论家并且独立训练，没有协作努力来最大化利润。这个结果说明：使用集中式评论家并结合反事实机制来评估每个子策略的优势值具备更优的性能。总的来说，这些发现突显了反事实机制在指导策略行为和提高框架在资产配置性能方面的有效性。



(a) (b) (c)

图 4-8 反事实策略和其他两种变式在三个数据集上的学习曲线。

4.5.4 模型参数分析

本小节的目标是通过改变不同超参数的值来理解交易策略的表现，包括训练窗口大小、n 步回报的长度 n^*，以及神经网络输出的隐藏特征维度 f。所有后续的比较实验都遵循相同的规则：当改变任何一个超参数时，他超参数保持不变。

1. 训练窗口大小分析

本方法采用了 30、120、250 和 370 的窗口大小，分别代表 1 个月、6 个月、1 年和 1.5 年的时长。图 4-9(a)-(c) 展示了改变窗口大小如何影响性能。从图中可以观察到随着窗口大小的增加，性能逐渐提高，但超过 250 后，增益变得不那么明显。因此，在使用本方法时建议用 250 的窗口大小来优化性能，并避免不必要的资源消耗。

2. n 步回报长度分析

中国知网 https://www.cnki.net

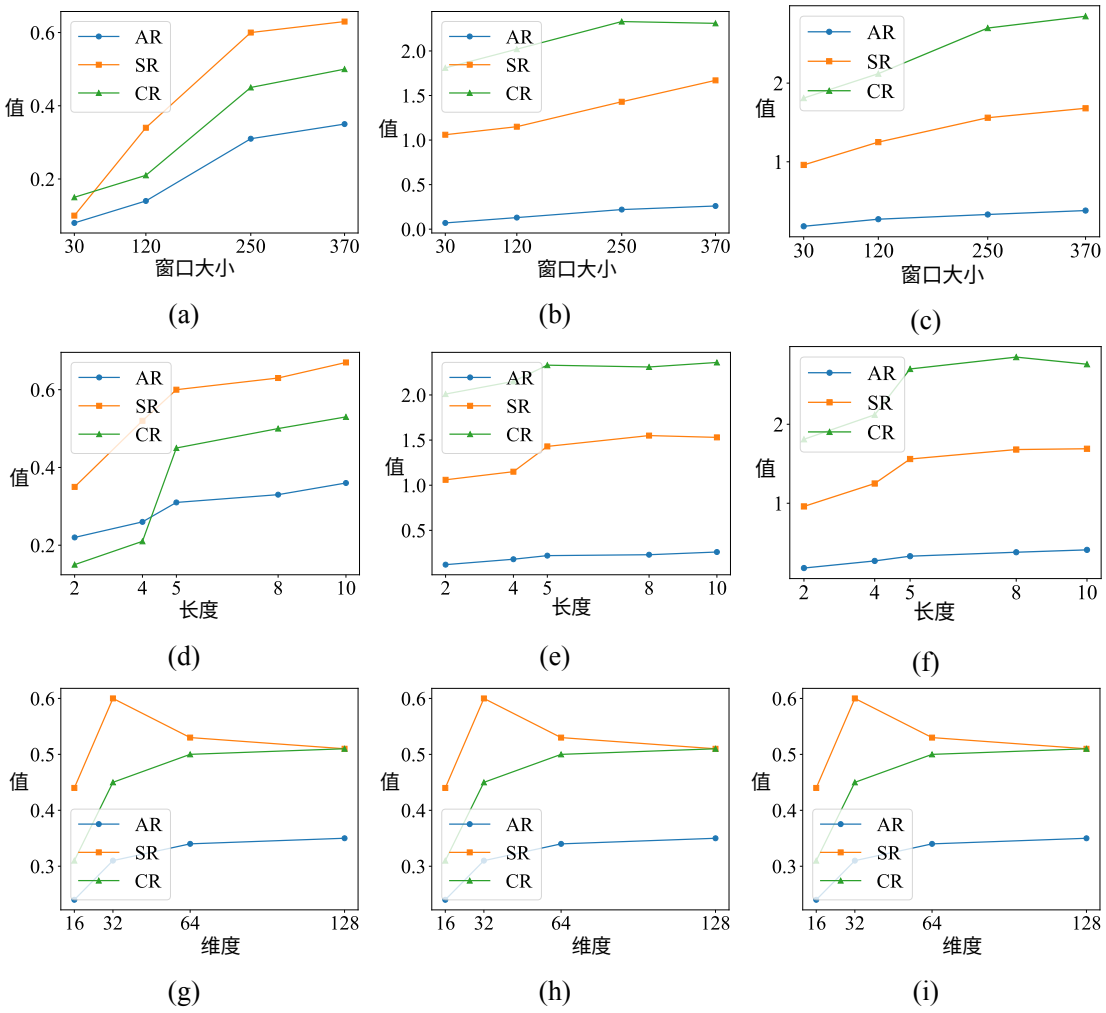


图 4-9 (a)-(c) 为模型在不同的窗口大小时在美国、香港和中国市场下的表现；(d)-(f) 为模型在不同的步回报长度 n 时在美国、香港和中国市场下的表现；(g)-(i) 为模型在不同的嵌入大小时在美国、香港和中国市场下的表现的训练中的性能比较。横轴表示超参数的取值，纵轴表示三个指标的值。

图 4-9(d)-(f) 展示了改变 n 步回报长度 n^* 如何影响性能。随着长度的增加，性能提高。这是因为较大的 n^* 使得评估者可以考虑更长时间范围内的奖励，从而导致状态价值的估计更加准确。然而，这种方法也导致了模型收敛速度变慢。因此，建议选择 $n^* = 5$ 以达到平衡，避免过大的 n^* 阻碍收敛而不带来显著的性能改善。

3. 嵌入维度分析

图 4-9(g)-(i) 展示了改变嵌入维度 f 如何影响性能。特别地，在 32 维时达到最佳性能。超过这一点后，随着维度的增加，性能呈下降趋势。这是因为过大的嵌入维度引入了冗余信息，对模型的性能产生了负面影响。

中国知网 https://www.cnki.net

4.5.5 案例分析

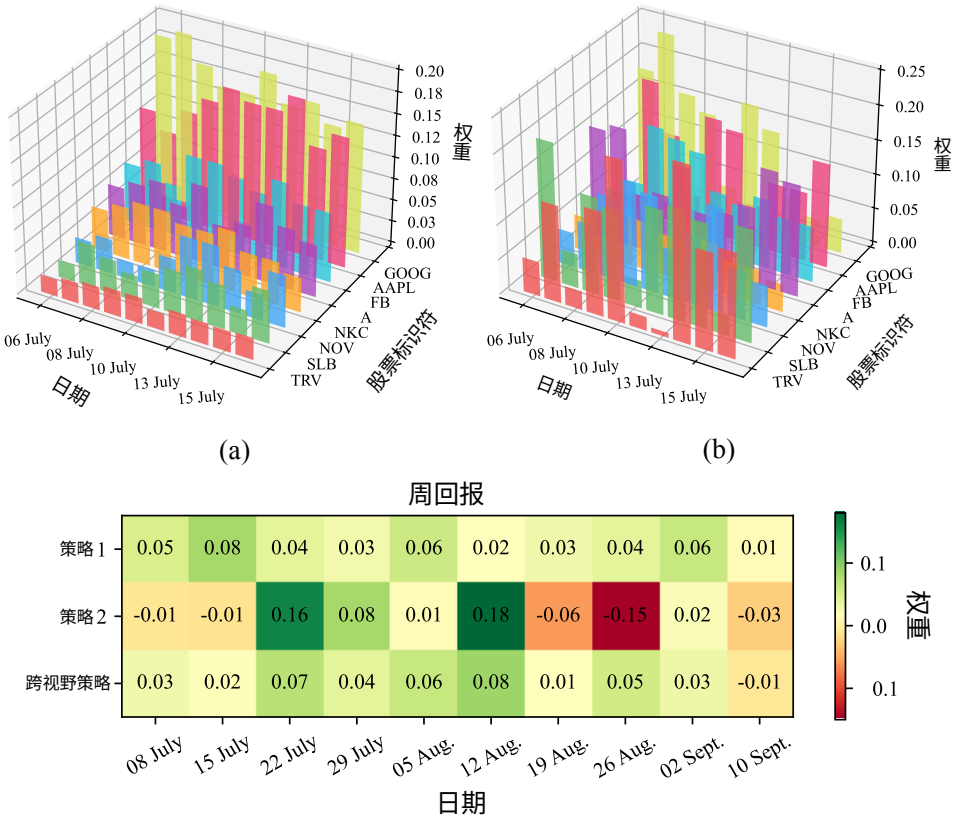


图 4-10 模型性能可视化。股票轴代表公司的符号，而权重轴代表分配给股票的资本权重。策略 1 和策略 2 分别对应于长期和短期视野策略。(a) 和 (b) 分别为两个策略在不同日期下的权重分配可视化；(c) 为两个策略以及跨视野策略在周回报率可视化。

本节通过可视化两个视野专属子策略（长期视野和短期视野）及其回报来定性分析本章提出的交易策略的表现。从图 4-10(a) 中可以观察到长期视野策略的决策相对稳定，每只股票的投资权重没有发生剧烈变化。这种策略决策的稳定性归因于其长期关视野。这意味着它保留了在长期内被认为有利的股票，避免了波动性投资决策。另一方面，如图 4-10(b) 所示，短期视野策略的决策表现出更显著的变化。随着时间步增加，每只股票的投资决策发生了实质性的变化。这是因为短期视野策略更旨在最大化短期回报而只关注短期股票趋势，导致更加波动的投资决策。图 4-10(c) 中展示了两个子策略和跨视野策略的回报。显然，长期策略的回报相对稳定，而短期视野策略的回报相对波动。跨视野策略整合了长期和短期视野策略的决策并考虑当前市场状况，展示了一个更稳定和有效的决策过程，导致

中国知网 https://www.cnki.net

了更有利的结果。

4.6 本章小结

本章提出了多视野融合的交易策略，一种两步骤的基于强化学习的资产配置方法，旨在通过整合具有不同投资视野的多个子策略来适应市场条件的变化。多视野融合策略在第一步中利用定制的视野信息学习的子策略。在第二步中，该方法学习一个跨视野策略，该策略考虑了子策略所做的投资预决策来做出最终的交易决定。跨视野策略基于 Q 值进行优化，而子策略则使用从反事实机制中产生的优势值进行优化，该机制衡量每个策略对回报的贡献。在三个市场上的实验结果显示，与现有的基准相比，多视野融合交易策略表现出更优越的性能。

进一步的分析验证了将不同投资视野纳入决策过程的有效性。然而，本文提出所提出方法的潜在局限性包括训练多个子策略以及同时优化跨视野策略的计算复杂性较高。未来的工作可以探索更高效的训练策略或利用先进的计算技术来解决这些局限性，并在现实世界的金融环境中测试本文提出的方法。

中国知网 https://www.cnki.net

第五章 被动投资策略：基于强化学习的资产选择

5.1 引言

主动投资和被动投资是基金管理中的两种基本策略。在主动投资中，基金经理旨在通过利用他们的专业知识和预测算法挑选有前景的资产来超越市场，力求获得更高的收益。相反，被动投资基于一个信念，即长期来看市场是无法被战胜的。它较少依赖基金经理对资产价值的评判，更多地基于一套严格的标准，例如市场或行业表现。被动投资的一种流行形式是指数投资，投资者寻求构建一个投资组合来复制（追踪）市场指数，并获得与市场一致的回报，这也被称为指数追踪[133]。

指数追踪可以分为两类：完全复制和部分（复制）指数追踪，这取决于构建追踪投资组合的方法。完全复制是最直观的方法，涉及持有构成指数的所有股票。然而在实践中，完全复制可能不是一个理想的策略，因为持有大量指数成分股会导致高昂的交易和管理费用[133,134]。相反，部分复制，顾名思义，只投资于少量的指数成分股，这种只能挑选少量成份股用于投资的操作也被称为“基数约束”。尽管部分复制会导致更大的追踪误差，但由于涉及的资产数量较少，交易成本也会相对较低。由于这一低交易成本的优势，部分复制方法近几年已经引起了学术界和金融行业的更多关注[23,135]。

部分指数追踪（Partial Index Tracking，PIT）需要定期执行资产选择和资产配置两个任务，以追踪特定市场指数的表现。资产选择决定了哪些资产应该被选入追踪组合，而资产配置则决定了如何最优地分配资本权重给这些选定资产。由于资产配置通常被构建为一个回归问题并通过二次规划（Quadratic Programming，QP）进行优化，现有研究的主要区别在于资产选择所采用的技术。资产选择中的基数约束（即从 N 个成份股中选择 K 个资产，$K < N$）使得该任务变成组合优化问题从而难以求解。当前在 PIT 中用于资产选择的方法通常基于启发式算法（Heuristic Algorithms），例如遗传算法[20,21]和模拟退火算法[22]是比较流行的用于寻找最优资产组合的算法。此外，基于神经网络的方法[23,136]近年来也日益受到欢迎。然而，这些方法通常以最小化追踪误差为优先目标，要么完全忽略交易成本，要么通过添加静态约束项使优化过程变得更加复杂。此外，随着成份股数量的增加，启发式算法面临更大规模的成份股指数时，计算可行解的时间变长甚至难以求解。

为了应对这些挑战，本章提出将深度强化学习应用于部分指数追踪（Reinforcement Learning for Partial Index Tracking, RLIT）的方法。RLIT 将资产选择问题

中国知网 https://www.cnki.net

构建为一个马尔可夫决策过程（Markov Decision Process, MDP），并通过深度强化学习（Deep Reinforcement Learning, DRL）进行优化，这一思路具有三个独特的优势可以解决以往方法遇到的挑战：（1）DRL 通过与环境的试错交互进行决策制定，以最大化预期奖励，这与指数追踪中的周期性决策需求高度契合。（2）DRL 能够在奖励函数中灵活整合实际需求，旨在权衡指数追踪误差和交易成本之间的平衡，而不是依赖于静态的约束项。（3）在 DRL 中引入深度神经网络，使模型能够基于学习经验迅速生成解决方案。然而，MDP 的建模提出了两个技术挑战：第一个挑战是设计一个能够处理资产选择中基数约束的策略。第二个挑战是设计一个奖励函数，不仅基于准确的指数追踪，还要有效控制交易成本。

为了解决这两个技术挑战，本章提出了关联性感知深度 Q 网络（Relevance-aware Deep Q-Network, RA-DQN）。该网络同时将资产之间的关系以及资产与指数的关系这两种的领域知识整合到资产选择中。此外，本章还引入了一个成本感知奖励函数来评估选择，引导模型做出既能最小化追踪误差又能减少交易成本的决策。本章的技术贡献有四个方面：

- 本研究探讨了深度强化学习在部分指数追踪上的应用，并提出了一个基于 DRL 的框架来解决部分指数追踪问题。
- 本章设计了一个关联性感知深度 Q 网络（RA-DQN），旨在克服资产选择中的基数约束问题，并在做出选择时综合考虑资产内部和指数间的关系。
- 本章介绍了一个成本感知奖励函数，以指导更精准的资产选择，实现最小的追踪误差和交易成本。
- 通过在 6 个基准数据集上的广泛实验展示了本章所提框架的有效性。此外，通过追踪超过 10 年的 S&P 500 指数，本研究证实了方法的实用性。

5.2 相关工作

本节将介绍指数追踪的相关研究和强化学习在金融领域的相关应用两个部分。

5.2.1 指数追踪

部分指数追踪（PIT）可以分解为两个子任务：资产选择和资产配置[21]，其中第一个子任务负责选择资产，第二个子问题完成在选定资产间分投资配资金。由于第二阶段通常被构建为一个通过二次规划（QP）优化的回归问题，当前研究的主要差异在于第一阶段用于资产选择的技术。资产的选择通常基于领域知识，如资产间关系（例如，协整关系[31]和收益相似性[137]）或指数与资产间的关系（例如，复制因子[138]）。各种启发式算法，如遗传算法[20,21]和模拟退火算法[22]，也

中国知网 https://www.cnki.net

相继被用于搜索最佳资产组合。正向选择和反向选择[23]是两种在商业实践中广泛使用的方法，其思路是将 PIT 建模为带约束的二次回归问题并使用 QP 求解，然后迭代地选择/排除权重最大/最小的资产。近期，各种基于深度神经网络（Neural Network, NN）的方法被提出用于资产选择。例如，Zheng 等人[23]提出了一种神经网络的重参数化方法，而 Zhang 等人[139]利用图卷积网络进行资产选择，并将其与混合整数规划结合用于资产配置。Kwak 等人[136]引入了固定噪声的深度神经网络（Deep Neural Network with Fixed Noise, NNF）模型，用于全部和部分复制指数追踪。然而，所有这些模型在追踪过程中都忽略了交易成本，这使得它们的实用性受到了限制。此外，随着成份股数量的增加，启发式算法面临可扩展性问题，导致寻找可行解的计算时间变长。

另一组方法通过添加稀疏性约束，如 ℓ_0 [133,140]、ℓ_2 [141] 或 ℓ_p（其中 $0 < p < 1$）[142]，共同优化了这两个子问题。ℓ_0 范数是构建稀疏追踪投资组合时最受欢迎的稀疏性约束。然而，施加 ℓ_0 约束使得回归问题变成 NP 难问题，难以求解[133]。一些工作尝试通过引入多步加权最小绝对收缩与选择算子（Multi-step Weighted Least Absolute Shrinkage and Selection Operator, MSW-Lasso）[143] 或线性近似（Linear Approximation Index Tracking, LAIT）[133] 来简化稀疏指数追踪。另一种选择是将优化问题转化为混合整数线性规划问题（Mixed-integer Linear Programming, MILP）[144,145]，但这需要使用商业优化求解器（例如，CPLEX、MOSEK 和 GUROBI），并且需要较长的计算时间才能得到解。此外，考虑交易成本约束使得优化问题更加难以求解[146,147]。

5.2.2 强化学习在金融领域的应用

本节主要介绍深度强化学习（DRL）在金融交易中的应用。

深度强化学习在金融市场的应用包括投资组合优化[18]、股票交易[64,148] 以及加密货币管理[83,95]。相同独立评估系统集合（Ensemble of Identical Independent Evaluations System, EIIE）[18] 通过使用三个相同的深度神经网络来实现特征提取。Jeong 和 Kim[148] 设计了一个自动化系统，结合了强化学习和深度神经网络回归器来预测股票的数量。Deng 等人[64] 使用深度神经网络表示金融信号，并用 RL 与这些信号表示进行交互并做出交易决策以最大化累积奖励。Xu 等人[83] 和 Zhang 等人[95] 在建模加密货币市场时都考虑了加密货币之间的关系。

上述所有研究仅关注深度强化学习在交易中的应用。然而，强化学习在指数追踪上的应用尚未被探索。从本质上讲，指数追踪是一个周期性的决策过程，非常适合应用强化学习算法。鉴于近年来深度强化学习取得的显著成就，探索其在

中国知网 https://www.cnki.net

指数追踪中的应用具有重要意义。

5.3 预备知识

部分指数追踪过程可以在图 5-1(a) 中展示。指数化投资的交易过程存在两种类型的周期：持股期和交易期。在交易期 t_1，所有成分股及指数的历史数据会作为状态 s 输入到框架中。数据首先经过第一阶段决定哪些资产被选为投资组合用以追踪指数，决策 a 将被输出到第二阶段。在第二阶段，问题转化为一个二次问题，使用二次规划计算预选资产之间的最优配置权重 w。这里，$a \in [0,1]$ 是一个长度为 N 的离散二元向量，其中 1 表示该资产被选入追踪投资组合，0 则相反。权重 $w \in \mathbb{R}^N$ 是实数向量，其中 $w_i \neq 0$ 表示资产 i 属于追踪投资组合，w_i 的值是其资本配置权重。在持股期间，管理者持有选好的资产，不进行任何买卖。由于市场波动，持股期间的资产价格和投资回报会随着时间变化，部分指数追踪的目的是所选资产的投资回报率与追踪的指数的回报率紧密对齐。在下一个交易期 t_2，将决定新的追踪资产及其权重。持股期和交易期的时间长度可以根据需要进行设置。在部分指数追踪的任务中有以下常用的定义：

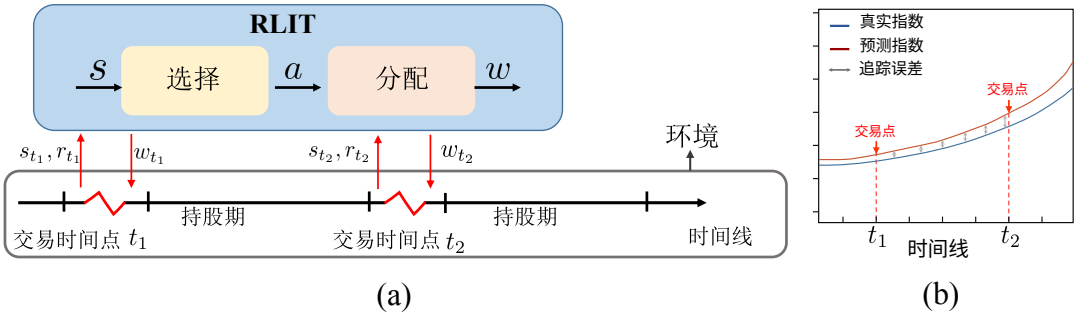


图 5-1 模型框架概述图。左边是模型的数据流程，右边是指数追踪的示例。左边的两个交易时间 t_1 和 t_2 对应右边的两个交易箭头 t_1 和 t_2。

定义 5.1 (约束条件) 二元向量 a 遵循基数约束，表示为 $\| a \|_0 = K$ 或 $\sum_{i=0}^{N} a_i = K$。这个约束意味着 a 中只有 K 个元素被设为 1，表示从 N 个成分中选择 K 个资产。为了控制交易成本，可能会使用换手率约束，如最常见的格式 $\| a_{t+1} - a_t \|_1 \leq C$，其中 C 是两个连续时期 $t+1$ 和 t 之间投资组合总变化的上限。换手率约束增加了解决方案的复杂性。对于 w，它满足非负性要求，即 $w_i \geq 0, \forall i \in \{1, 2, \cdots, N\}$ 与和为一约束 $\sum_{i=0}^{N} w_i = 1$，分别表示只做多策略和资本始终被充分利用。

