“知识”嵌入型深度强化学习在多元资产配置中的应用

——“学海拾珠”系列之二百三十

报告日期：2025-04-02

分析师：严佳炜
执业证书号：S0010520070001
邮箱：yanjw@hazq.com

分析师：钱静闲
执业证书号：S0010522090002
邮箱：qianjx@hazq.com

相关报告
1.《分解动量：被遗忘的成分 HTP——“学海拾珠”系列之二百二十九》
2.《基于树模型的有效前沿扩展——“学海拾珠”系列之二百二十八》
3.《使用深度强化学习解决高维多期环境下的组合配置——“学海拾珠”系列之二百二十七》
4.《风险规避型强化学习模型在投资组合优化中的应用——“学海拾珠”系列之二百二十六》
5.《贝塔异象的波动性之谜——“学海拾珠”系列之二百二十五》
6.《ETF 的资产配置与再平衡：样本协方差对比 EWMA 与 GARCH 模型——“学海拾珠”系列之二百二十四》
7.《市场对投资者情绪的反应——“学海拾珠”系列之二百二十三》

主要观点：

本篇是“学海拾珠”系列第二百三十篇，过往强化学习模型通常从随机初始化的权重开始训练，并且对其生成的输出缺