定义 5.2 (投资组合回报) 假设持股期内总共有 T 个时间步骤。在时间步骤

中国知网 https://www.cnki.net

$t \in \{1, 2, ..., T\}$ 的投资组合回报等于该时间步骤所有选定资产的回报之和:

$$y'_t = x_t \cdot w_t^\top = \sum_{i=0}^{K} w_{i,t} \cdot x_{i,t}, \tag{5-1}$$

其中 $x_t \in \mathbb{R}^{1\times K}$ 是时间步骤 t 所有 K 个资产的回报率。因此，在持股期结束时可以获得 T 个回报值，即 $Y' = [y'_1, y'_2, ..., y'_T]$。

定义 5.3 **(追踪误差)** 追踪误差用于衡量预测的投资组合回报与追踪指数的实际回报之间的差异，如图 5-1(b) 所示。欧几里得距离是比较常用的用于衡量追踪误差的指标[23,133,149]。数学上欧几里得距离可以表示为 $\epsilon = \frac{1}{T} \parallel Y - Y' \parallel_2^2$，其中 $Y = [y_1, y_2, ..., y_T]$ 是持有期间指数的回报。

5.4 模型: 基于强化学习的指数追踪模型

5.4.1 资产选择

本研究将资产选择建模为一个马尔可夫决策过程，该过程由四元组定义：MDP = (S, A, P, R)，其中 S 是状态空间，A 是动作空间，P 是转移概率函数，R 是奖励函数。在 MDP 中，智能体在时间步 t 接收到状态 $s_t \in S$，然后根据智能体的策略 π_θ 做出响应 $a_t \in A$，策略由 θ 参数化。上述描述可以表示为 $a_t = \pi_\theta(s_t)$。环境根据智能体的动作 a_t 提供奖励，智能体随后根据接收到的奖励 r_t 在下一个周期调整其动作。智能体的最终目标是最大化其预期的未来奖励。在部分指数追踪（PIT）的背景下，MDP 可以定义为：

- 状态 s：描述 N 个成分股和指数当前状态的一组历史特征。通常，可以使用不同类型的信息作为当前状态，如交易量、历史价格和情绪评分。这项工作使用前 L 天的成分股和指数回报率序列作为当前状态，以保持与其他对比模型的一致性。
- 动作 a：动作由二元向量 $a_t = [a_{t,1}, ..., a_{t,N}] \in A$ 指定，且 $\parallel a_t \parallel_0 = K$，其中 $a_{t,i} = 1$ 表示资产 i 在时间步 t 被选入追踪投资组合，反之 $a_{t,i} = 0$。A 是包含 C_N^K 元素的组合集合，其中每个元素代表从 N 个成分中选择 K 个资产的可能组合，也被称为组合动作空间。
- 奖励 r：本章定义了可以考虑交易成本的奖励函数——成本感知奖励函数，用于评估动作以权衡交易成本和最小化追踪误差。
- 策略 π：策略是一个指定在当前状态 s 作为输入时采取哪些动作的状态动作转换函数。与状态-动作值相关的最优策略可以通过 Q 学习算法[123]优化：$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta_t \cdot (r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, A) - Q(s_t, a_t))$，其中 η_t 是学习

中国知网 https://www.cnki.net

率，γ 是折扣因子。状态-动作值函数 Q 是关键，因为准确估计动作-状态值与 Q 学习算法的有效性相关。因此，下一节会详细论述 Q 网络。

1. 关联性感知深度 Q 网络

本节介绍关联性感知深度 Q 网络（RA-DQN）如何解决资产选择。RA-DQN 的设计侧重于解决两个关键方面：（1）在选择资产时有效处理基数约束；（2）利用金融领域知识。金融领域知识包括个别资产与指数之间的关系以及不同资产之间的相互关系，这些通常也是工业领域所考虑的因素[31,135,138]。这要求 RA-DQN 为与指数关系密切的资产分配更高的 Q 值，而具有密切相互关系的资产保持相似的 Q 值。下面将详细说明如何在 RA-DQN 的设计中考虑这些因素。

DQN 通过最大化 Q 值来选择动作，表示为 $a_t = \text{argmax}_A Q(s_t, A)$。然而，难点在于 A 是从 N 个成分中抽取的 K 个资产的组合集合，当 N 很大时，A 的可能组合数（即，C_N^K）是巨大的，因此通过遍历所有可能的组合来计算每个组合的 Q 值是难以做到的。受到 DQN 在处理自然语言动作空间[150,151] 的启发，RA-DQN 通过估计资产级别的值而不是组合级别的值来应对组合动作空间的挑战。具体来说，Q 值函数旨在估计每个资产的值 $Q(s_t, i), i \in \{1, .., N\}$，其中 $Q(s_t, i)$ 代表通过选择资产 i 获得的预期累积未来奖励。在选择资产时，贪婪地选择具有最高值 $Q(s_t, i), i \in \{1, .., N\}$ 的前 K 个资产以获得最高的 $Q(s_t, a_t)$，换句话说，获取最大状态-动作值 $Q(s_t, a_t)$ 等同于贪婪地选择前 K 个资产：

$$Q(s_t, a_t) = \sum_{i}^{|a_t|_K} Q(s_t, i), \tag{5-2}$$

其中 $|a_t|_K$ 表示具有最高值 $Q(s_t, i)$ 的前 K 个资产。这种方法使神经网络能被用来估计每支资产的价值，而不是去评估大量的资产组合价值，从而解决了关键方面（1）。

为了将资产内部和指数间关系融入 $Q(s_t, i)$ 的建模中，本研究提出了一种双流 RA-DQN 结构。这种结构包含了一个局部敏感模块和一个指数感知模块分别用来估计资产与资产之间的估计值 Q_i^S 和资产与指数间的值 Q_i^I，即相似的资产就有相近的 Q_i^S，与指数关系大的应具有更大的 Q_i^I。资产 i 的最终 Q 值计算为 Q_i^S 和 Q_i^I 的加权组合，表示为：

$$Q(s_t, i) = (1 - \lambda) Q_i^I + \lambda Q_i^S. \tag{5-3}$$

在双流结构之前，有一个预处理模块将原始回报率数据编码为更稳健的表示。因此，RA-DQN 包含三个组成部分，预处理模块（图 5-2(a)），局部敏感模块（图 5-2(b)）和指数感知模块（图 5-2(c)）。下文中将详细介绍每个模块。

中国知网 https://www.cnki.net

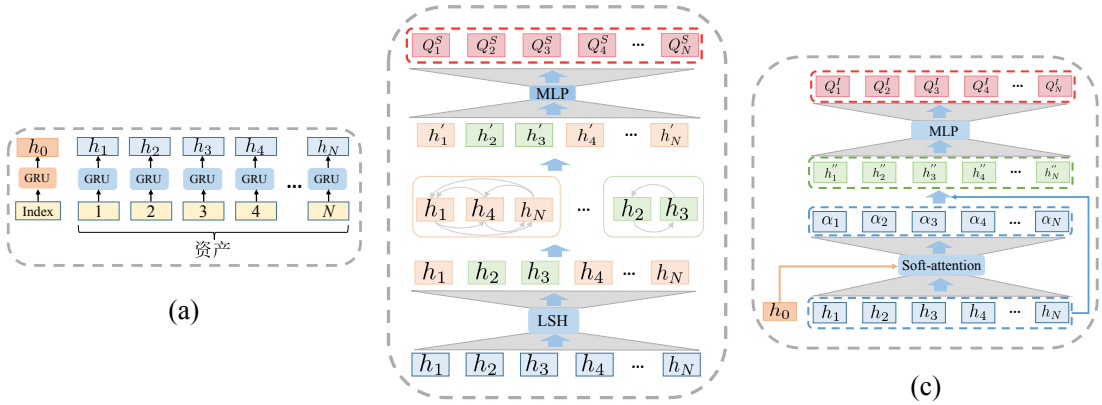


图 5-2 RA-DQN 的三个模块。(a) 为预处理模块；(b) 为局部敏感模块；(c) 为指数感知模块。

（i）预处理模块：预处理模块用于将资产和指数的原始序列数据（即回报率）编码为更稳健的表示。本模块采用门控循环单元（GRU）来编码序列数据。具体来说，指数数据 $p_{index} \in \mathbb{R}^{1\times T}$ 和资产数据 $p_{assets} \in \mathbb{R}^{N\times T}$ 分别被编码成表示 $h_0 \in \mathbb{R}^{1\times d}$ 和 $h_{i\in\{1,..,N\}} \in \mathbb{R}^{N\times d}$，其中 d 是表示的维度。值得注意的是，本模块采用两个不同的 GRU 分别对指数和资产的数据进行编码（在图 5-2(a) 中以不同颜色区分），以生成更具区分性的指数和资产表示。

（ii）局部敏感模块：为了在建模 Q_i^S 时考虑资产内部关系，即表现相似的资产应展现相似的表示并产生接近的 Q 值，本模块引入了局部敏感注意力，该注意力通过整合局部敏感哈希（Locality-sensitive Hash, LSH）和自注意力生成每个资产的表示，如图 5-2(b) 所示。首先，使用 LSH 将资产分组到不同的组中。在每个组内为每个资产计算自注意力。数学上，资产 i 的注意力可以写为：

$$h_i' = \sum_{j\in P_i} softmax(\frac{q_i \cdot k_j}{\sqrt{d_k}})v_j, \tag{5-4}$$

其中 q_i、k_j、v_i 是通过对原始输入 h 进行线性变换获得的查询、键和值向量，P_i 是与资产 i 哈希到同一个桶的资产集合，表示为 $P_i = \{j \mid hash(q_i) = hash(k_j)\}$。本研究采用角度 LSH[152]，这是一种 LSH 变种，具有提简单、高效和易于实现的优点，表示为：

$$hash(x) = argmax[xM; -xM], \tag{5-5}$$

其中 $M \in \mathbb{R}^{d\times \frac{m}{2}}$ 是从正态分布中抽取的随机矩阵，m 是桶的数量。在为每个资产 i 获取局部敏感表示 h_i' 之后，通过将 h_i' 输入多层感知器（Multi-layer Perceptron,

中国知网 https://www.cnki.net

MLP）来计算 Q 值 Q_i^S。

（iii）指数感知模块：该模块旨在建模 Q_i^I 时考虑资产与指数间关系。通过采用软注意力机制来计算指数 h_0 与资产 h_i 之间的系数，如公式 5-6所示，其中 $W_1, W_2 \in \mathbb{R}^{d\times d}, \mu, b \in \mathbb{R}^{d\times 1}$ 是可训练的参数。系数 α_i 代表指数与资产 i 之间的相关性，α_i 越大，他们之间的相关性越强。获得系数 α_i 后直接将其与 h_i 相乘，以获得指数感知表示 h_i''，如公式 5-7所示。最终的 Q 值 Q_i^I 也通过一个 MLP 计算得到（与局部敏感模块不同的一个）。

$$\alpha_i = \mu^T \sigma(W_1 h_0 + W_2 h_i + b), \tag{5-6}$$

$$h_i'' = \alpha_i h_i. \tag{5-7}$$

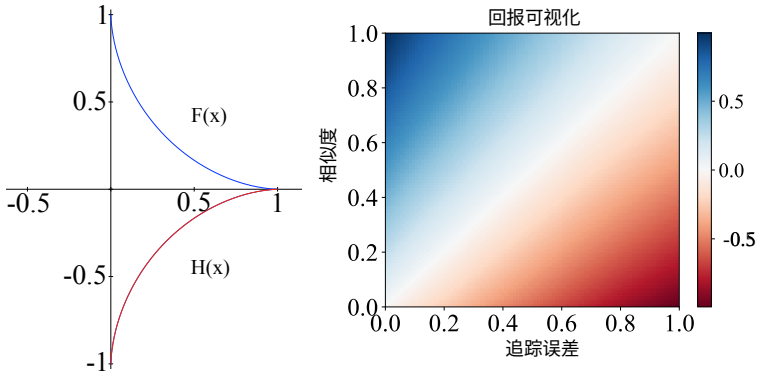
2. 成本感知奖励函数

与之前基于强化学习的交易任务 [17,64,153] 不同，这些任务旨在最大化投资利润，部分指数追踪的目标是紧密追踪选定的指数并解决其他实际需求，即最小化交易成本。因此，直接应用现有的奖励函数不适合这个任务。本研究引入了一个自定义奖励函数，通过映射函数从而同时考虑追踪误差和交易成本的双重目的。

奖励映射是指用一个映射函数重塑原始的评价准则，以更好地评估动作。在实践中，最小化追踪误差很重要，但控制交易成本同样重要，这些都是基金管理者需要重点考虑的因素。盲目追求最小的追踪误差而不考虑交易成本是不切实际的，也不符合现实交易的需求。为了反映动作与追踪误差 ϵ（即更小的误差意味着更好的策略响应，应该通过更大的奖励来鼓励）之间的关系，本研究引入了一个奖励映射函数 $F(x)$（蓝色曲线）如图 5-3(a) 所展示的那样来映射 ϵ。图 5-3(a) 表明，ϵ 的较小值导致更高的奖励 $F(\epsilon)$。本研究采用两个连续动作 a_t 和 a_{t-1} 的杰卡德相似度（记为 j）来作为衡量两个时刻内交易动作的差异并用这个差异值来评估交易成本。更高的相似度意味着更低的成本，因为在两个连续时间步骤之间不同的动作涉及到在不同资产之间切换投资，不可避免地会产生交易成本。为了控制交易动作产生更小的交易成本，本研究采用映射函数 $H(x)$（图 5-3(a) 红色曲线）来映射 j。显然，较小的 j 值将会导致较小的奖励 $H(j)$。$F(x)$ 和 $H(x)$ 都被称为半星型映射函数，它们的数学表示如公式 5-8所示：

$$\begin{cases} x^{\frac{3}{5}} + F(x)^{\frac{3}{5}} = 1 \\ x^{\frac{3}{5}} - H(x)^{\frac{3}{5}} = 1 \end{cases}. \tag{5-8}$$

中国知网 https://www.cnki.net



(a) 半星型映射函数　　(b) 奖励函数可视化

图 5-3 映射函数与奖励函数可视化。

最终，成本感知奖励函数的定义为 $R = F(\epsilon) + H(j)$，如图 5-3(b) 中所示。成本的奖励函数鼓励智能体做出能够导致较小 ϵ 和较大 j（更深蓝色）的动作，并在造成较大 ϵ 和较小 j（更深红色）时对智能体进行惩罚。显然，奖励函数提供了一种更灵活的方法来指导策略选择更合适的资产组合来权衡追踪误差和交易成本。

5.4.2 资产配置

在资产被选定之后，指数追踪可以被定义为二次回归如公式 5-9所示：

$$\min_{w \geq 0, \sum_i w_i = 1} \| X'w - y \|_2^2, \tag{5-9}$$

其中 $X' \in \mathbb{R}^{T \times K}$ 包含了选定资产的回报率。另外两个约束，即非负性与和为一约束，可以通过转换为以下二次规划问题从而高效求解：

$$\begin{aligned} \min_{w} \quad & \tfrac{1}{2} w^\top P w + q^\top w, \\ \text{subject to:} \quad & Gw \leq h \quad \text{and} \quad Aw = b \end{aligned} \tag{5-10}$$

其中 $P = 2(X'^\top X')$, $q = -2X'^\top y$, $G = -I$, $h = 0$, $A = 1^\top$ 和 $b = 1$。更详细的转换过程如下所示：

命题 5.4 (最小二乘回归转化为二次规划) 设 $f(x)$ 表示最小二乘目标函数，$g(x)$ 表示二次规划目标函数。转换的目标是在相同约束条件下，找到一个二次规划函数 $g(x)$，使其与 $f(x)$ 有相同的最优解，表示为 $f(x) \propto g(x)$。

从展开最小二乘函数的目标函数 $f(x)$ 开始：

$$\begin{aligned} \| X'w - y \|_2^2 &= (X'w - y)^\top (X'w - y) \\ &= w^\top X'^\top X w - y^\top X' w - w^\top X'^\top y + y^\top y. \end{aligned}$$

中国知网 https://www.cnki.net

表 5-1 6 个基准数据集情况统计。

指数名字	股票数量 N	训练周数	测试周数	选择数量
Hang Seng	31	146	145	10
DAX 100	85	146	145	10
FTSE 100	89	146	145	10
S&P 100	98	146	145	10
Nikkei 225	225	146	145	10
Russell 2000	1318	146	145	90

注意 $y^\top X'w = w^\top X'^\top y$，舍弃常数项 $y^\top y$，因为它不依赖于优化向量 w，因此不影响问题的最优解。因此，$f(x)$ 可以转化为以下形式的二次问题 $g(x)$：

$$\begin{aligned} \| X'w - y \|_2^2 &\propto w^\top X'^\top X'w - 2y^\top X'w \\ &= \tfrac{1}{2} w^\top P w + q^\top w, \end{aligned}$$

其中 $P = 2(X'^\top X')$，$q = -2X'^\top y$。这是一个凸优化问题，可以由大多数现成的求解器处理。

在完成二次规划（QP）过程后就可以计算出配置权重 w。随后，w 被提供给环境以获得奖励，这作为反馈指导策略 π 的优化。

5.5 实验结果与分析

5.5.1 实验设置

实验部分评估了模型在 6 个基准数据集上的性能。随后，通过消融实验与参数分析，探讨了各个模块与参数的贡献。最终，模型在真实世界数据集——标准普尔 500 指数（S&P 500）上的应用有效性得到了验证。

1. 数据集介绍

六个基准数据集来自于开源库 OR-Library[①]，这六个数据集常用于指数追踪问题。每个数据集对应一个特定的市场指数及其成分股。具体来说，这六个市场指数分别是：Heng Seng 指数（香港）、DAX 100 指数（德国）、FTSE 100 指数（英国）、S&P 100 指数（美国）、Nikkei 225 指数（日本）和 Russell 2000 指数（美国）。对于每个成分股分别包含了 291 个周的价格。关于数据更多的细节和选择的资产数目 K 在表 5-1中进行了总结。

2. 对比模型

本模型与学术界和工业界中使用的七种流行的指数追踪方法进行了全面比较。

① http://people.brunel.ac.uk/mastjjb/jeb/orlib/indtrackinfo.html

中国知网 https://www.cnki.net

前三个方法联合优化资产选择和配置，而后四个方法则依次优化它们。

- MILP [144]：MILP 将 PIT 转换为混合整数线性规划问题，随后使用 Gurobi 商业求解器解决，该求解器对于学术研究目的免费。
- LAIT [133]：LAIT 引入 ℓ_0 正则化以确保投资组合的稀疏性，但采用近似函数来计算 ℓ_0 范数。
- MSW-Lasso [143]：MSW-Lasso 使用多步加权 Lasso 作为稀疏性约束，以提高优化效率。
- Hybrid-GA [20]：该方法利用遗传算法进行资产选择，然后使用 QP 进行配置优化。
- Forward& Backward [23]：这两个方法都是从使用 QP 拟合所有资产开始。Forward 方法迭代选择权重最大的资产，直到达到所需的资产数目 K。另一方面，Backward 方法迭代排除权重最小的资产，直到剩余的资产数等于 K。
- NNF [136]：NNF 采用了一个带有固定噪声的神经网络，用于资产选择和配置，以解决 PIT 问题。

3. 评测指标

模型性能最直观的评估指标是真实指数与复制指数表现之间的测试时的追踪误差（Tracking Error, TE）。除了 TE，本实验还使用另外两个额外的指标：（i）夏普比率（Sharp Ratio, SR）：SR 考虑了投资组合的回报和投资风险；（ii）换手率（Turnover, TO）：换手率可以解释为在 T 个交易周期内，N 个资产交易的平均比率。较低的 TO 意味着较低的交易成本。本研究要求指标 SR 尽可能与指数保持一致，而不是像其他主动投资策略一样主要关注虑高 SR 指标。因此，以下实验报告了两个预测指数与真实指数之间的指标绝对差异，表示为 Δ_{SR}。

4. 评测程序

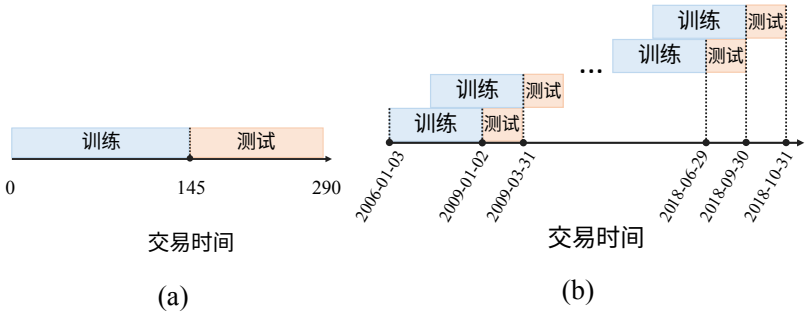


图 5-4 两种不同的数据分割方法。(a) 为 6 个基线数据集的分割方法，(b) 为 S&P 500 数据集的分割方法。

中国知网 https://www.cnki.net

表 5-2 所有模型在六个基线数据集上的表现。‘-’ 表示模型结果无法在规定的时间内被计算出来。

数据集	指标	MILP [144]	LAIT [133]	MSW-Lasso [143]	Hybrid-GA [20]	Forward [23]	Backward [23]	NNF [136]	RLIT	p 值
Hang Seng	TE↓	**7.7e-4**	1.5e-3	4.2e-3	8.1e-4	8.3e-4	1.5e-2	8.2e-4	7.9e-4	6.1e-4
	Δ_{SR}↓	**0.13**	0.72	0.50	0.59	0.67	0.34	1.25	0.19	4.3e-4
	TO↓	0.23	0.25	0.29	0.30	0.28	0.33	0.39	**0.20**	3.1e-4
DAX 100	TE↓	1.3e-3	1.5e-3	2.4e-3	1.5e-3	1.3e-3	3.1e-2	9.8e-4	**9.6e-4**	9.6e-5
	Δ_{SR}↓	1.49	2.51	0.92	0.45	**0.01**	0.53	1.93	0.38	1.6e-4
	TO↓	0.32	0.31	0.35	0.29	0.43	0.31	0.32	**0.25**	7.6e-4
FTSE 100	TE↓	1.1e-3	1.4e-3	2.1e-3	8.6e-4	8.9e-4	2.4e-2	9.0e-4	**8.3e-4**	4.3e-4
	Δ_{SR}↓	0.67	3.44	0.65	1.07	0.54	0.96	1.25	**0.40**	4.8e-4
	TO↓	0.28	0.30	0.25	0.24	0.26	0.27	0.33	**0.22**	1.5e-4
S&P 100	TE↓	1.2e-3	1.5e-3	1.7e-3	9.8e-4	1.1e-3	3.1e-2	9.8e-4	**9.6e-4**	6.1e-4
	Δ_{SR}↓	1.82	3.73	0.81	0.77	0.85	0.84	3.21	**0.70**	9.1e-5
	TO↓	0.37	0.40	0.31	0.32	0.32	0.33	0.30	**0.28**	4.8e-4
Nikkei 225	TE↓	3.3e-3	1.6e-3	1.9e-3	9.1e-4	9.5e-4	4.5e-2	9.2e-4	**8.2e-4**	6.1e-5
	Δ_{SR}↓	1.37	0.24	0.30	0.45	1.41	0.20	1.09	**0.17**	3.1e-4
	TO↓	0.38	0.46	0.34	0.27	0.34	0.32	0.39	**0.23**	4.3e-4
Russell 2000	TE↓	3.9e-3	3.9e-3	6.3e-3	4.5e-3	2.1e-2	-	3.7e-3	**3.1e-3**	7.6e-4
	Δ_{SR}↓	0.48	0.20	2.25	1.30	0.11	-	0.39	**0.09**	6.0e-5
	TO↓	0.48	0.54	0.49	0.44	0.51	-	0.48	**0.41**	3.0e-4

本实验参考其他指数追踪工作 [22,154,155] 的评测标准，采用日调整策略来评估框架的性能。本实验还按照文献 [22,154,155] 的做法，将前 146 周的周价格作为训练集，接下来的 145 周作为测试集来验证模型，如图 5-4(a) 所示。

5. 参数设置

所有实验均在一台配备 32GB 运行内存和一块 NVIDIA RTX 3090 GPU 的机器上进行。为了消除随机性并获得统计上显著的结果，所有实验 10 次并计算了平均性能。本方法使用 PyTorch 实现了所提框架并采用用 Adam 优化器，学习率为 $1.0e^{-4}$。框架的训练步骤对所有数据集均为 50000，批量大小为 32。折扣因子 γ 设为 0.99。QP 问题使用 CVXOPT 1.3[①]解决，这是一款开源的凸优化软件。考虑到每个指数在现实世界中的行业构成，指数 HSI、DAX 100、FTSE100、S&P 100、Nikkei 225 和 Russell 2000 的组数 M 分别为：$\{4, 11, 11, 10, 33, 11, 10\}$。更多的参数详见表 5-3。

5.5.2 实验结果

1. 在基准数据集上的结果分析

性能比较结果如表 5-2所报告，其中最佳性能模型以粗体表示，最强基准以下划线突出显示。虽然 MILP 在应用于恒生指数（成分数最少）时表现良好，但在应用于成分数较多的其他指数时性能下降。值得注意的是，RLIT 模型超越了四种方法（即 Hybrid-GA、forward、backward 和 NNF），凸显了通过 RL 进行资产选择的优越性。TO 指标在所有数据集中的所有方法中取得了最小值，展示了本章提出

① https://cvxopt.org

中国知网 https://www.cnki.net

的成本感知奖励函数的优越性。总的来说，RLIT 的成功可归因于设计的 RA-DQN 能够准确评估资产价值，以及指导选择过程的奖励函数。p 值也证实了实验结果具有统计意义。

5.5.2.1 运行时间比较

2. 时间比较

测试时间比较：为了进行公正的比较，本研究为不同数据集设置了不同的时间限制，分别为 10 秒、50 秒、50 秒、50 秒、300 秒和 1200 秒。尽管本方法支持 GPU 加速，实验中仅采用 CPU 进行运算。如图 5-5展示，随着成分股数量的增加，求解时间也相应增长。Forward、backward、Hybrid-GA 和 MILP 方法耗时较多，前两者通过迭代选择达到指定数量的股票，对 N 和 K 的变化特别敏感。而后两者通过不断搜索直至达到时间限制或其他停止条件。与此相对，RLIT 为指数追踪任务提供了一种效率高且性能不受牺牲的解决策略，显示出其在实际应用中的实用性和可行性。特别是，随着成分股数量增长，求解所需时间亦增加。例如，DAX 100 数据集所需时间超过 Hang Seng 数据集，因为 DAX 100 包含更多成分股，需更长探索时间。相似地，由于 Russell 2000 数据集具有最多的成分股，因此求解时间最长。

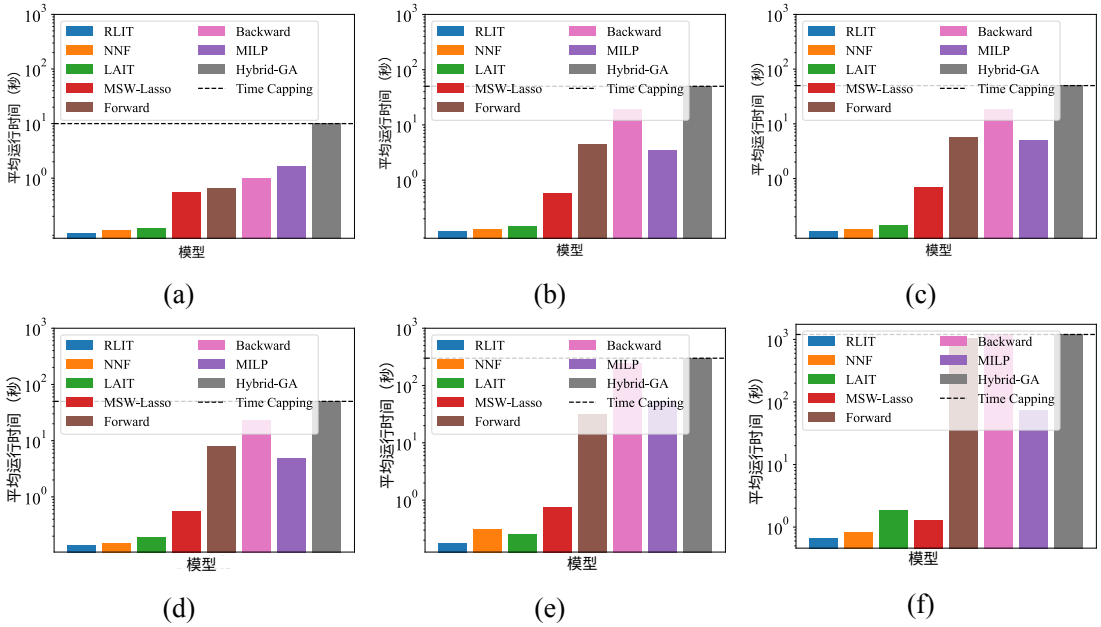


图 5-5 模型在六个基准指数数据集上的测试时间对比。(a)-(f) 分为为 Hang Seng、DAX 100、FTSE 100、S&P 100、Nikkei 225 和 Russell 2000 数据集。

训练时间比较：本章的模型根据表 5-3中列出的参数进行训练。训练时间因不

中国知网 https://www.cnki.net

表 5-3 训练参数表格。

参数	值
批量大小	32
缓冲区大小	$10e^5$
学习率	$10e^{-4}$
折扣因子	0.99
优化器	Adam
训练步数	$10e^4$
探索概率	0.05
软更新系数	0.005

同数据集而异，因为构成要素的规模决定了处理特征所需的时间。随着构成要素规模的增加，训练时间相应增长。本章的方法在六个数据集上的训练时长大约是：Heng Seng 指数 0.5 小时，DAX 100、FTSE 100 和 S&P 100 指数各 1 小时，Nikkei 225 指数 1.5 小时，以及 Russell 2000 指数 3 小时。如图 5-6所示，随着成份股规模的增大，收敛变得更具挑战性。方差随成份股规模的增加而增加，因为更大规模的成份股需要对动作空间进行更广泛的探索。

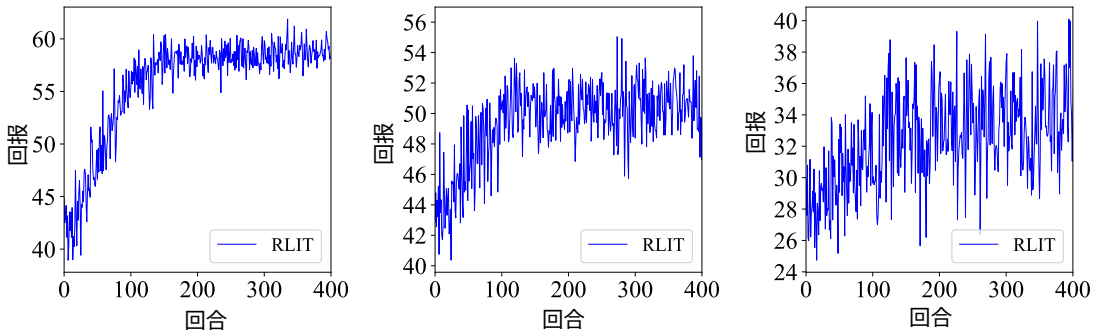
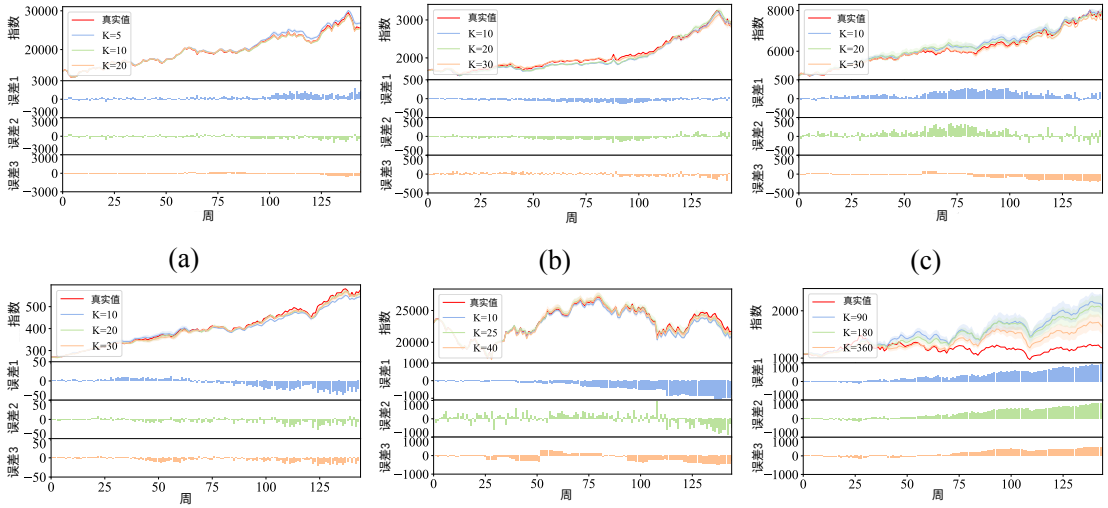


图 5-6 模型在不同规模的成份股数据集上的训练收敛可视化。(a)-(c) 分别为 Hang Seng、Nikkei 225 和 Russell 2000 数据集。

3. 可视化结果

图 5-7可视化了 RLIT 在几种部分复制投资组合大小 K 下的指数追踪性能。随着选定资产数量 K 的增加，追踪误差变小。真实指数由所有资产组成，因此资产选择越多，它就越接近真实指数。另外，指数中的成分股规模越大，追踪起来越具挑战性。这是因为随着成分股数量的增加，需要探索的行动空间也更大，找到最优组合也更加困难。尤其从图 5-7(a) 和图 5-7(f) 的第一栏可以看出，Russell 2000 与真实指数的偏差比 Hang Seng 指数更大。

中国知网 https://www.cnki.net



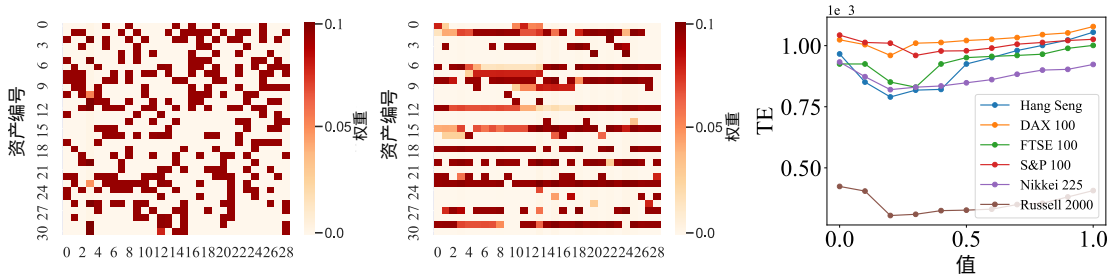
(a) (b) (c)

(d) (e) (f)

图 5-7 在六个指数数据集的追踪可视化。(a)-(f) 分为为 Hang Seng、DAX 100、FTSE 100、S&P 100、Nikkei 225 和 Russell 2000 数据集。

5.5.3 模型消融实验

通过去除奖励函数 R 内的交易成本约束 $H(j)$，本研究评估了该约束对奖励函数的影响。此评估仅在 Hang Seng 数据集上进行，因为其资产数量较少，更便于进行可视化分析。图 5-8(a)-(b) 展示了在有/无交易成本约束的条件下，部分交易过程的对比情况。从图 5-8(a) 中可见，每个交易日选择的资产并不一致（这一点可以从每列的颜色变化中清晰观察到）。相比之下，图 5-8(b) 中的颜色变化更稳定，表明资产 2 在几乎所有交易日中均有权重。这表明该资产在多数交易日被选中且持有固定的投资组合权重，因此无需为其重新平衡支付额外的交易成本。



(a) (b) (c)

图 5-8 奖励函数分析与参数 λ 值分析结果。(a) 为无交易成本约束；(b) 为有交易成本约束；(c) 不同 λ 值下的表现。

中国知网 https://www.cnki.net

5.5.4 模型参数分析

图 5-8(c) 展示了通过改变 λ 值对性能的影响。$\lambda = 1$ 意味着模型忽略指数与资产之间的关系，仅在模型 Q 时考虑资产内部关系，反之 $\lambda = 0$ 则表示忽略资产与资产之间的关系。在所有数据集上，当 λ 值在 0.1 到 0.4 范围内时，本章所提出的方法达到了最小的追踪误差。当 λ 等于 0 或 1 时，追踪误差达到最大。根据实验结果，建议 λ 设置为 0.2 或 0.3 以获得最佳追踪性能。

5.5.5 案例分析

通过追踪 S&P 500 指数，本研究进一步评估了所提出的方法在实际场景中的性能。评估所采用的数据集来自证券价格研究中心，涵盖 2006 年 1 月 3 日至 2022 年 12 月 30 日共 4280 天、430 个资产的日价格数据。回测的时间为从 2009 年 1 月 2 日至 2022 年 12 月 30 日，以前三年数据作为历史序列 L。采用的策略是季度重新调整，旨在对所有参与比较的方法进行长期投资效果的全面评估。

1. 评测程序

本次实验采取了针对长期投资的季度再平衡策略，并适用于所有进行比较的方法。回测的时间跨度从 2009 年 1 月 2 日起至 2022 年 12 月 30 日止。利用滑动窗口策略，如图 5-4(b) 所示，进行数据回测，具体步骤如下：每个季度末，例如 2009 年 3 月 31 日，利用之前三年的数据（2006 年 1 月 3 日至 2009 年 1 月 2 日）对模型进行训练，同时应用季度滑动窗口策略。随后，根据模型利用前一个季度数据（即 2009 年 1 月 2 日至 2009 年 3 月 31 日）得出的权重来调整投资组合。在 2009 年 6 月 29 日，使用过去三年的数据（2006 年 3 月 1 日至 2009 年 3 月 31 日）对模型参数进行微调，再次以前一个季度的数据（2009 年 3 月 31 日至 2009 年 6 月 29 日）作为输入进行再次调整。此过程每季度重复一次，直到 2022 年 12 月 30 日的最后一个交易日。

2. 实验分析

图 5-9可视化了 S&P 500 指数的追踪表现，其中 $K = 50$。显然，RLIT 在最小化追踪误差方面始终实现了优越的追踪表现。MILP、Hybrid-GA、LAIT 和 MSW-Lasso 的表现与指数有显著偏离。

表 5-4中展示了 K 分别取值为 30、40 和 50 的实验结果，对应于不同实际交易情况下可能选取的不同投资资产数量。表中的 p 值表明了结果的显著性水平，基于本研究提出的模型与最强对比方法之间的 10 对实验运行。每个模型运行 10 次，参数设置一致。在此数据集上决策的时间上限为 500 秒。从表 5-4中有以下发现：（i）就三个指标（TE、Δ_{SR} 和 TO）而言，本研究的方法对于不同 K 的选择均

中国知网 https://www.cnki.net

表 5-4 S&P 500 指数上不同 K 值情况下的指数追踪效果。

	K=30			K=40			K=50		
	TE↓	Δ_{SR} ↓	TO↓	TE↓	Δ_{SR} ↓	TO↓	TE↓	Δ_{SR} ↓	TO↓
MILP	4.3e-2	0.89	0.36	3.6e-2	0.88	0.34	2.0e-2	0.88	0.33
LAIT	3.7e-2	0.70	0.45	2.9e-2	0.39	0.39	2.4e-2	0.25	0.35
MSW-Lasso	2.3e-2	0.55	0.42	1.9e-2	0.57	0.39	1.5e-2	0.41	0.32
Forward	9.9e-3	0.24	0.47	9.3e-2	0.23	0.43	9.0e-3	0.16	0.40
Hybrid-GA	2.3e-2	1.52	0.37	2.0e-2	0.80	0.33	1.8e-2	1.61	0.37
NNF	9.2e-3	**0.14**	0.39	9.1e-3	0.14	0.36	8.8e-3	**0.07**	0.34
RLIT	**8.8e-3**	0.16	**0.33**	**8.6e-3**	**0.05**	**0.31**	**8.2e-3**	0.08	**0.29**
p 值	4.8e-4	7.6e-4	6.8e-4	3.0e-4	4.8e-4	6.0e-4	9.1e-4	6.1e-4	1.8e-4

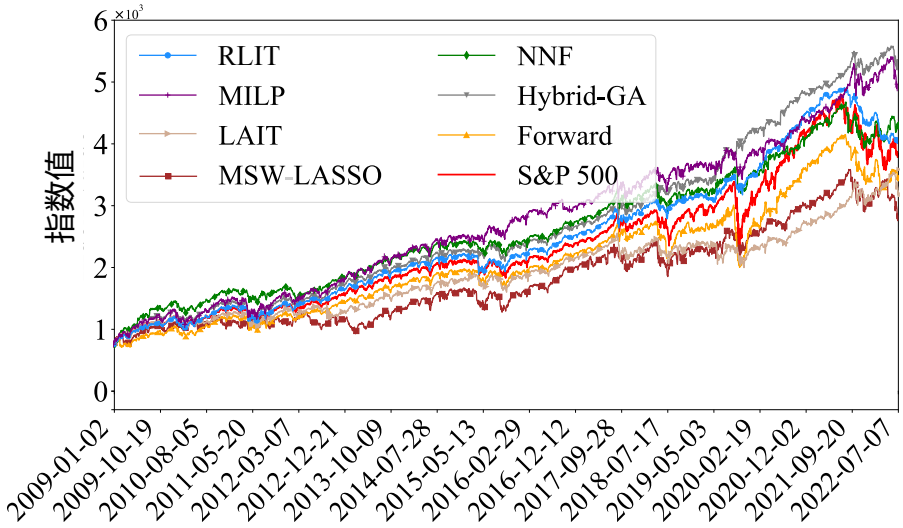


图 5-9 在 S&P 500 数据上的指数追踪可视化。

展示出与大多数对比模型相当的性能。（ii）总体而言，采用序列优化方法（例如，Forward、NNF 和 RLIT）的方法在 S&P 5000 指数上的表现优于采用联合优化方法（例如，MILP、LAIT 和 MSW-Lasso）的方法。鉴于 S&P 500 的大规模成分，联合优化导致了巨大的组合空间，难以有效处理。顺序解决资产选择和分配问题显著降低了复杂性，允许更好地管理指数追踪过程。（iii）所有方法倾向于随着投资组合大小 K 的增加而展现出改善的表现（较小的追踪误差）。例如，随着 K 的增加，TO 结果倾向于减少。

5.6 本章小结

本章讨论了如何将强化学习技术运用于部分指数追踪的挑战中。鉴于在资产选择阶段遇到的基数约束问题，本研究引入了一个基于深度强化学习的解决框架，旨在通过挖掘资产间以及资产与指数间的相关性来进行资产选择。具体而言，资

中国知网 https://www.cnki.net

产选择问题被建模为一个马尔可夫决策过程，通过探索和试错的方式进行求解。本文设计了一个关联性感知的深度 Q 网络（RA-DQN），该网络不仅能够有效应对基数约束问题，还通过集成局部敏感模块和指数感知模块来同时考虑资产内部和资产与指数间的相互作用。此外，引入了一个以成本为考量的奖励函数来评价资产选择的效果。实验测试表明，本研究提出的框架具有显著性，并在超过 10 年的 S&P 500 指数数据测试中表现优异，进一步验证了框架的优越性与实用性。

中国知网 https://www.cnki.net

第六章 被动投资策略：基于分层强化学习的指数追踪

6.1 引言

主动投资和被动投资是基金经理人遵循的两种基本策略。在主动投资中，基金经理人通过利用他们的专业知识选择有前景的资产，旨在超越市场表现。然而，这种方法通常需要频繁的资产再平衡，导致高昂的交易和管理费用[156]。实证分析表明，主动管理的基金通常难以在长时间内持续超越基准指数[140,157]。相反，被动投资策略建立在市场长期内不能被持续超越的前提上。这些策略较少依赖于基金经理的直觉，更多地依赖于预定的标准，如整体市场或行业表现。被动投资的一种流行形式是指数投资，投资者寻求构建一个与特定市场指数表现密切相符的投资组合，目标是实现与整体市场趋势一致的回报。这种方法通常被称为指数追踪[133]。

指数追踪可以根据构建追踪投资组合的方法分为全复制和部分复制两组。全复制是指将资金投资到指数的所有成份股中，与部分复制相比，它可以更准确地复制指数，具有最小的追踪误差。然而，由于持有大量指数成分股所造成的高额交易和管理费用等因素，全复制在实际中可能不太实用[133]。相反，部分复制顾名思义，只投资于指数成分股的一小部分，这种只选择小规模数量的做法也被称为“基数约束”。尽管它会导致更大的追踪误差，但通过选择更少数量的资产作为追踪组合，部分复制具有更低交易成本的优势。这些优势使得学术界和金融行业对部分复制方法的关注日益增加[23,135]。

部分指数追踪（Partial Index Tracking, PIT）涉及周期性地执行资产选择和资产配置两个子任务，以追踪特定市场指数的表现。资产选择的目标是从大量成分中选择指定数量的资产，由于其组合属性，优化该子任务具有一定难度。例如，通过从恒生指数的 31 个成分股中选择 10 个资产进行复制，需要探索超过 4000 万种可能的组合，这对计算机来说是一个艰难的计算挑战。当处理具有更多成分股的指数时，如标准普尔 500（S&P 500）和罗素 2000 (Russell 2000) 指数，这种复杂性将会进一步加大。相比之下，资产配置涉及在选定资产之间分配资金，其复杂性相对较低，因为它关注的资产数量较少。传统方法解决 PIT 通常将其构建为最小二乘回归问题，并根据它们是联合还是顺序优化资产选择和配置两个任务，将这些方法分为两类。第一类通过使用稀疏性约束（例如，ℓ_0 范数[133,158]）或将 PIT 转化为混合整数线性规划（Mix-integer Linear Programming, MILP）问题[144] 来联合优化资产选择和资产配置。这一种方法面临**复杂度高**问题，使得优化问题变为 NP

中国知网 https://www.cnki.net

难或需要使用商业整数规划求解器。一种更有效的方法是将资产选择和配置视为两个独立的任务并顺序优化它们，这些方法被归为第二类。这一类的流行方法包括使用指标（例如，协整关系[31]、指数-资产相似性[137,159]和复制因子[138]）或采用启发式算法（例如，遗传算法[21,154,160]、禁忌搜索[21]和模拟退火[22]）进行资产选择。随后，资产配置被建模为一个带约束项的回归问题进行处理，并使用二次规划（Quadratic Programming, QP）进行优化。然而，虽然与联合优化相比，顺序解决 PIT 的两个任务确实减轻了计算复杂性，但它引入了**独立优化**的问题——独立处理两个任务而没有机制保证两个任务的协同优化可能会损害模型的整体性能。

为了像顺序优化方法一样高效处理部分指数追踪，同时保留联合优化方法的协作优势，本章提出了一个分层模型用于部分指数追踪（Hierarchical Model for Partial Index Tracking, HIT）。为了实现 HIT，本研究利用 PIT 的周期性决策特性，将 PIT 重新构为一个分层马尔可夫决策过程（Hierarchical Markov Decision Process, HMDP）并通过分层强化学习（Hierarchical Reinforcement Learning, HRL）进行优化。HIT 包含一个高层策略用于学习选择资产，以及一个低层策略用于学习在所选资产中分配资本。两个策略都在奖励函数的指导下进行优化，以确保协同优化。然而，将 PIT 重构为 HMDP 除了定义状态和动作外，还需要解决两个重大技术挑战。第一个挑战涉及高层策略的资产选择。因为组合动作空间过于复杂，因此高层策略需要有能力挑选出合适的资产组合来追踪指数。第二个挑战涉及设计一个奖励函数，以促进两个策略的协同优化，旨在精确追踪指数并控制交易成本。由于本工作是首次将 HRL 应用于 PIT，还没有现成的奖励函数能满足 PIT 的特定需求。为了解决第一个挑战，本研究将资产选择重构为在资产树上的路径搜索过程。高层策略基于资产树上进行搜索，从根节点开始选择搜索到树的第 K 层的叶节点来挑选出 K 个资产。为了应对第二个挑战，本研究引入了一个成本敏感奖励函数。这个奖励函数旨在指导两个策略的协同优化，追求高精度指数复制的同时考虑交易成本的现实需求。本章的贡献有三个方面：

- 与将部分指数追踪视为回归问题的传统方法不同，本研究将 PIT 构建为一个分层马尔可夫问题并使用分层强化学习进行了优化。鉴于以往的部分指数追踪任务都被定义为回归问题，本工作从一个新的视角来解决 PIT 问题。
- 在 HIT 中，本研究设计了高层策略并引入了一个新颖的成本敏感奖励函数。高层策略对于有效处理复杂的动作空间至关重要，而成本敏感奖励函数鼓励两个策略之间的协同优化，优先考虑精确指数复制的同时考虑交易成本。
- 通过在 6 个基准数据集和超过 10 年的真实市场指数上进行了大量的实验，以证明该模型的有效性、效率和可扩展性。

中国知网 https://www.cnki.net

6.2 相关工作

6.2.1 指数追踪

传统的方法在解决部分指数追踪问题时都将其构建为一个受约束的最小二乘回归问题如公式 6-1，这个受约束的最小二乘回归问题已被证明是 NP 难[23,161]，无法在多项式时间内求得最优解：

$$\min_{w\geq 0,\sum_i w_i=1,\|w\|_0=K} \| Xw - Y \|_2^2 . \tag{6-1}$$

其中 $X \in \mathbb{R}^{T\times N}$ 是一个回报率矩阵，包含 N 个候选资产和 T 个时间步，$Y \in \mathbb{R}^T$ 是基准指数的真实回报率，$w \in \mathbb{R}^N$ 是分配给每个资产的权重向量。$\| w \|_0$ 是基数约束，要求 w 中非零元素的数量是一个固定数字 K，代表从 N 个成分中选出的资产数量。$w \geq 0$ 和 $\sum_i w_i = 1$ 分别是非负性与和为一约束，代表只做多策略和资本始终被充分利用。如前所述，PIT 可以被分解为两个任务：资产选择（解决基数约束）和资产配置（解决非负性与和为一约束）[21]。根据这两个任务是顺序解决还是联合解决，基于回归的研究方法可以被大致分为两组。

第一组方法将 PIT 构建为一个受约束的回归问题，并通过在投资组合权重上引入稀疏性约束来联合优化资产选择和配置。ℓ_0 范数和 ℓ_1 范数（也称为 LASSO）是构建稀疏追踪投资组合的两个广泛使用的稀疏性约束[133,143,158]。然而，引入 ℓ_0 和 ℓ_1 约束可能使回归问题在计算上复杂度变高或导致追踪性能不佳。近年来的工作如线性近似方法（Linear Approximation for Index Tracking, LAIT）[133] 和多步加权最小绝对值收缩和选择算子（Multi-step Weighted Least Absolute Shrinkage and Selection Operator, MSW-LASSO）[143] 以及基于混合整数规划问题[144,147] 已经在上一章的相关工作中介绍，这些方法由于搜索空间迅速增长而面临找不到最优解或求解的时间过长等问题，此处不再赘述。

第二组方法通过两步顺序解决资产选择和资产配置任务，以避免出现第一组方法所遇到的计算复杂度高的问题。在这些方法中，第二步（资产配置）通常被构建为一个二次回归问题，并通过二次规划进行优化。第二组的方法之间的关键区别在于第一步资产选择所采用的技术。常用的资产选择方法如基于领域知识[31,138,159]、基于启发式搜索算法[20,22,160]、基于神经网络算法[23,136,162,163] 等已经在上一章的相关工作中提及，此处不再赘述。然而，所有上述方法都独立优化资产选择和配置，没有进行协同优化。这项工作将部分指数追踪建模为层级马尔可夫模型，并通过分层强化学习进行优化，其中奖励函数作为两个任务之间的连接，促进协同优化，避免孤立优化。上一章的方案注重于将领域知识融入于资产选择

中国知网 https://www.cnki.net

过程中，而本章的方案更注重于如何将资产选择与配置两种方法的协同优化。

6.2.2 强化学习

1. 强化学习在金融领域的应用

深度强化学习已广泛应用于金融市场的多个领域，包括投资组合优化、股票交易和加密货币管理。在投资组合优化方面，Jiang 等人[18] 提出了相同独立评估系统集合，该系统利用三个不同的深度神经网络进行特征提取，并利用 DRL 技术优化三个网络以提高投资组合性能。对于股票交易，Jeong 和 Kim[148] 开发了一个自动化系统，该系统将强化学习与深度神经网络回归器结合起来预测股票数量。这一系统能够根据学习到的交易策略做出智能的交易决策。在加密货币管理的背景下，Xu 等人[83] 和 Zhang 等人[95] 都考虑了建模加密货币市场时加密货币之间的相互关系。他们采用深度强化学习技术捕捉市场动态并做出明智的交易决策。

2. 分层强化学习在金融领域的应用

分层强化学习为处理复杂的交易任务提供了一个强大的框架。它将具有复杂动作空间的任务分解为多个层次的子任务，每个层次负责一个特定的子任务。这种分层结构通过利用来自不同层次的策略的优势，使学习和决策更加高效。在金融和投资领域，HRL 已被应用于各种问题，包括投资组合管理[11,164] 和算法交易[165,166]。例如，Wang 等人[11] 提出了一个用于投资组合管理的分层强化系统，他们的方法涉及一个高层策略，以较低的频率确定投资组合权重以最大化长期利润，同时一个低层策略被调用以在较短的时间窗口内以较高的频率执行交易动作以最小化交易成本。Han 等人[165] 设计了一个分层强化学习框架，其中高层策略选择两种资产，低层策略执行一系列交易动作。为了清晰地概述 HRL 在金融中的应用，表 6-1中对相关研究进行了总结。显然，目前已有的分层强化学习在金融上的应用都集中在主动投资上，这些应用具有共同的奖励函数定义和相对简单的高层任务。然而，与这些应用不同的是，指数追踪由于独特的被动投资需求需要不同的奖励定义（与指数的距离），并涉及更加复杂的高层任务。

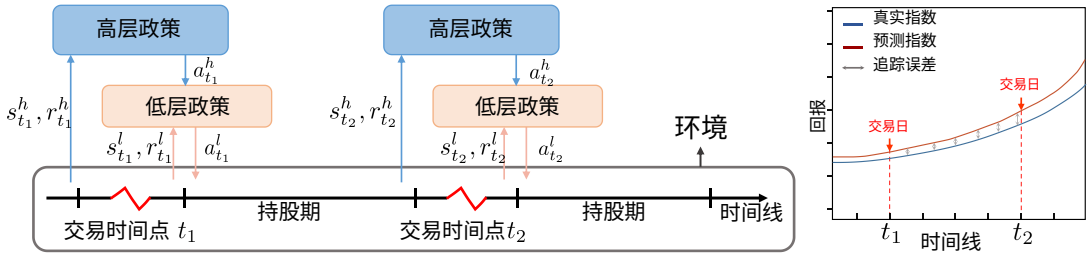
表 6-1 分层强化学习在金融领域的研究总结。

方法	高层策略	低层策略	奖励	任务 (策略类型)
文献 [11]	资产配置	订单执行	交易回报	资产管理 (主动投资)
文献 [164]	订单确定	订单执行	交易回报	资产管理 (主动投资)
文献 [165]	股票对选择	资产配置	交易回报	算法交易 (主动投资)
文献 [166]	代理选择	交易	交易回报	算法交易 (主动投资)
本研究	资产选择	资产配置	距离	指数追踪 (被动投资)

中国知网 https://www.cnki.net

6.3 预备知识

本节提供了关于部分指数追踪的一些初步知识。指数追踪是金融领域的一个基本任务，投资者旨在复制特定市场指数的表现，使他们能够将投资与特定行业或整体市场的表现对齐。PIT 通过周期性地选择少量指数成分股形成追踪资产组合并在它们之间分配资金来实现对齐指数回报这一目标。该过程包含两个不同的周期：持股期和交易期，如图 6-1(a) 所示。在持股期，选定的资产被持有，不进行任何买卖操作。由于市场动态，这些资产的价格在此期间可能会波动。如果投资组合权重被准确估计，回报将接近于指数的回报；否则，回报可能会偏离。在交易期，投资者有机会通过选择新的资产并在它们之间重新分配资本来调整追踪投资组合。持股期和交易期的长度可以根据具体设置而变化。



(a) (b)

图 6-1 模型框架概述图。(a) 是模型的数据流程；(b) 是指数追踪的示例。左边的两个交易时间 t_1 和 t_2 对应右边的两个交易箭头 t_1 和 t_2。

假设需要选择 K 个资产来复制具有 N 个成分股的指数。设 $x_t = [p_{0,t}, p_{1,t}, \cdots, p_{N,t}]^\top \in \mathbb{R}^N$ 表示时间步骤 t 所有资产的回报率。时间步骤 t 选择的资产用二元向量表示，记为 $a_t^h \in \mathbb{B}^N$，向量中的每个位置对应一个资产。如果 a_t^h 的第 i 个元素为 1，则表示选择了第 i 个资产；否则为 0。时间步骤 t 的分配权重表示为 $a_t^l \in \mathbb{R}^N$，其中 $N-K$ 个未选中资产的权重被设置为 0。在 PIT 中有以下定义：

定义 6.1 (约束条件) 二元向量 a_t^h 遵循基数约束，表示为 $\| a_t^h \|_0 = K$ 或 $\sum_{i=0}^{N} a_{i,t}^h = K$。这一约束意味着 a_t^h 中只有 K 个元素等于 1，表示从 N 个成分中选择 K 个资产。假定一个仅限买入的策略，这意味着 a_t^l 遵守非负性约束，即 $a_{i,t}^l \geq 0, \forall i \in \{1, 2, \cdots, N\}$，且资本始终被充分利用，这表明 a_t^l 遵循和为一约束 $\sum_{i=0}^{N} a_{i,t}^l = 1$。

定义 6.2 (投资组合回报) 假设持股期总共有 T 个时间步骤。在时间步骤

中国知网 https://www.cnki.net

$t \in \{1, 2, ..., T\}$ 的投资组合回报等于该时间步骤内所有选定资产的回报之和。

$$y'_t = x_t \cdot a_t^{l\top} = \sum_{i=0}^{N} a^l_{i,t} \cdot x_{i,t}. \tag{6-2}$$

因此，在持股期结束时可以获得 T 个回报值，即 $Y' = [y'_1, y'_2, ..., y'_T]$。

定义 6.3 (追踪误差) 追踪误差用于量化预测的投资组合回报与实际指数回报之间的差异，如图 6-1(b) 所示。可以使用多种标准来计算追踪误差，如均方根误差 [136] 和均方误差 [139]。本文采用欧几里得距离作为追踪误差。数学上，它可以表示为 $\epsilon = \frac{1}{T} \parallel Y - Y' \parallel_2^2$，其中 $Y = [y_1, y_2, ..., y_T]$ 是持股期间指数的回报。欧几里得距离是评估追踪投资组合性能的常用度量 [133, 155]，也作为将 PIT 建模为受约束回归问题时的目标函数 [23, 133, 147]。

定义 6.4 (交易成本) 在 t_2 时期交易期间，选择动作 $a^l_{t_2}$ 后，需要将当前投资组合从 $\hat{a}_{t_2} = \frac{a^l_{t_1} \odot \mathrm{p}_{t_2/t_1}}{a^{l\top}_{t_1} \mathrm{p}_{t_2/t_1}} \in \mathbb{R}^N$ 调整为 $a^l_{t_2}$，其中 p_{t_2/t_1} 代表 t_2 和 t_1 之间的相对收盘价，$\odot$ 表示元素间的乘积。如果 $a^l_{t_2}$ 与 $\hat{a}_{t_2}$ 之间的差异更大，则需要在资产之间进行更多的再平衡，这会导致交易成本的增加。因此，限制 $a^l_{t_2}$ 与 $\hat{a}_{t_2}$ 之间的差异是减少交易成本的有效方法。

上述基础知识提供了对部分指数追踪（PIT）的基本理解。接下来将详细阐述如何将 PIT 建模为一个分层的马尔可夫决策过程（MDP）。

6.4 模型：基于分层强化学习的指数追踪模型

当面对像 PIT 这样涉及复杂动作空间的决策任务时，层级强化学习（HRL）在有效管理任务的复杂性方面具有三个关键优势：

- **复杂决策空间的分解**：HRL 允许将复杂问题分解为多个层次，每个层次解决决策制定过程的一个具体方面。例如，PIT 问题可以被分解为高层的离散动作空间子任务进行资产选择，以及低层的连续动作空间子任务进行选定资产的权重分配。这种分解降低了动作空间的复杂性，使得有效解决 PIT 成为可能。
- **决策的层级结构**：通过学习高层和低层策略的决策层级结构，促进了决策的有效性。具体来说，高层策略学习资产选择，而低层策略学习权重分配。这种层级学习方法使得两个任务各司其职，决策制定更加有效。
- **任务间的协同优化**：HRL 通过奖励函数鼓励高层和低层任务间的协同优化。这种协作帮助克服了之前解决方案遇到的孤立优化的局限性。

因此，通过将 PIT 建模为 MDP 并通过 HRL 进行优化，这种方法在克服传统基于回归的解决方案面临的限制方面显示出了前景。

中国知网 https://www.cnki.net

6.4.1 高层马尔可夫决策过程-资产选择

资产选择过程被建模为一个高层的马尔可夫决策过程，表示为 MDP^h。它由四个元素的元组定义：MDP$^h = (\mathcal{S}^h, \mathcal{A}^h, \mathcal{P}^h, \mathcal{R}^h)$。这里，$\mathcal{S}^h$ 代表高层状态空间，$\mathcal{A}^h$ 是高层动作空间，$\mathcal{P}^h$ 是高层转移概率函数，而 $\mathcal{R}^h$ 是高层奖励函数。在这个 MDP 中，高层策略 π^h 在时间步 t 接收一个高层状态 $s_t^h \in \mathcal{S}^h$ 作为输入，并根据高层策略 π^h（由 θ^h 参数化）输出 $a_t^h \in \mathcal{A}^h$（即，选定的 K 个资产）。数学上，这可以表示为 $a_t^h = \pi^h(s_t^h, \theta^h)$。然后环境基于其动作 a_t^h 对策略 π^h 进行奖励，π^h 根据它接收到的奖励 r_t^h 更新其参数 θ^h。π^h 的最终目标是获得更高的累积奖励。

状态 高层状态 $s_t^h \in \mathcal{S}^h$ 由所有资产的历史基本特征表示。对于 N 个资产，特征是使用大小为 z 的时间窗口构建的：$s_t^h = [x_{t-z}, x_{t-z+1}, ..., x_t]$，其中 x_t 代表第 t 天 N 个资产的回报率。通过添加更多特征可以进一步丰富状态以改善模型性能。然而，在这项工作中，为了简洁和公平比较的目的，仅使用基本回报率来描述高层状态。

动作 一旦高层策略接收到高层状态 s_t^h，它就会根据状态 s_t^h 中的信息，做出一个高层动作 $a_t^h \in \mathcal{A}^h$ 来作为高层决策。这个动作 a_t^h 对应于从 N 个成分中选择 K 个资产。高层动作 a_t^h 可以表示为一个长度为 N 的二进制向量 $\mathbb{B}^N$。在这个向量中，第 i 个资产 $a_{i,t}^h = 1$ 表示该资产被选择，而 $a_{i,t}^h = 0$ 表示该资产未被选择。高层动作 a_t^h 遵守基数约束，如定义 1 中所述，表示为 $\| a_t^h \|_0$。这个约束确保被选择的资产数量不超过预定义的值 K。

奖励 与之前旨在最大化投资利润的基于强化学习的交易任务 [18,64,148] 不同，PIT 的主要目标是尽可能精确地复制选定的指数，同时控制交易成本。因此，直接应用现有的奖励函数可能不适用于 PIT。这项工作介绍了一个新的奖励函数，该函数借用映射函数使得奖励函数能够同时考虑追踪误差和交易成本。奖励映射是指使用一个融合了领域知识的函数转换原始的衡量标准，以更好地评估动作。

在 PIT 的背景下，高层动作不仅与追踪误差相关，还与交易成本相关。例如，选择两组完全不同的资产将需要完全重新配置，从而导致更高的交易成本。因此，高层奖励必须在追踪误差和交易成本之间找到平衡。高层奖励 r^h 可以如下形式化：

$$r^h = \frac{1}{2}[F(\epsilon) + F(1 - \mathcal{J}(a_{t_1}^h, a_{t_2}^h))], \tag{6-3}$$

其中 $F(\cdot)$ 是如图 6-2(a) 所示的奖励映射函数，通过数学表达式 $F(x)^{\frac{3}{5}} + x^{\frac{3}{5}} = 1$ 定义，适用于 $x > 0$ 的情况。这个映射函数根据其大小将变量值（如追踪误差 ϵ）反

中国知网 https://www.cnki.net

映为奖励。具体来说，追踪误差越小，它作为奖励的值就越高。使用杰卡德相似度评估 $a^h_{t_1}$ 和 $\hat{a}^h_{t_2}$ 之间的差异，表示为 $\mathcal{J}(a^h_{t_1}, a^h_{t_2})$。奖励的可视化显示在图 6-2(b) 中，其中高层策略被鼓励最小化追踪误差并最大化两个决策之间的差异以获得更多奖励（移向红色区域）。

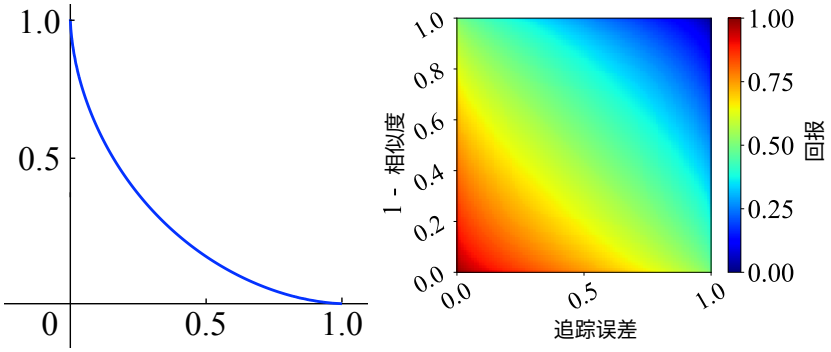


图 6-2 映射函数与奖励函数可视化。(a) 为映射函数；(b) 为奖励函数可视化。

6.4.2 低层马尔可夫决策过程-资产配置

在选择了资产之后，下一步是在所选资产之间分配资金权重。这一步骤被建模为一个低层马尔可夫决策过程（表示为 MDP^l），同样由四个元素的元组定义：MDP$^l = (\mathcal{S}^l, \mathcal{A}^l, \mathcal{P}^l, \mathcal{R}^l)$。这里，$\mathcal{S}^l$ 代表低层状态空间，$\mathcal{A}^l$ 是低层动作空间，$\mathcal{P}^l$ 是低层转移概率函数，$\mathcal{R}^l$ 是低层奖励函数。在 MDP^l 中，一个低层策略 π^l，由 θ^l 参数化，在时间步骤 t 接收一个低层状态 $s^l_t \in \mathcal{S}^l$，然后以一个低层动作 $a^l_t \in \mathcal{A}^l$ 做出响应。这个过程可以表示为 $a^l_t = \pi^l(s^l_t, \theta^l)$。环境根据策略 π^l 的动作 a^l_t 提供奖励，π^l 根据它获得的奖励 r^l_t 调整其参数，目的是在未来获得更高的奖励。

状态 低层状态 $s^l_t \in \mathcal{S}^l$ 用 K 个选定资产的基本历史特征来描述。对于这 K 个资产，特征是通过时间窗口 z 构建的：$s^l_t = \{x'_{t-z}, x'_{t-z+1}, \cdots, x'_t\}$，其中 x'_t 代表第 t 天 K 个资产的回报率。

动作 在接收到低层状态 s^l_t 后，低层策略将基于状态 s^l_t 给出低层动作 $a^l_t \in \mathcal{A}^l$，并在 K 个选定资产之间分配权重。因此，低层动作 $a^l_t \in \mathbb{R}^N$ 被定义为一个长度为 N 的连续动作空间，其中代表未选定资产的 $N-K$ 个元素被设为 0。a^l_t 遵循非负性和和为一的约束，即所有 i 的 $a^l_{i,t} \geq 0$ 且 $\sum_i a^l_{i,t} = 1$，如定义 1 所述。

中国知网 https://www.cnki.net

奖励 低层动作在控制追踪误差和交易成本方面扮演着至关重要的角色。当高层策略选择资产时，资金在它们之间的分配显著影响追踪表现和交易成本。即使高层策略选择相同的资产组合，为它们分配不同的权重可以由于投资组合的重新分配而导致不同的交易成本。低层奖励的设计受到最小化追踪误差和最大化当前投资组合权重 $\hat{a}^l_{t_1}$ 与最新权重 $a^l_{t_2}$ 的相似性的目标启发，以减少重新分配所产生的成本：

$$r^l = \frac{1}{2}[F(\epsilon) + F(1 - \mathcal{C}(a^l_{t_1}, \hat{a}^l_{t_2}))], \tag{6-4}$$

其中 $\mathcal{C}(\cdot)$ 代表余弦相似度，$\hat{a}^l_{t_1}$ 是当前投资组合权重，如定义 4 所述。奖励热图在图 6-2(b) 中展示，策略被鼓励做出导致更高奖励的决策，如红色区域所示。

6.4.3 高层策略的优化

在指数追踪中的资产选择问题涉及从更大的 N 个成分中挑选出一个 K 资产的子集。然而，由于其组合性质，探索所有可能的 K 资产组合变得不切实际。为了解决这个问题，本研究将资产选择重新构思为在树结构内的路径搜索过程，从根到第 K 层的叶子，其中树的深度限制为 K 个资产，确保了可管理的搜索复杂度。通过采用蒙特卡洛树搜索（Monte Carlo Tree Search, MCTS）来导航这一路径搜索过程。MCTS 利用模拟和统计结果来指导搜索，使其能够快速聚焦于预期奖励更高的有希望的路径。为了提高搜索效率，本研究结合了神经网络指导 [167] 和核回归（Kernel Regression, KR）在 MCTS 中。神经网络在树搜索中提供启发式指导，而 KR 为未访问的资产提供估算值，即使在处理大规模指数追踪时也能有效探索。以下将详细说明如何利用神经引导的 MCTS 进行资产选择，并阐述在过程中整合 KR 的方法。

1. 基于 MCTS 的资产选择

MCTS 利用大量的模拟来分析的当前所面临的选择，从而指导最有希望的下一步动作。最近，将 MCTS 与神经网络的结合进一步增强了其搜索能力并提高了搜索效率，其重大进展并在各个领域引起了广泛关注 [167]。在资产选择方面，本研究采用基于神经引导的 MCTS 方法来确定从树的根到第 K 层叶子的路径。具体来说，在每个时间步骤 t 和高层状态 s^h_t 下，初始化一棵以根节点 R 和边为起点的树，其中节点代表所选资产的状态，边代表节点下可用的资产。一般来说，每条边存储诸如先验概率 $P(n,i)$、访问计数 $N(n,i)$ 和动作值 $Q(n,i)$（从节点 n 选择资产 i 时的预期奖励）等属性来记录访问历史。这里，n 代表树中的一个节点，i 对应于节点 n 下的可用资产。从树的根节点 R 开始，MCTS 算法在神经网络 f_{θ_m} 的指导下模

中国知网 https://www.cnki.net

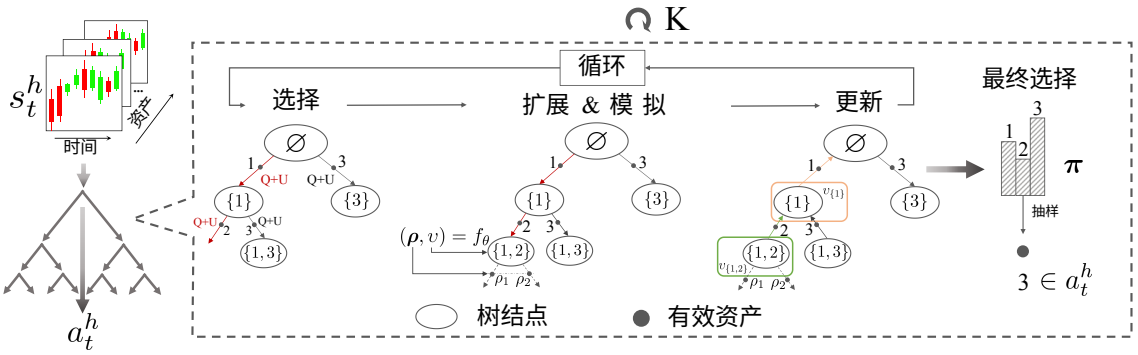


图 6-3 高层策略（MCTS）在资产选择上的示意图。MCTS 上方的循环标志表示选择资产直至达到所需数量 K 的迭代过程。

拟资产选择过程，随着模拟轮次的增加逐渐构建树。每轮模拟包括 4 个可重复的阶段（为了方便理解，这里的扩展和模拟阶段是写在一起的）和一个最终的选择阶段，如图 6-3所示：

- 选择：从根节点开始，迭代选择最大化树的上限置信区间（UCT）[168] 的子节点，通过边 i：$\arg\max_i(Q+U) = \arg\max_i(Q(n,i) + c \times P(n,i) \times \frac{\sqrt{N(n)}}{1+N(n,i)})$，直到达到一个可扩展的节点 L。UCT 公式的左部分（Q）鼓励开发已经显示出有利结果的资产，而右部分（U）鼓励探索与其父节点 n 相比访问较少的资产。探索与开发之间的权衡由可调节的探索常数 c 控制。一个可扩展的节点 L 指的是仍有未探索子节点的任何节点。例如，在图 6-3的选择阶段，可扩展节点 $\{1\}$ 有未探索的资产（如资产 2）可以导致创建一个新的未探索节点 $\{1,2\}$。
- 扩展与评估：如果可扩展的节点不代表树的决定性结束（即，没有达到树的 K 级），它将被扩展并由神经网络评估：$(\rho, \upsilon) = f_{\theta_m}(L)$，其中 ρ 是在 L 下可用资产的一组先验概率，υ 是节点 L 的预期奖励。然后完全扩展可扩展节点，并且每个边 i 用 $N(L,i) = 0, Q(L,i) = 0, P(L,i) = \rho_i$ 初始化，其中 $\rho_i \in \rho$。
- 更新：在这一阶段，从根节点到可扩展节点 L 的每条边通过增加其访问计数 $N(n,i)$ 并更新其平均奖励 $Q(n,i) = \frac{1}{N(n,i)} \sum_{(n,i)=(R \to L)} \upsilon_n$ 来更新，其中 $(R \to L)$ 表示从根节点到节点 L 遍历的所有节点和边 (n,i) 的元组，υ_n 是由 f_{θ_m} 评估的节点 n 的值。这个更新过程在图 6-3的更新阶段展示，其中节点 $\{1,2\}$ 的值 υ 由 $f_{\theta_m}(L)$ 评估，并迭代传播以更新前面的边。

一旦更新阶段完成，模拟再次重复，直到达到预定义的轮数 m。经过 m 轮模拟后，在最终的选择阶段，根据根节点下子节点的访问次数统计出选择概率 π。最频繁访问的节点被选为新的根节点，且与此节点相对应的资产（边）成为高层动作 a_t^h 的第一个元素。从新的根节点开始，执行另外 m 轮模拟以选择 a_t^h 的下一个元素。这

中国知网 https://www.cnki.net

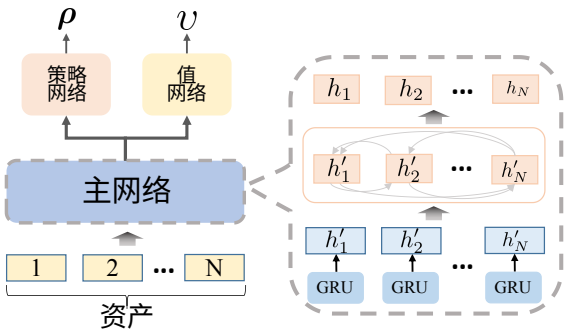


图 6-4 MCTS 的网络结构。

个过程持续进行，直到选定所需数量的资产 K。

网络结构：神经网络的结构 f_{θ_m} 如图 6-4所示。模型的核心是将资产的原始序列数据（即，回报率）编码成上下文感知的表示，即每个资产的表示能够捕捉资产之间的相似性。为了实现这一点，本模块采用门控循环单元来编码资产的原始序列数据，表示为 $s^h \in \mathbb{R}^{N\times z}$，编码成更稳健的表示 $h'_{i\in\{1,..,N\}} \in \mathbb{R}^{N\times d}$。这里，$N$ 代表资产数量，z 表示序列的长度。表示的维度用 d 表示。考虑到资产间隐性关系的重要性，本模块进一步通过利用 Transformer 编码器 [75] 增强了表示 h'_i。这使得表示 h'_i 能够捕捉资产间的相互依赖性，使它们更加具有上下文感知能力。增强后的表示记为 h_i。随后，使用两个独立的多层感知机作为策略层和价值层。策略层生成先验概率 ρ，而价值层产生价值估计 υ。

网络训练：在第 k 次选择时，选择概率 π_k 在执行 m 轮模拟后生成。π_k 是网络 f_{θ_m} 预测的先验概率的训练目标，因为它是通过多步前瞻搜索产生的，即通过考虑未来场景并确定哪些资产选择将导致有利结果。选择过程持续到 $k = K$，此时已抽样出所需数量的 K 个资产，并获得高层奖励 r_t^h。每次资产选择的数据存储为 $(\mathbf{x}_k, \pi_k, r_k)$，其中 $r_k = r_t^h$ 且 $\mathbf{x}_k$ 是基于高层状态 s_t^h 构建的节点状态。准确来说，$\mathbf{x}$ 通过将选定资产的特征掩码为 0 来表示一个节点状态。例如，为了在选择阶段表示图 6-3中的节点 $\{1,3\}$，资产 1 和 3 的特征被掩码为 0，表示为 $\mathbf{x} = s_t^h|_{1,3=0}$。初始根节点不需要掩码，因为在初始阶段没有资产被选中。通过之前时间步存储的数据 $(\mathbf{x}, \pi, r)$ 训练神经网络的参数 θ_m。高层策略本质上可以理解为一个树策略，在给定当前节点状态 $\mathbf{x}$ 时选择最有利的资产。神经网络 $f_{\theta_m}(\mathbf{x}) = (\rho, \upsilon)$ 的调整旨在最小化预测奖励 υ 与实际奖励 r 之间的误差，并最大化先验概率 ρ 与目标搜索概率 π 之间的相似性。通过梯度下降在损失函数 ℓ 上进行神经网络参数 θ_m 的调整，该损失函数结合了均方误差损失和交叉熵损失：

$$\ell = (r - \upsilon)^2 - \pi^T \log \rho + \lambda \|\theta\|^2, \tag{6-5}$$

中国知网 https://www.cnki.net

其中 λ 是正则化权重。

2. 结合核回归的 MCTS

在蒙特卡洛树搜索（MCTS）中，根节点下大量可用资产导致了探索不完全和选择阶段决策空间有限的挑战。具体来说，UCT 公式优先探索未访问的节点，但由于 MCTS 的回合数有限（即，$m < N$），每个节点只能被探索几次。增加搜索回合将会带来高昂的计算成本。对每个可用资产的不足探索可能导致不准确和非最优的资产选择。此外，在最终选择阶段，未访问的节点没有被选中的可能性，导致决策空间受限。

为了在拥有大规模成分股的指数中应用 MCTS，本研究提出了一个解决方案，该方案利用核回归（KR）基于已访问节点的属性来估计未访问节点的伪属性，如图 6-5所示。具体而言，在根节点 R 下，通过将未访问的资产视为查询点（$\mathcal{U}$），将已访问的资产视为观察数据点（$\mathcal{V}$），利用核函数 $K(h_a, h_b)$ 对 UCT 算法所需的未访问节点的伪计数 $N(R, h_a)$ 和伪值 $Q(R, h_a)$ 进行估计，其中 $h_a \in \mathcal{U}$ 和 $h_b \in \mathcal{V}$ 是节点表示。通过利用已访问节点的密度来近似估计未访问节点的访问次数（伪计数 $N(R, h_a)$），如等式 (6-6)所述，以评估其探索潜力。对于伪值 $Q(R, h_a)$ 则通过等式 (6-7)所述进行估计。通过核回归分析已访问与未访问节点间的相互关系来评估资产价值，这样便能预测未探索资产的潜在价值，为 MCTS 的节点选择提供有力参考。

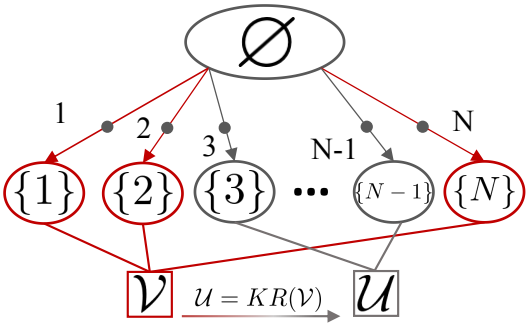


图 6-5 蒙特卡罗搜索树与核回归功能结合示意图。

$$N(R, h_a) = \sum_{h_b \in \mathcal{V}} K(h_a, h_b) N(R, h_b), \tag{6-6}$$

$$\mathbb{E}(Q(R, h_a)|\mathcal{V}) = \frac{\sum_{h_b \in \mathcal{V}} K(h_a, h_b) Q(R, h_b) N(R, h_b)}{\sum_{h_b \in \mathcal{V}} K(h_a, h_b) N(R, h_b)}. \tag{6-7}$$

KR 估计在模拟过程中以特定频率执行，例如每 20 轮模拟一次。随着访问节

中国知网 https://www.cnki.net

点数量的增加，对未访问节点中两个属性的估计变得更加准确。这些估计的属性随后被整合到 UCT 公式中，允许 UCT 算法在面对大规模成分股时，优先利用有利好的资产而不是仅仅专注于探索。此外，在最终选择阶段，基于它们的伪访问次数，未访问节点被分配样本概率。

6.4.4 低层策略的优化

低层策略将低层状态 s^l 作为输入，并输出配置权重 a^l，表示为 $a^l = \pi^l(s^l)$。由于资产配置涉及连续动作空间，本研究利用近端策略优化（Proximal Policy Optimization, PPO）算法 [132] 来优化低层策略，因为它在各种连续动作任务上展现出了卓越的性能。PPO 算法采用演员-评论家结构，其中演员网络使用与图 6-4中的高层策略相同的主干网络，后接一个策略层以生成动作 a^l。然后，动作被归一化以满足非负性与和为一的约束。评论家是一个两层感知机网络。

6.5 实验结果与分析

6.5.1 实验设置

1. 数据集介绍

本研究中使用了 6 个在指数追踪问题中常用的基准数据集，可以从 OR-Library 资源库[①]下载。每个数据集对应一个特定的市场指数及其成分股。这 6 个市场指数分别是：Hang Seng 指数（香港）、DAX 100 指数（德国）、FTSE 100 指数（英国）、Nikkei 225 指数（日本）、S&P 500 指数（美国）和 Russell 2000 指数（美国）。这些指数的成分股数量从构成 Hang Seng 指数的 31 个到构成 Russell 2000 指数的 1318 个不等。对于每个成分股，提供了 291 个周价格。数据的更详细统计信息在表 6-2中总结。

表 6-2 基准数据集的介绍。

规模	指数	训练周数	测试周数	N	K
小	Hang Seng	146	145	31	10
中	DAX 100	146	145	85	10
	FTSE 100	146	145	89	10
	Nikkei 225	146	145	225	10
大	S&P 500	146	145	457	40
	Russell 2000	146	145	1318	90

① http://people.brunel.ac.uk/mastjjb/jeb/orlib/indtrackinfo.html

中国知网 https://www.cnki.net

2. 对比模型

本实验与六种基准方法进行了比较。前三种基准方法同时优化资产选择和配置，而后三种基准方法则依次优化它们。(1) MILP [144]：这种基准方法将指数追踪形式化为一个混合整数问题，并利用 Gurobi 求解器（对学术研究免费）来找到解决方案。(2) LAIT [133]：LAIT 引入了 ℓ_0 正则化以确保投资组合的稀疏性，但采用了一个近似函数来计算 ℓ_0 范数。(3) MSW-LASSO [143]：MSW-LASSO 使用多步加权 LASSO 作为优化效率的稀疏性约束。(4) Hybrid-GA [20]：这个方法利用遗传算法进行资产选择，然后使用 QP 进行配置优化。(5) Forward [23]：Forward 首先使用 QP 拟合所有资产，然后迭代选择权重最大的资产，直到达到所需的资产数量。(6) NNF [136]：NNF 在部分指数追踪（PIT）中采用了一个包含固定噪声的神经网络进行资产选择和配置。这个神经网络在训练期间包括固定噪声。

3. 评测指标

本实验采用与其他先前研究 [22, 154, 155] 相同的评测指标，包括样本内追踪误差（Tracking Error Insample, TEI）和样本外追踪误差（Tracking Error Outsample, TEO）。TEI 测量训练阶段的追踪误差，而 TEO 测量测试阶段的追踪误差。追踪误差提供了追踪器跟随指数的密切程度的见解。追踪误差的数学定义如定义 6.3所表示：$\epsilon = \frac{1}{T} \parallel Y - Y' \parallel_2^2$。

4. 评测程序

本实验采用了 [22, 154, 155] 中的每日调整策略来评估模型的性能。此外，数据被划分为训练集和测试集，训练集包括前 146 周的周价格，测试集包括接下来的 145 周，以验证模型的性能，分割方法与第五章的基准模型分割方法一致，此处不再进行描述。

5. 参数设置

所有实验均在配备英特尔 (R) Xeon(R) Silver 4214R CPU、256GB 运行内存及 NVIDIA GeForce RTX 3090（32GB 内存）的机器上完成。为保证结果的可靠性并降低随机性影响，每个模型运行 10 次，并计算平均性能。该框架与多个基准进行了比较，这些基准的结果根据其原始论文中描述的最优设置，通过开源代码复现得到。该框架的实现基于 PyTorch，并采用 Adam 优化器进行优化。在训练过程中，进行了 50,000 次训练步骤，学习率设置为 $2.0e^{-4}$。为了提高蒙特卡洛树搜索过程的效率，采用虚拟损失技术 [169] 并行化搜索过程，每次模拟分配 300 个线程。训练阶段的模拟轮数设置为 300，而测试阶段则为 800。

中国知网 https://www.cnki.net

表 6-3 所有模型在 6 个基准数据集上的表现。

数据集	Hang Seng		DAX 100		FTSE 100		Nikkei 225		S&P 500		Russell 2000	
	TEI	TEO	TEI	TEO	TEI	TEO	TEI	TEO	TEI	TEO	TEI	TEO
MILP [144]	7.6e-4	**7.8e-4**	1.3e-3	1.1e-3	1.2e-3	9.7e-4	9.1e-4	3.1e-3	2.7e-3	6.8e-3	4.9e-3	3.5e-3
MSW-LASSO [143]	4.7e-3	4.2e-3	2.9e-3	2.4e-3	2.4e-3	2.1e-3	1.6e-3	1.7e-3	3.7e-3	3.0e-3	3.5e-3	6.3e-3
LAIT [133]	1.9e-3	1.5e-3	2.6e-3	1.6e-3	2.5e-3	1.5e-3	2.1e-3	1.7e-3	9.7e-3	9.4e-3	4.2e-3	4.0e-3
Hybrid-GA [20]	8.1e-4	8.7e-4	7.5e-4	8.1e-4	8.9e-4	8.6e-4	8.7e-4	9.7e-4	9.4e-4	9.8e-4	3.8e-3	4.3e-3
Forward [23]	8.3e-4	8.9e-4	8.8e-4	7.6e-4	8.3e-4	8.9e-4	8.9e-4	9.3e-4	9.0e-4	1.4e-3	5.9e-3	5.8e-3
NNF [136]	7.9e-4	8.4e-4	7.8e-4	7.7e-4	8.5e-4	8.4e-4	8.4e-4	8.6e-4	8.9e-4	9.6e-4	4.7e-3	2.9e-3
HIT	**7.4e-4**	8.2e-4	**7.2e-4**	**7.1e-4**	**7.8e-4**	**8.2e-4**	**8.2e-4**	**8.3e-4**	**7.8e-4**	**8.1e-4**	**1.6e-3**	**2.2e-3**
p 值	3.0e-3	3.0e-3	5.0e-3	3.0e-3	6.0e-3	2.0e-3	3.0e-3	3.0e-3	3.0e-3	2.0e-3	2.0e-3	3.0e-3

6.5.2 实验结果

1. 在基准数据集上的结果分析

所有模型在基准数据集上的实验结果展示在表 6-3中，其中表现最好的模型以粗体标记，次优则以下划线标记。尽管 MILP 方法在应用于成分股较少的 Hang Seng 指数时表现良好，但在应用于成分股数量较多的其他指数时性能下降。LAIT 和 MSW-LASSO 采取了简化 ℓ_0 和 ℓ_1 约束以减少计算复杂性的方法。然而，这种简化以牺牲有效性为代价，导致在所有数据集上的解决方案都不是最优的。相反，如 Forward、NNF、Hybrid-GA 和 HIT 等方法，它们顺序地处理资产选择和配置，始终展示出稳健的性能。这突显了顺序优化方法在解决 PIT 问题中的实用性。值得注意的是，HIT 模型超越了所有其他基于顺序的方法，展示了其优越性。

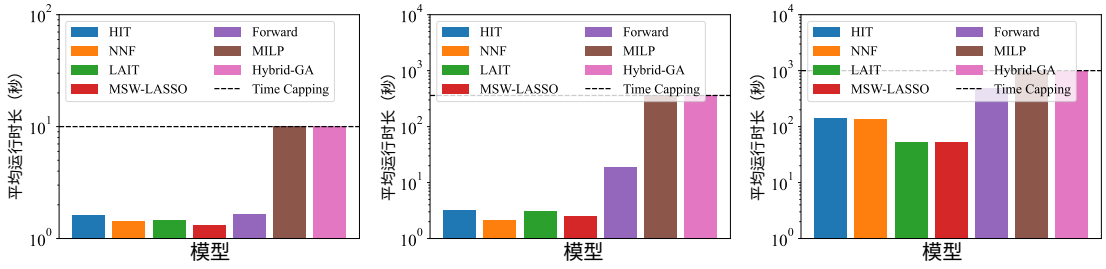
此外，本实验还进行了配对 Wilcoxon 符号秩检验，以验证方法超越次优模型的统计性意义。表 6-3中报告的 p 值反映了结果的统计性意义。统计性意义分析是基于提出的模型次优的 10 对实验结果，每个模型采用相同的参数设置运行 10 次进行的。在众多研究中，通常采用 0.05 的置信水平 [89,90]。当 p 值小于 0.05 时，意味着模型性能与次优模型相比具有统计意义的显著性。在表 6-3中，可以看到所有的 p 值均小于 0.05，证明实验结果具有统计学上的意义，并且提出的模型表现优于对比模型。

2. 运行时间对比

本实验还比较了每个交易周期做出决策时的平均时间消耗，如图 6-6所示。对于小、中和大规模成份股的指数，决策的时间限制分别为 10 秒、50 秒和 1000 秒，可以观察到以下趋势：（1）决策所需的时间随问题规模的增加而增加。（2）Hybrid-GA 和 MILP 是最耗时的方法。它们持续搜索最优资产，直到达到时间上限或获得最优解。虽然 LAIT 和 MSW-LASSO 与其他方法相比消耗的时间较少，但如前所述，这种效率是以牺牲有效性为代价的。尽管 HIT 不是最节省时间的，但

中国知网 https://www.cnki.net

它可以在控制时间成本的情况下提供解决方案。这一方面使 HIT 更适合实际应用，更具实用性。

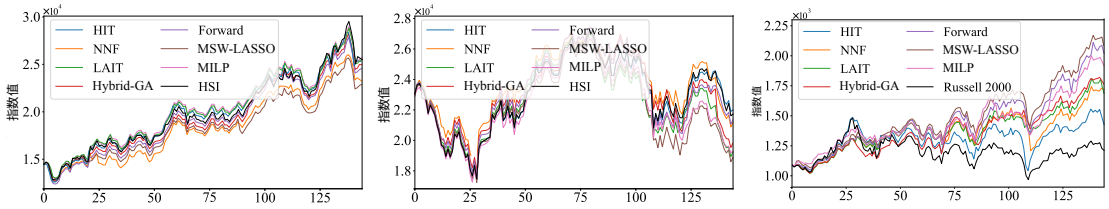


(a) (b) (c)

图 6-6 平均决策时间对比。(a)-(c) 分别为 Hang Seng、Nikkei 225 和 Russell 2000 数据集。

3. 可视化分析

图 6-7展示了所有模型在具有三种不同成分规模的数据集上的指数追踪性能。通过图中可以观察到以下趋势：（1）在处理小规模指数时，所有模型的表现相似，如图 6-7(a) 所示，导致的差异最小。（2）随着数据集大小的增加，如图 6-7(b) 中展示的中等规模指数，模型的表现开始表现出多样性，但它们仍然能够有效追踪指数。（3）然而，在尝试追踪大规模指数时，显著的差异变得明显。例如，MSW-LASSO 模型表现最差，而 HIT 模型紧密跟随指数，如图 6-7(c) 所见。



(a) Hang Seng (b) Nikkei 225 (c) Russell 2000

图 6-7 模型在小、中、大规模成份股的指数上的追踪效果。

6.5.3 模型消融实验

本节进行了消融研究，以调查各个模块对模型性能的影响。具体而言，本节研究了（1）MCTS 中核回归的影响，（2）不同高层策略对模型总体性能的影响，以及（3）层次结构的优势。下文提供了每项消融研究的详细解释。

1. 核回归功能的影响

本实验评估了在蒙特卡洛树搜索中应用核回归的影响，结果展示在表 6-4的第 1 和第 2 列。对于小规模和中等规模的数据集，KR 并没有带来显著改进。然而，

中国知网 https://www.cnki.net

表 6-4 在有/无核回归以及有/无成本约束的情况下模型的表现。

数据集	无核回归		有成本约束和核回归		无成本约束	
	TEI	TEO	TEI	TEO	TEI	TEO
Hang Seng	9.5e-4	9.1e-4	8.7e-4	8.5e-4	7.4e-4	8.2e-4
DAX 100	8.6e-4	8.0e-4	7.8e-4	7.5e-4	7.2e-4	7.1e-4
FTSE 100	9.9e-4	9.6e-4	8.6e-4	8.9e-4	7.8e-4	8.2e-4
Nikkei 225	1.2e-3	9.6e-4	9.5e-4	9.0e-4	8.2e-4	8.3e-4
S&P 500	1.0e-3	1.2e-3	9.5e-4	9.2e-4	7.8e-4	8.1e-4
Russell 2000	2.7e-3	3.3e-3	1.9e-3	2.9e-3	1.6e-3	2.2e-3

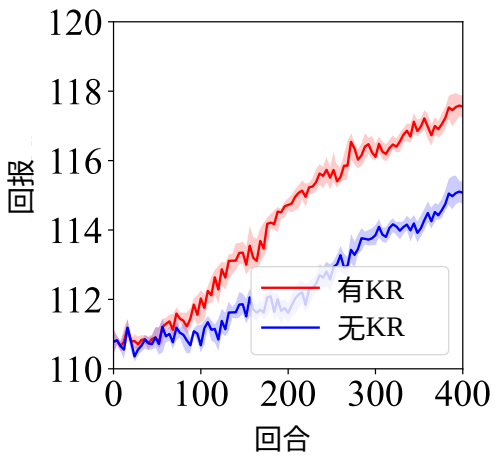


图 6-8 有 (红色) 和无 (蓝色)KR 情况下的回报比较。

对于大规模数据集，KR 模块导致了实质性的改善。KR 模块旨在估计根节点下未访问节点的伪值，允许树在开发和探索之间取得平衡，而不是仅仅侧重于探索。图 6-8进一步表明，配合 KR 的 MCTS 能够探索到更有前景的资产，从而获得更高的奖励。随着数据集规模的增加，KR 模块的有效性变得更加明显，因为它有助于更好地估计未访问节点的值，并指导 MCTS 选择更高回报的资产。

2. 高层策略的影响

本实验通过用深度 Q 网络（Deep Q-network, DQN）策略替换 MCTS，并且 DQN 使用与 MCTS 相同的网络结构来评估 MCTS 的影响。DQN 使用相同的高层状态作为输入来预测每个资产的值。前 K 个资产被选为追踪投资组合以分配权重。如图 6-9所示，当使用 MCTS 作为高层策略时，追踪误差收敛到更低的值。这表明 MCTS 在资产选择方面比 DQN 做得更好。这种差异可以归因于 MCTS 的模拟机制，它模拟了每个资产的选择，并基于模拟结果选择最有前景的资产。相比之下，DQN 仅估计每个资产的 Q 值，难以有效处理复杂的资产选择过程。

中国知网 https://www.cnki.net

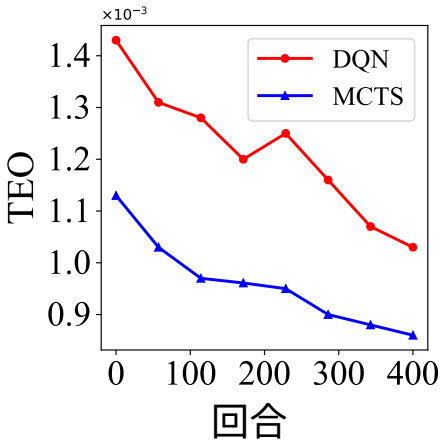


图 6-9 两种不同的高层策略情况下 TEO 性能的比较。

3. 分层结构的影响

为了评估分层架构的优势，本实验开发了一个非分层的强化学习框架变种。在这个变种中，策略直接输出权重向量 w（如公式 (6-1)中定义），而不是将其分解为两个动作空间（a_t^h 和 a_t^l）。这个变种的策略从环境中取状态 s_t（等同于 s^h）并直接输出动作 w。强化学习的环境配备了等式 (6-4)中显示的奖励函数，根据其动作 w_t 奖励策略。向量 w 表示为 $w_t = [w_{t,1}, ..., w_{t,N}] \in \mathbb{R}^{1\times N}$，其中 $w_{t,i} > 0$ 表示在时间步 t，资产 i 被选入追踪投资组合，其值代表分配给资产 i 的权重。通过剪裁或标准化等操作，w_t 向量满足等式 (6-1)中指定的三个约束。由于权重向量 w 涉及连续动作空间，采用 PPO 算法来优化这个变种策略，利用其在处理连续动作任务中的强大性能。PPO 参数的实现细节与 Stable-Baselines3 [170] 保持一致。

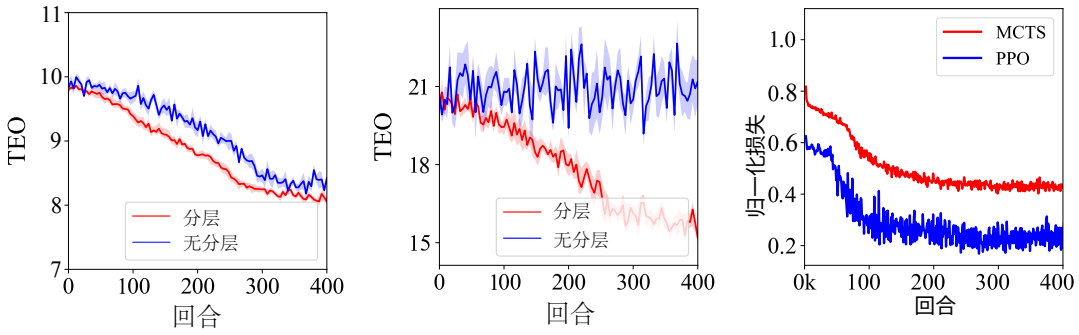


图 6-10 (a) 和 (b) 是在 Hang Seng 和 Nikkei 225 数据集下的误差变化图，误差随着训练回合变化而变化。(c) 是在 Russell 2000 数据集下高层和低层两个策略的策略损失。

图 6-10(a) 和图 6-10(b) 展示了分层结构的优势。总的来说，两个模型在 Hang

中国知网 https://www.cnki.net

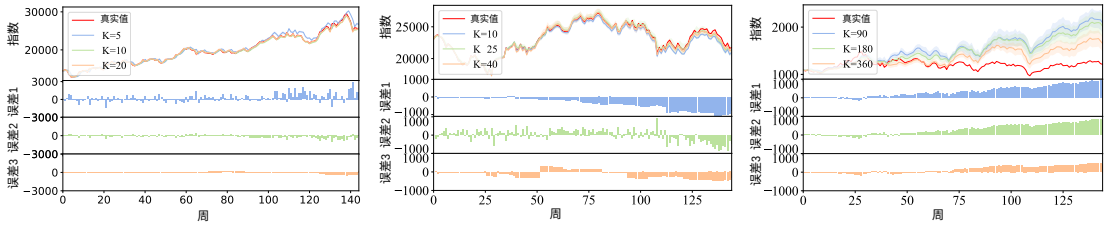
Seng 指数（成分股数量最小的指数）上的追踪误差表现相似。然而，变种在大规模成分股下难以减少追踪误差，这是由于过大的动作空间和难以探索到有效动作（权重）造成的。相比之下，HIT 型在降低追踪误差方面显示出显著的减少，因为分层架构简化了动作空间，有效地探索了每个子任务的空间。图 6-10(c) 还反映了两个分层策略的有效学习，其中高层（MCTS）和低层（PPO）的损失都同步减少，展示了两个策略的协同优化。

6.5.4 模型参数分析

本节进行了模型参数的探索，以探究：（1）所选资产数量 K 对性能的影响，（2）奖励函数中成本约束的影响。下文提供了每个参数项的探究结果。

1. 不同选择数量 K 的影响

图 6-11展示了 HIT 在不同部分复制投资组合大小 K 下的指数追踪性能。从图 6-11可以观察到随着所选资产数量 K 的增加，追踪误差变小。通过包含投资组合中的所有资产可以更逼近真实指数，所以在有更多资金究可以做更多的资产选择（即更大的 K）的前提下越可以更接近真实指数曲线。



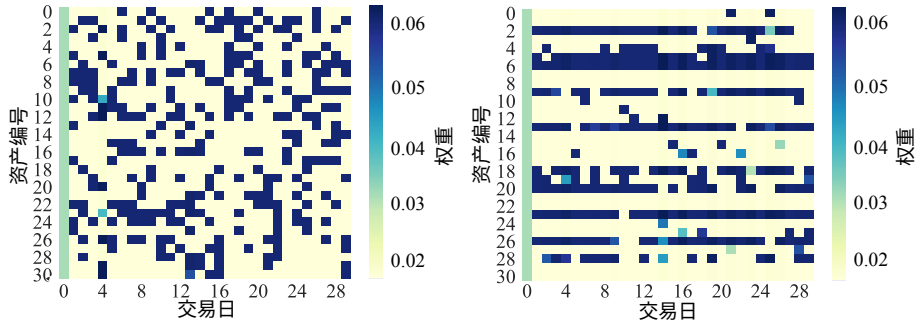
(a) (b) (c)

图 6-11 不同 K 值的情况下，3 个基准数据集的追踪效果可视化。误差 1、误差 2 和误差 3 分别对应三个不同的 K 值（与颜色相对应）。(a)-(c) 分别为 Hang Seng、Nikkei 225 和 Russell 2000 数据集。

2. 奖励函数中成本约束的影响

为了评估奖励函数中交易成本约束的效果，本实验在 Hang Seng 数据集上进行了有和无交易约束的实验。图 6-12分别展示了不带和带有交易费用约束的一小部分交易过程。从图 6-12(a) 中可以观察到每个交易日所选资产在变化（每行的颜色变化所示），而图 6-12(b) 显示了更稳定的颜色过渡，表明某些资产在大多数交易日都被选中。这种稳定性减少了频繁交易，从而降低了成本。实验还比较了奖励函数中有无成本约束时的追踪误差，如表 6-4（第 2 和 3 列）所示。带有成本约束的追踪误差更高，因为交易成本限制了重新调整投资组合的灵活性。较高的交易成本通常会阻止频繁的资金分配调整，导致无法紧密追踪目标指数。

中国知网 https://www.cnki.net



(a) (b)

图 6-12 有、无成本约束情况下的交易决策可视化。(a) 为无成本约束；(b) 为有成本约束。

6.5.5 案例分析

考虑到用于与基准进行比较的 6 个评价标准已经不再是最新的，且评价指标较为有限，本研究对 HIT 模型在现实市场中的表现进行了更为全面和严格的评估。因此，采用了 HIT 模型来追踪 S&P 500 指数，并引入了更多的经验评价指标来衡量模型的表现。此次实验旨在为模型在实际市场环境中的性能提供一个更精确和相关的评估。

1. 数据集介绍

本实验从证券价格研究中心（Center for Research in Security Prices, CRSP）获取了研究所需的数据，该中心因其在金融数据准确性和可靠性方面的声誉而被广泛认可。数据集涵盖了从 2006 年 1 月 3 日至 2022 年 12 月 30 日，共 4218 天的日价格数据。经过过滤掉数据缺失的资产后，用于分析的资产总数达到了 430 个。这一高质量的数据集保证了实验结果的可信性。

2. 评测程序

在本实验中，对所有比较方法采用了季度调整策略，专注于长期投资的表现。回测周期从 2009 年 1 月 2 日延续至 2022 年 10 月 31 日。采用了滑动窗口策略以提取数据，具体过程如下：于每季度末（例如，2009 年 3 月 31 日），使用前三年的历史数据（即 2006 年 1 月 3 日至 2009 年 1 月 2 日）对模型进行训练。接着，利用训练完成的模型，以前一季度的数据（即 2009 年 1 月 2 日至 2009 年 3 月 30 日）作为输入，计算下一季度的投资权重。然后，在调整日（即 2009 年 3 月 31 日）按照计算得到的权重调整投资组合。到了 2009 年 6 月 29 日，同样采用过去三年的数据（即 2006 年 3 月 1 日至 2009 年 3 月 31 日）调整模型参数，并使用前一季度的数据（即 2009 年 3 月 31 日至 2009 年 6 月 28 日）作为输入调整投资组合。此流程一直重复至 2022 年 10 月 31 日的最后一个交易日，确保模型始终以最新数据

中国知网 https://www.cnki.net

进行训练。

3. 评测指标

本实验最直观的性能评价指标为测试期间指数与部分复制投资组合之间的追踪误差（TEO）。除 TEO 外，本实验还报告了两个附加指标：一）夏普比率（Sharp Ratio, SR），该指标同时考虑投资组合的回报与风险，定义为 $\mathrm{SR} = \frac{\mathrm{Average}(Y')}{\mathrm{Standard\ Deviation}(Y')}$，$Y'$ 为回报向量；二）最大回撤（Max Drawdown, MDD），MDD 衡量一定时间内投资组合从最高点到最低点的最大比例下跌，作为下行风险的衡量指标，定义为 $\mathrm{MDD} = \max_{\tau_1:\tau_2>\tau_1} \frac{W_{\tau_1} - W_{\tau_2}}{W_{\tau_1}}$，$W_\tau = \sum_{t=1}^{\tau} Y'_t$ 为时间步 τ 的回报总和。选择这两个指标的主要原因是使其与指数尽可能保持一致，而非优先追求高 SR 和低 MDD，后者可能是其他主动投资应用的焦点。

4. 实验结果与分析

表 6-5中展示了两种 $K \in \{30, 50\}$ 的结果，代表了在各种实际交易场景中可能希望的持股数量。表中的 p 值表示本章提出的模型与次优模型之间的 10 对实验运行的比较的统计性意义。每个模型运行 10 次，参数设置保持一致。在这个数据集上，决策的时间上限是 300 秒。

从表 6-5可以观察到以下情况：（i）就三个指标（TEO, SR, 和 MDD）来说，本章的方法在两种 K 的选择上展示了与大多数基线相媲美的性能。尽管在 MDD 指标上本模型的表现略低于 NNF，但仍然非常有竞争力。（ii）总的来说，采用序列优化方法（例如，Forward, NNF 和 HIT）的方法的表现优于采用联合优化方法（例如，MILP, LAIT 和 MSW-LASSO）的方法在 S&P 500 指数上。考虑到 S&P 500 的大规模成分，联合优化它会导致一个庞大的组合空间，难以有效处理。将任务分解为独立的子任务显著降低了复杂性，从而更好地管理指数追踪过程。采用序列优化方法的性能展现出极小的变化，进一步确认了序列优化方式在处理大规模成分指数追踪方面的有效性。此外，一些方法（例如，MILP 和 Hybrid-GA）展示出高 SR，这在被动投资中可能不受青睐因为被动投资旨在与指数表现保持一致而不是追求高 SR。（ii）所有方法倾向于在更大的投资组合规模 K 下展现出更好的表现。这是因为指数的表现是由所有成份股共同影响的，而成分投的数量越多，越可以准确复制指数的表现。

图 6-13中可视化了 S&P 500 指数在 $K = 50$ 时的追踪表现。线上的每个标记点代表一个交易日。从图中可以看出，与其他模型相比，HIT 实现了最佳的追踪表现，追踪误差最小。另一方面，MILP、Hybrid-GA、LAIT 和 MSW-LASSO 的结果展示了与指数的显著偏离，如表 6-5中所示的 TEO 指标。尽管 Forward 或 NNF 的表现相对可接受，但仍不如本章的模型那样令人满意。

中国知网 https://www.cnki.net

表 6-5 不同模型在 S&P 500 数据集上的表现。Δ 是模型预测指数与真实指数之间的绝对值误差。

模型	K=30			K=50		
	TEO	SR (Δ)	MDD (Δ)	TEO	SR (Δ)	MDD (Δ)
S&P 500	0	1.09	-0.19	0	1.09	-0.19
MILP	4.3e-2	1.89 (0.89)	-0.24 (0.05)	2.0e-2	1.97 (0.88)	-0.26 (0.07)
LAIT	2.7e-2	0.39 (0.70)	-0.54 (0.35)	1.2e-2	0.92 (0.17)	-0.58 (0.39)
MSW-LASSO	2.3e-2	0.54 (0.55)	-0.61 (0.42)	1.5e-2	0.68 (0.41)	-0.45 (0.26)
Hybrid-GA	2.1e-2	2.61 (1.52)	-0.25 (0.06)	1.8e-2	2.70 (1.61)	-0.12 (0.07)
Forward	9.9e-3	0.85 (0.24)	-0.25 (0.06)	9.4e-3	0.93 (0.16)	-0.29 (0.10)
NNF	9.2e-3	0.96 (0.13)	-0.28 (0.09)	8.8e-3	1.01 (0.08)	**-0.17 (0.02)**
HIT	**7.6e-3**	**1.14 (0.05)**	**-0.15 (0.04)**	**5.6e-3**	**1.05 (0.04)**	-0.16 (0.03)
p 值	5.0e-3	3.0e-3	1.0e-3	6.0e-3	3.0e-3	3.0e-3

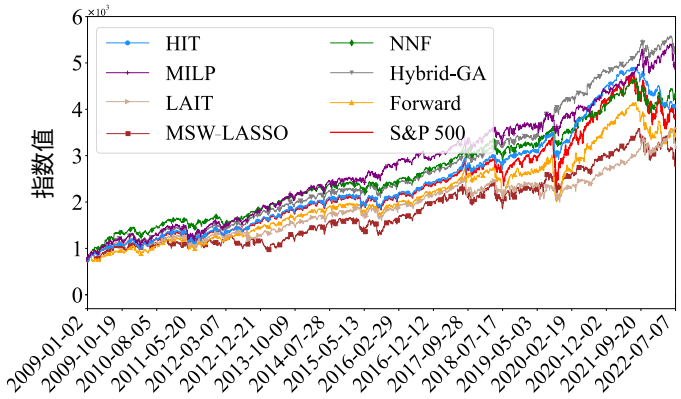


图 6-13 当 $K = 50$ 时，在 S&P 500 数据集上的追踪效果可视化。

6.5.6 与第五章节模型对比分析

第五章和本章中都聚焦于部分指数追踪这一关键任务。基于两个模型的实验成果，本节旨在深入探讨和分析它们的性能差异。通过对比表 6-6的数据可以观察到，在绝大多数的数据集中，本章介绍的 HIT 模型在追踪效能上显著超越了第五章提出的 RLIT 模型。HIT 模型之所以表现更佳，主要归功于其对资产选择与配置这两项任务进行的协同优化设计。这种协同优化策略确保了两个任务的优化效果显著高于 RLIT 模型仅针对单一任务的优化效果。此外，HIT 模型采用蒙特卡洛树搜索（MCTS）作为资产选择的核心机制，这也是其另一个显著优势。MCTS 通过模拟资产选择过程并基于这些模拟的结果来做出选择，理论上，只要模拟次数足够，就能够识别出最优的资产组合。相比之下，RLIT 模型则依赖于神经网络来估

中国知网 https://www.cnki.net

表 6-6 与上一章节所提出的 RLIT 模型在 TEO 指标上的对比。

	Hang Seng	DAX 100	FTSE 100	Nikkei 225	Russell 2000
RLIT	**7.9e-4**	9.6e-4	8.3e-4	**8.2e-4**	3.1e-3
HIT	8.2e-4	**7.1e-4**	**8.2e-4**	8.3e-4	**2.2e-3**

算资产的价值，这种估算方式往往会带来较大的误差。

在时间效率方面，通过比较第五章的图 5-5和本章的图 6-6，可以明显看出 HIT 模型在处理同样任务时的时间消耗显著高于 RLIT 模型。正如前文所述，MCTS 需要大量的时间来进行资产选择的模拟，而 RLIT 模型则是直接通过神经网络评估资产价值，显然前者在效率上的表现不如后者。因此，尽管 HIT 模型在多数数据集上的表现都优于 RLIT 模型，但它在时间消耗上的劣势也不容忽视。

总结来说，尽管 RLIT 模型在执行效率上具有优势，但其追踪效果稍逊一筹；而 HIT 模型虽然在大多数情况下都能提供更好的追踪效果，但相对较高的时间消耗也使其在效率方面不及 RLIT 模型。

6.6 本章小结

本章首次探索了深度分层强化学习来解决部分指数追踪问题。具体而言，本章提出了一个名为 HIT 的分层模型，该模型将 PIT 构建为一个分层马尔可夫决策过程，并使用强化学习进行优化。HIT 包含两级策略：高层策略专注于资产选择，而低层策略处理所选资产之间的资本分配。为了鼓励这两个策略之间的协作学习，本章还提出了一个成本敏感的奖励函数，强调准确复制指数的同时考虑交易成本。通过大量的实验表明，与最先进方法相比，本模型展现了卓越的有效性、效率和可扩展性。

虽然 HIT 模型在资产选择和资本配置方面展现出显著的性能，但高层策略需要不断模拟以决定资产选择，这导致了较长的消耗时间。未来的主要工作将是研究更高效的高层策略来解决资产选择问题，以提高整体模型的效率和响应速度，进一步优化部分指数追踪的性能。通过改进高层策略，期望能够减少模拟的次数和时间，同时保持或提高模型在资产选择精度上的表现。

中国知网 https://www.cnki.net

第七章 总结与展望

7.1 总结

随着中国居民私人财富的快速增长，资产管理（Portfolio Management, PM）正迅速成为一个极为关键的领域。随着大数据、互联网、云计算等信息技术的发展，人工智能逐步应用于各个领域，在资产管理领域也广泛应用智能技术更好地管理资产。资产管理的策略主要分为主动投资和被动投资两大类别。在主动投资策略中，投资者或基金经理积极寻求超越市场平均水平的回报，因此在资产选择和配置时，重点是通过精准的市场分析和资产评估来最大化收益。而在被动投资策略中，目标是实现与整体市场同步的稳定增长。每类策略都包括两个关键任务：资产选择和资产配置。资产选择的目的是挑选合适的资产进行投资，而资产配置任务则着眼于在选定资产间实现最优的资本分配。然而现有的基于资产选择与配置算法在资产管理中的应用仍然存在一定缺陷，主要体现为：

（1）在主动投资中的资产选择任务上，以往的解决方案忽略了资产（例如股票）间的关系，只专注于分析资产的时间模式，将每只资产视作独立个体，却忽略了资产之间的相互关系。虽然有些模型考虑到资产间的联系，但采用了两步处理方式（即先捕捉时间模式再考虑关系依赖），这种方法复杂且效率低下。

（2）在主动投资中的资产配置任务上，尽管深度强化学习因其在序列决策中的能力而成为资产配置领域的一个有前景的方法，但将深度强化学习仍然无法应对金融环境的非平稳性质，其背后的主要原因是由于具有不同投资视野和投资策略的多样化交易者的参与。现有基于深度强化学习的解决方案未能充分考虑到这种非平稳性背后的原因。

（3）在被动投资中的部分指数追踪问题上，主要复杂性在于资产选择阶段，需要从更大的资产池中选出有限数量的资产，也称为“基数约束”，从而实现最小化追踪误差与交易成本。现有的方法通常基于启发式算法，然而这些方法在处理基数约束和交易成本约束时效率低下，尤其是在大规模指数复制方面，无法在有效时间内找到可行解。

（4）在被动投资的部分指数追踪问题上，资产选择任务往往受到基数约束，而资产配置往往受到“非负约束”与“和为一约束”。现在的解决方案往往将资产选择和资产配置分开独立优化来解决不两个任务中的约束问题。然而这些方法却没有提供一个有效的机制来保证这两个任务之间的协同优化，影响了整体的优化效率和效果。

中国知网 https://www.cnki.net

本博士论文针对以上四个问题，提出以下四个解决方案：

（1）本文提出了一种基于关系-时序图卷积网络的股票排名预测模型，旨在充分利用股票间的关系，预测股票的收益率排名。首先，本论文将股票间的关系及其每日特征建模为一个关系-时序图。然后，应用图卷积网络和三种提出的关系感知策略，提取每个节点（股票）的关系-时序特征。最后，将这些特征用于学习排名方式计算排名得分，得分最高的股票代表未来投资收益最高。

（2）本文提出了跨视野交易模型，通过整合具有不同投资视野的多个交易策略来适应市场条件的变化。该模型共分为两步，在第一步中利用专属视野信息学习不同视野的策略。在第二步中，本论文通过学习一个跨视野策略，该策略考虑了专属视野策略所做的投资预决策来做出最终的交易决定。本文还提出一个反事实机制来推理跨视野策略对回报的贡献以更好地优化各个不同视野的策略。

（3）本文提出了一个基于深度强化学习的框架，该框架通过利用资产内部和指数间的关系来选择资产。具体来说，本论文首先将资产选择定义为一个马尔可夫决策过程并以试错的方式进行优化。精心设计的关联性感知深度 Q 网络不仅能够处理基数约束，而且包含一个局部敏感模块和一个指数感知模块，分别考虑资产内部和指数间的关系。本论文还提出了一个考虑成本的奖励函数来评估选择

（4）本文中探索了深度分层强化学习来解决部分指数跟踪问题。具体而言，本论文提出了一个名为分层模型，该模型将部分指数跟踪构建为一个分层马尔可夫决策过程，并使用强化学习进行优化。模型包含两层策略：高层策略专注于资产选择，而低层策略处理所选资产之间的资本分配。为了鼓励这两个策略之间的协作学习，本论文提出了一个成本敏感的奖励函数，在引导两层策略准确复制指数的同时考虑交易成本。

尽管这些研究成果在理论和方法上具有一定的前瞻性和创新性，但仍存在一些不足之处，需要在未来的研究中加以改进和深化：

（1）资源开销大：深度强化学习等先进方法虽然提高了决策的精确性，但也带来了更大的计算和存储需求，例如采用 GPU 来加速模型的训练与优化。这在资源有限的环境中可能是一个制约因素。相对于其他模型，本研究工作的资源消耗更加大。

（2）模型可解释性不足：尽管所提模型在实验中展示了优良的性能，但由于采用了复杂的神经网络结构，这些模型的决策过程缺乏透明度，难以解释其内部逻辑和决策依据。这在金融投资领域尤为重要，因为投资决策需要向监管机构和投资者明确表明其合理性和安全性。

（3）金融知识融入尚浅：当前模型虽然在处理数据和预测方面表现出色，但

中国知网 https://www.cnki.net

在融入深层次的金融专业知识方面仍有待加强。例如，金融市场的微观结构、宏观经济因素与资产价格的关系等复杂因素在模型中尚未得到充分利用。

（4）模型泛化能力差：目前的研究主要集中在特定市场或特定经济周期内的数据分析，模型的泛化能力到其他市场或不同经济周期的有效性尚未充分验证。金融市场波动性大，不同市场和经济周期的特性可能会对模型的性能产生重大影响。

（5）实战投资经验缺乏：尽管理论研究为理解市场行为和指导实际投资提供了重要视角，但论文中的模型和方法大多停留在理论和模拟测试阶段。实际市场中的不确定性和复杂性要求模型在真实的投资环境中得到测试和验证。

7.2 后续工作展望

针对以上提到的不足，未来工作可以从以下四个方向进行深入研究和发展：

（1）优化模型的资源效率：针对资源开销的问题，未来工作将探索高效的模型架构和算法，减少模型的计算复杂度和内存需求。这可能包括研究模型压缩技术、参数剪枝、量化技术等，以及利用硬件加速如 GPU 和 TPU 来提高计算效率。同时，将评估不同模型的性能与资源消耗之间的平衡，寻找最优的配置方案。

（2）增强模型的可解释性：未来研究应着重于提高模型的可解释性，以增强金融模型在实际应用中的可靠性和透明度。通过采用可解释的机器学习技术，例如决策树、规则学习或可解释的神经网络框架，来设计新的模型。

（3）融合更深层次的金融知识：为了提高模型的实用性和精确度，未来工作可以集中在更深层次地融合金融专业知识。将经济周期理论、市场微观结构分析以及风险管理策略等纳入模型构建过程中，开发能够反映复杂市场行为和经济因素影响的高级模型。

（4）验证模型的泛化能力：鉴于金融市场的多样性和动态变化特性，未来的研究需要在不同市场环境和经济周期中测试模型的泛化能力。这包括跨市场测试、跨国测试以及在极端市场条件下的测试。这样的测试将有助于评估模型的稳健性，并指导实际的策略调整以适应不同市场的需求。

（5）结合实际交易进行实证研究：将理论研究与市场实际操作相结合是未来工作的一个重要方向。将研究成果应用于实际交易中，不仅可以验证模型的实际效果，也能根据实际市场反馈进行模型的优化和调整。此外，实证研究还应考虑交易成本、市场冲击和法规限制等实际因素，以确保研究成果的实用性和操作性。

通过对这些方向的深入探索，未来的研究能够更全面地应对当前模型和方法中存在的不足，同时提升金融技术在理论与实践中的应用效果。

中国知网 https://www.cnki.net

参考文献

[1] 李斌, 林彦, 唐闻轩. Ml-tea: 一套基于机器学习和技术分析的量化投资算法 [J]. 系统工程理论与实践, 2017, 37(05): 1089-1100.

[2] 陈鹤丹, 聂挺进, 罗戈, et al. 人工智能在资管行业的应用——基于基本面投资研究 [J]. 新金融, 2024, 39-45.

[3] 李占军. 量化投资在我国投资市场中的应用 [J]. 投资与合作, 2022, 22-24.

[4] 李斌, 屠雪永. 基于机器学习和资产特征的投资组合选择研究 [J]. 系统工程理论与实践, 2024, 44: 338-359.

[5] 黎明辉. 基于 hurst 指数的量化投资策略研究 [J]. 现代信息科技, 2023, 7: 83-87.

[6] 叶望春. 金融科技与银行智能化转型 [J]. 中国金融, 2017, 67-68.

[7] 徐鹏程. 金融科技对我国商业银行盈利能力的影响和对策 [J]. 中国市场, 2024, 03(36-39).

[8] 文学舟, 周志慧, 范歆婷, et al. 金融科技赋能下商业银行小微企业信贷供给精准性的影响因素研究——基于双中介模型 [J]. 金融理论与实践, 2024, 97-109.

[9] 曾福生, 胡特. 金融科技创新的行动者网络与农村普惠金融发展——以蚂蚁金服为例 [J]. 湖南农业大学学报 (社会科学版), 2024, 25: 1-10.

[10] Lee J, Kim R, Yi S, et al. MAPS: multi-agent reinforcement learning-based portfolio management system[C]. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, 4520-4526.

[11] Wang R, Wei H, An B, et al. Commission fee is not enough: A hierarchical reinforced framework for portfolio management[C]. Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, 2021, 626-633.

[12] Zhang L, Aggarwal CC, Qi G. Stock price prediction via discovering multi-frequency trading patterns[C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017, 2017, 2141-2149.

[13] Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction[C]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, 2017, 2627-2633.

[14] Liu G, Mao Y, Sun Q, et al. Multi-scale two-way deep neural network for stock trend prediction[C]. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, 4555-4561.

中国知网 https://www.cnki.net

[15] Feng F, He X, Wang X, et al. Temporal relational ranking for stock prediction[J]. ACM Trans. Inf. Syst., 2019, 37(2): 27:1-27:30.

[16] Sawhney R, Agarwal S, Wadhwa A, et al. Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach[C]. Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, 2021, 497-504.

[17] Wang Z, Huang B, Tu S, et al. Deeptrader: A deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding[C]. Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, 2021, 643-650.

[18] Jiang Z, Xu D, Liang J. A deep reinforcement learning framework for the financial portfolio management problem[J]. CoRR, 2017, abs/1706.10059.

[19] Niu H, Li S, 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, 2022, 1573-1583.

[20] de Amorim TW, Silva JCS, de Almeida Filho AT. Assessing the interactions amongst index tracking model formulations and genetic algorithm approaches with different rebalancing strategies[J]. Soft Comput., 2023.

[21] García F, Guijarro F, Oliver J. Index tracking optimization with cardinality constraint: a performance comparison of genetic algorithms and tabu search heuristics[J]. Neural Comput. Appl., 2018, 30(8): 2625-2641.

[22] Woodside-Oriakhi M, Lucas C, Beasley JE. Heuristic algorithms for the cardinality constrained efficient frontier[J]. Eur. J. Oper. Res., 2011, 213(3): 538-550.

[23] Zheng Y, Chen B, Hospedales TM, et al. Index tracking with cardinality constraints: A stochastic neural networks approach[C]. The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, 2020, 1242-1249.

[24] 姜海燕, 吴长凤. 智能投顾的发展现状及监管建议 [J]. 证券市场导报, 2016, : 4-10.

[25] 孙清云, 赵艳群. 国内传统金融机构智能投顾业务发展探讨——以摩羯智投为例 [J]. 国际金融, 2017, : 34-39.

[26] 康佳伦. 国内商业银行智能投顾平台对比分析 [J]. 北京印刷学院学报, 2020, 28: 23-25.

[27] Murphy JJ. Technical analysis of the financial markets: A comprehensive guide to trading methods and applications[M]. Penguin, 2019.

中国知网 https://www.cnki.net

[28] Wang J, Leu J. Stock market trend prediction using arima-based neural networks[C]. Proceedings of International Conference on Neural Networks (ICNN'96), Washington, DC, USA, June 3-6, 1996, 1996, 2160-2165.

[29] Li B, Zhao P, Hoi SCH, et al. PAMR: passive aggressive mean reversion strategy for portfolio selection[J]. Mach. Learn., 2012, 87(2): 221-258.

[30] Li B, Hoi SCH, Zhao P, et al. Confidence weighted mean reversion strategy for online portfolio selection[J]. ACM Trans. Knowl. Discov. Data, 2013, 7(1): 4:1-4:38.

[31] Alexander C, Dimitriu A. Indexing and statistical arbitrage[J]. The Journal of Portfolio Management, 2005, 31(2): 50-63.

[32] Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory[J]. PloS one, 2017, 12(7): e0180944.

[33] Popoola A, Ahmad K. Testing the suitability of wavelet preprocessing for TSK fuzzy models[C]. IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2006, Vancouver, BC, Canada, July 16-21, 2006, 2006, 1305-1309.

[34] Deng S, Zhang N, Zhang W, et al. Knowledge-driven stock trend prediction and explanation via temporal convolutional network[C]. Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019, 2019, 678-685.

[35] Li Q, Tan J, Wang J, et al. A multimodal event-driven LSTM model for stock prediction using online news[J]. IEEE Trans. Knowl. Data Eng., 2021, 33(10): 3323-3337.

[36] Otter DW, Medina JR, Kalita JK. A survey of the usages of deep learning for natural language processing[J]. IEEE Trans. Neural Networks Learn. Syst., 2021, 32(2): 604-624.

[37] Dabre R, Chu C, Kunchukuttan A. A survey of multilingual neural machine translation[J]. ACM Comput. Surv., 2021, 53(5): 99:1-99:38.

[38] Deng Y, Lei W, Lam W, et al. A survey on proactive dialogue systems: Problems, methods, and prospects[C]. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China, 2023, 6583-6591.

[39] Bengio Y, Simard PY, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Trans. Neural Networks, 1994, 5(2): 157-166.

[40] Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. CoRR, 2018, abs/1803.01271.

中国知网 https://www.cnki.net

[41] Li W, Bao R, Harimoto K, et al. Modeling the stock relation with graph network for overnight stock movement prediction[C]. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, 4541-4547.

[42] Carreira MCS. Exponential kernels with latency in hawkes processes: Applications in finance[J]. CoRR, 2021, abs/2101.06348.

[43] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks[C]. 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, 2017.

[44] Song Y, Suganthan PN, Pedrycz W, et al. Ensemble reinforcement learning: A survey[J]. CoRR, 2023, abs/2303.02618.

[45] Perepu SK, Balaji BS, Tanneru HK, et al. Reinforcement learning based dynamic weighing of ensemble models for time series forecasting[J]. CoRR, 2020, abs/2008.08878.

[46] Chen X, Wang C, Zhou Z, et al. Randomized ensembled double q-learning: Learning fast without a model[C]. 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, 2021.

[47] Carta S, Ferreira A, Podda AS, et al. Multi-dqn: An ensemble of deep q-learning agents for stock market forecasting[J]. Expert Syst. Appl., 2021, 164: 113820.

[48] Elliott DL, Anderson CW. The wisdom of the crowd: Reliable deep reinforcement learning through ensembles of q-functions[J]. IEEE Trans. Neural Networks Learn. Syst., 2023, 34(1): 43-51.

[49] Li G, Dong M, Ming L, et al. Deep reinforcement learning based ensemble model for rumor tracking[J]. Inf. Syst., 2022, 103: 101772.

[50] Saadallah A, Morik K. Online ensemble aggregation using deep reinforcement learning for time series forecasting[C]. 8th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2021, Porto, Portugal, October 6-9, 2021, 2021, 1-8.

[51] Sharma G, Singh A, Jain S. Deepevap: Deep reinforcement learning based ensemble approach for estimating reference evapotranspiration[J]. Appl. Soft Comput., 2022, 125: 109113.

[52] Schubert F, Benjamins C, Döhler S, et al. POLTER: policy trajectory ensemble regularization for unsupervised reinforcement learning[J]. Trans. Mach. Learn. Res., 2023, 2023.

[53] Shen M, How JP. Robust opponent modeling via adversarial ensemble reinforcement learning[C]. Proceedings of the Thirty-First International Conference on Automated Planning and Scheduling, ICAPS 2021, Guangzhou, China (virtual), August 2-13, 2021, 2021, 578-587.

中国知网 https://www.cnki.net

[54] Qin Y, Wang Z, Chen C. HRL2E: hierarchical reinforcement learning with low-level ensemble[C]. International Joint Conference on Neural Networks, IJCNN 2022, Padua, Italy, July 18-23, 2022, 2022, 1-7.

[55] Ferreira PVR, Paffenroth RC, Wyglinski AM, et al. Multiobjective reinforcement learning for cognitive satellite communications using deep neural network ensembles[J]. IEEE J. Sel. Areas Commun., 2018, 36(5): 1030-1041.

[56] Cully A, Demiris Y. Quality and diversity optimization: A unifying modular framework[J]. IEEE Trans. Evol. Comput., 2018, 22(2): 245-259.

[57] Stulp F, Schaal S. Hierarchical reinforcement learning with movement primitives[C]. 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011), Bled, Slovenia, October 26-28, 2011, 2011, 231-238.

[58] Sutton RS, Precup D, Singh S. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning[J]. Artif. Intell., 1999, 112(1-2): 181-211.

[59] Parr R, Russell S. Reinforcement learning with hierarchies of machines[C]. Advances in Neural Information Processing Systems 10, [NIPS Conference, Denver, Colorado, USA, 1997], 1997, 1043-1049.

[60] Dietterich TG. Hierarchical reinforcement learning with the MAXQ value function decomposition[J]. J. Artif. Intell. Res., 2000, 13: 227-303.

[61] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, 2016, 785-794.

[62] Medsker L, Jain LC. Recurrent neural networks: Design and applications[M]. CRC Press, 1999.

[63] Hsu Y, Tsai Y, Li C. Fingat: Financial graph attention networks for recommending top-k profitable stocks[J]. IEEE Trans. Knowl. Data Eng., 2021.

[64] Deng Y, Bao F, Kong Y, et al. Deep direct reinforcement learning for financial signal representation and trading[J]. IEEE Trans. Neural Networks Learn. Syst., 2017, 28(3): 653-664.

[65] Liu Y, Liu Q, Zhao H, et al. Adaptive quantitative trading: An imitative deep reinforcement learning approach[C]. The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, 2020, 2128-2135.

中国知网 https://www.cnki.net

[66] Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks?[C]. 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, 2019.

[67] Rhee S, Seo S, Kim S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification[C]. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, 2018, 3527-3534.

[68] Schlichtkrull MS, Kipf TN, Bloem P, et al. Modeling relational data with graph convolutional networks[C]. The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings, 2018, 593-607.

[69] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, 2018, 7444-7452.

[70] Chen J, Lécué F, Pan JZ, et al. Learning from ontology streams with semantic concept drift[C]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, 2017, 957-963.

[71] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE Trans. Neural Networks Learn. Syst., 2021, 32(1): 4-24.

[72] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs[C]. Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, 2016, 2014-2023.

[73] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[C]. 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.

[74] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[C]. Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain, 2016, 3837-3845.

中国知网 https://www.cnki.net

[75] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, 5998-6008.

[76] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks[C]. 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, 2018.

[77] Lea C, Vidal R, Reiter A, et al. Temporal convolutional networks: A unified approach to action segmentation[C]. Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III, 2016, 47-54.

[78] van den Oord A, Dieleman S, Zen H, et al. Wavenet: A generative model for raw audio[C]. The 9th ISCA Speech Synthesis Workshop, Sunnyvale, CA, USA, 13-15 September 2016, 2016, 125.

[79] Waibel AH, Hanazawa T, Hinton GE, et al. Phoneme recognition using time-delay neural networks[J]. IEEE Trans. Acoust. Speech Signal Process., 1989, 37(3): 328-339.

[80] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, 2016, 770-778.

[81] Ba LJ, Kiros JR, Hinton GE. Layer normalization[J]. CoRR, 2016, abs/1607.06450.

[82] Srivastava N, Hinton GE, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. J. Mach. Learn. Res., 2014, 15(1): 1929-1958.

[83] Xu K, Zhang Y, Ye D, et al. Relation-aware transformer for portfolio policy learning[C]. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, 4647-4653.

[84] Zhang Y, Zhao P, Wu Q, et al. Cost-sensitive portfolio selection via deep reinforcement learning[J]. IEEE Trans. Knowl. Data Eng., 2022, 34(1): 236-248.

[85] Vrandecic D, Krötzsch M. Wikidata: a free collaborative knowledgebase[J]. Commun. ACM, 2014, 57(10): 78-85.

[86] Dixon M, Klabjan D, Bang JH. Classification-based financial markets prediction using deep neural networks[J]. Algorithmic Finance, 2017, 6(3-4): 67-77.

[87] Feng F, Chen H, He X, et al. Enhancing stock movement prediction with adversarial training[C]. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, 2019, 5843-5849.

[88] Rey D, Neuhäuser M. Wilcoxon-signed-rank test[M]. Springer, 2011, 1658-1659.

中国知网 https://www.cnki.net

[89] He Y, Chu X, Wang Y. Neighbor profile: Bagging nearest neighbors for unsupervised time series mining[C]. 36th IEEE International Conference on Data Engineering, ICDE 2020, 2020, 373-384.

[90] Yürüten O, Zhang J, Pu P. Decomposing activities of daily living to discover routine clusters[C]. Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, 1348-1354.

[91] Li B, Hoi SCH. Online portfolio selection: A survey[J]. ACM Comput. Surv., 2014, 46(3): 35:1-35:36.

[92] Li B, Hoi SCH. On-line portfolio selection with moving average reversion[C]. Proceedings of the 29th International Conference on Machine Learning, ICML 2012, 2012.

[93] Neely CJ, Rapach DE, Tu J, et al. Forecasting the equity risk premium: The role of technical indicators[J]. Manag. Sci., 2014, 60(7): 1772-1791.

[94] Ye Y, Pei H, Wang B, et al. Reinforcement-learning based portfolio management with augmented asset movement prediction states[C]. The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, 2020, 1112-1119.

[95] Zhang Y, Zhao P, Wu Q, et al. Cost-sensitive portfolio selection via deep reinforcement learning[J]. IEEE Trans. Knowl. Data Eng., 2022, 34(1): 236-248.

[96] Kristoufek L. Fractal markets hypothesis and the global financial crisis: Wavelet power evidence[J]. Sci. Rep., 2013, 3: 2857.

[97] Peters EE. Fractal market analysis: Applying chaos theory to investment and economics[M]. Wiley, 1994.

[98] Nuij W, Milea V, Hogenboom F, et al. An automated framework for incorporating news into stock trading strategies[J]. IEEE Trans. Knowl. Data Eng., 2014, 26(4): 823-835.

[99] Li Z, Tam VWL. Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes[C]. 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017, 2017, 1-8.

[100] Saadallah A, Tavakol M, Morik K. An actor-critic ensemble aggregation model for time-series forecasting[C]. 37th IEEE International Conference on Data Engineering, ICDE 2021, 2021, 2255-2260.

[101] Liu P, Zhang Y, Bao F, et al. Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading[J]. Appl. Intell., 2023, 53(2): 1683-1706.

中国知网 https://www.cnki.net

[102] Wu X, Chen H, Wang J, et al. Adaptive stock trading strategies with deep reinforcement learning methods[J]. Inf. Sci., 2020, 538: 142-158.

[103] Helmbold DP, Schapire RE, Singer Y, et al. On-line portfolio selection using multiplicative updates[J]. Math. Financ., 1998, 8(4): 325-347.

[104] Li Y, Zheng X, Chen C, et al. Exponential gradient with momentum for online portfolio selection[J]. Expert Syst. Appl., 2022, 187: 115889.

[105] Agarwal A, Hazan E, Kale S, et al. Algorithms for portfolio management based on the newton method[C]. Machine Learning, Proceedings of the Twenty-Third International Conference, ICML 2006, 2006, 9-16.

[106] Borodin A, El-Yaniv R, Gogan V. On the competitive theory and practice of portfolio selection (extended abstract)[C]. LATIN 2000: Theoretical Informatics, 4th Latin American Symposium, Proceedings, 2000, 173-196.

[107] Cover TM, Gluss DH. On-line portfolio selection using multiplicative updates[J]. Adv. Appl. Math., 1986, 7(2): 170-181.

[108] Borodin A, El-Yaniv R, Gogan V. Can we learn to beat the best stock[C]. Advances in Neural Information Processing Systems 16 Neural Information Processing Systems, NIPS 2003, 2003, 345-352.

[109] Blum A, Kalai A. Universal portfolios with and without transaction costs[J]. Mach. Learn., 1999, 35(3): 193-205.

[110] Huang D, Zhou J, Li B, et al. Robust median reversion strategy for online portfolio selection[J]. IEEE Trans. Knowl. Data Eng., 2016, 28(9): 2480-2493.

[111] Neuneier R. Optimal asset allocation using adaptive dynamic programming[C]. Advances in Neural Information Processing Systems 8, NIPS 1995, 1995, 952-958.

[112] Jang J, Seong N. Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory[J]. Expert Syst. Appl., 2023, 218: 119556.

[113] Zhang R, Ren X, Gu F, et al. MDAEN: multi-dimensional attention-based ensemble network in deep reinforcement learning framework for portfolio management[C]. International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2022, 2022, 143-151.

[114] Lien Y, Li Y, Wang Y. Contrastive learning and reward smoothing for deep portfolio management[C]. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 2023, 3966-3974.

中国知网 https://www.cnki.net

[115] Ma C, Zhang J, Li Z, et al. Multi-agent deep reinforcement learning algorithm with trend consistency regularization for portfolio management[J]. Neural Comput. Appl., 2023, 35(9): 6589-6601.

[116] Liu X, Yang H, Gao J, et al. Finrl: deep reinforcement learning framework to automate trading in quantitative finance[C]. ICAIF'21: 2nd ACM International Conference on AI in Finance, 2021, 1:1-1:9.

[117] Saad M, Bernstein A, Böhlen MH, et al. Single point incremental fourier transform on 2d data streams[C]. 37th IEEE International Conference on Data Engineering, ICDE 2021, 2021, 852-863.

[118] Samiee K, Kovács P, Gabbouj M. Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform[J]. IEEE Trans. Biomed. Eng., 2015, 62(2): 541-552.

[119] Xu M, Han M, Lin H. Wavelet-denoising multiple echo state networks for multivariate time series prediction[J]. Inf. Sci., 2018, 465: 439-458.

[120] Fang Y, Qin Y, Luo H, et al. When spatio-temporal meet wavelets: Disentangled traffic forecasting via efficient spectral graph attention networks[C]. 39th IEEE International Conference on Data Engineering, ICDE 2023, 2023, 517-529.

[121] Lahmiri S. Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks[J]. J. King Saud Univ. Comput. Inf. Sci., 2014, 26(2): 218-227.

[122] Williams RJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Mach. Learn., 1992, 8: 229-256.

[123] Watkins CJCH, Dayan P. Technical note q-learning[J]. Mach. Learn., 1992, 8: 279-292.

[124] Weaver L, Tao N. The optimal reward baseline for gradient-based reinforcement learning[C]. UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, 2001, 538-545.

[125] Sutton RS. Learning to predict by the methods of temporal differences[J]. Mach. Learn., 1988, 3: 9-44.

[126] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]. 4th International Conference on Learning Representations, ICLR 2016, Conference Track Proceedings, 2016.

[127] Wolpert DH, Tumer K. Optimal payoff functions for members of collectives[J]. Adv. Complex Syst., 2001, 4(2-3): 265-280.

中国知网 https://www.cnki.net

[128] Foerster JN, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients[C]. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), 2018, 2974-2982.

[129] Sutton RS, McAllester DA, Singh SP, et al. Policy gradient methods for reinforcement learning with function approximation[C]. Advances in Neural Information Processing Systems 12, NIPS 1999, 1999, 1057-1063.

[130] Konda VR, Tsitsiklis JN. Actor-critic algorithms[C]. Advances in Neural Information Processing Systems 12, [NIPS Conference 1999], 1999, 1008-1014.

[131] Lillicrap TP, Hunt JJ, Pritzel A, et al. Continuous control with deep reinforcement learning[C]. 4th International Conference on Learning Representations, ICLR 2016, 2016.

[132] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[J]. CoRR, 2017, abs/1707.06347.

[133] Benidis K, Feng Y, Palomar DP. Sparse portfolios for high-dimensional financial index tracking[J]. IEEE Trans. Signal Process., 2018, 66(1): 155-170.

[134] Strub O, Baumann P. Optimal construction and rebalancing of index-tracking portfolios[J]. Eur. J. Oper. Res., 2018, 264(1): 370-387.

[135] Karlow D. Comparison and development of methods for index tracking[D]. Frankfurt School of Finance & Management (Frankfurt/Main), 2013, 1-177.

[136] Kwak Y, Song J, Lee H. Neural network with fixed noise for index-tracking portfolio optimization[J]. Expert Syst. Appl., 2021, 183: 115298.

[137] Dose C, Cincotti S. Clustering of financial time series with application to index and enhanced index tracking portfolio[J]. Physica A: Statistical Mechanics and its Applications, 2005, 355(2): 145-151.

[138] Corielli F, Marcellino M. Factor based index tracking[J]. Journal of Banking & Finance, 2006, 30(8): 2215-2233.

[139] Zhang Z, Zhao Z. A deep learning-aided approach to portfolio design for financial index tracking[C]. 54th Asilomar Conference on Signals, Systems, and Computers, ACSCC 2020, 2020, 200-204.

[140] Beasley JE, Meade N, Chang T. An evolutionary heuristic for the index tracking problem[J]. Eur. J. Oper. Res., 2003, 148(3): 621-643.

[141] Takeda A, Niranjan M, Gotoh J, et al. Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios[J]. Comput. Manag. Sci., 2013, 10(1): 21-49.

中国知网 https://www.cnki.net

[142] Fastrich B, Paterlini S, Winker P. Cardinality versus q-norm constraints for index tracking[J]. Quantitative Finance, 2014, 14(11): 2019-2032.

[143] Shi F, Shu L, Luo Y, et al. High-dimensional sparse index tracking based on a multi-step convex optimization approach[J]. Quantitative Finance, 2023, 23(9): 1361-1372.

[144] Canakgoz NA, Beasley JE. Mixed-integer programming approaches for index tracking and enhanced indexation[J]. Eur. J. Oper. Res., 2009, 196(1): 384-399.

[145] Scozzari A, Tardella F, Paterlini S, et al. Exact and heuristic approaches for the index tracking problem with UCITS constraints[J]. Ann. Oper. Res., 2013, 205(1): 235-250.

[146] Li XP, Shi Z, Leung C, et al. Sparse index tracking with k-sparsity or deviation constraint via norm minimization[J]. IEEE Trans. Neural Networks Learn. Syst., 2023, 34(12): 10930-10943.

[147] Anis HT, Costa G, Kwon RH. Risk-allocation-based index tracking[J]. Comput. Oper. Res., 2023, 106219.

[148] Jeong G, Kim HY. Improving financial trading decisions using deep q-learning: Predicting the number of shares, action strategies, and transfer learning[J]. Expert Syst. Appl., 2019, 117: 125-138.

[149] Xu F, Lu Z, Xu Z. An efficient optimization approach for a cardinality-constrained index tracking problem[J]. Optimization Methods and Software, 2016, 31(2): 258-271.

[150] He J, Chen J, He X, et al. Deep reinforcement learning with a natural language action space[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Volume 1: Long Papers, 2016.

[151] He J, Ostendorf M, He X, et al. Deep reinforcement learning with a combinatorial action space for predicting popular reddit threads[C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, 2016, 1838-1848.

[152] Andoni A, Indyk P, Laarhoven T, et al. Practical and optimal LSH for angular distance[C]. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, 2015, 1225-1233.

[153] Gao S, Wang Y, Yang X. Stockformer: Learning hybrid trading machines with predictive coding[C]. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 2023, 4766-4774.

[154] Li Q, Sun L, Bao L. Enhanced index tracking based on multi-objective immune algorithm[J]. Expert Syst. Appl., 2011, 38(5): 6101-6106.

中国知网 https://www.cnki.net

[155] Xu F, Lu Z, Xu Z. An efficient optimization approach for a cardinality-constrained index tracking problem[J]. Optim. Methods Softw., 2016, 31(2): 258-271.

[156] Filippi C, Guastaroba G, Speranza MG. A heuristic framework for the bi-objective enhanced index tracking problem[J]. Omega, 2016, 65: 122-137.

[157] Barber BM, Odean T. Trading is hazardous to your wealth: The common stock investment performance of individual investors[J]. The Journal of Finance, 2000, 55(5): 773-806.

[158] Benidis K, Feng Y, Palomar DP. Optimization methods for financial index tracking: From theory to practice[J]. Found. Trends Optim., 2018, 3(3): 171-279.

[159] Focardi SM, 3 FJF. A methodology for index tracking based on time-series clustering[J]. Quantitative Finance, 2004, 4(4): 417-425.

[160] Ni H, Wang Y. Stock index tracking by pareto efficient genetic algorithm[J]. Appl. Soft Comput., 2013, 13(12): 4519-4535.

[161] Nesterov YE, Nemirovskii A. Interior-point polynomial algorithms in convex programming[M]. SIAM, 1994.

[162] Zheng Y, Li Y, Xu Q, et al. Index tracking with differentiate asset selection[C]. ICAIF '20: The First ACM International Conference on AI in Finance, 2020, 12:1-12:8.

[163] Tong H, Liu Y, Liu L, et al. LSTM-DGMDH: high-dimensional index tracking based on LSTM and adaptive deep evolutionary GMDH neural network[J]. IEEE Access, 2023, 11: 115654-115667.

[164] Suri K, Shi XQ, Plataniotis KN, et al. Trader: Practical deep hierarchical reinforcement learning for trade execution[J]. CoRR, 2021, abs/2104.00620.

[165] Han W, Zhang B, Xie Q, et al. Select and trade: Towards unified pair trading with hierarchical reinforcement learning[C]. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, 2023, 4123-4134.

[166] Millea A. Hierarchical model-based deep reinforcement learning for single-asset trading[J]. Analytics, 2023, 2(3): 560-576.

[167] Kemmerling M, Schmitt RH. Beyond games: A systematic review of neural monte carlo tree search applications[J]. CoRR, 2023, abs/2303.08060.

[168] Kocsis L, Szepesvári C. Bandit based monte-carlo planning[C]. Machine Learning: ECML 2006, 17th European Conference on Machine Learning, Proceedings, 2006, 282-293.

中国知网 https://www.cnki.net

[169] Mirsoleimani SA, Plaat A, van den Herik HJ, et al. An analysis of virtual loss in parallel MCTS[C]. Proceedings of the 9th International Conference on Agents and Artificial Intelligence, ICAART 2017, Volume 2, 2017, 648-652.

[170] Raffin A, Hill A, Gleave A, et al. Stable-baselines3: Reliable reinforcement learning implementations[J]. J. Mach. Learn. Res., 2021, 22: 268:1-268:8.

中国知网 https://www.cnki.net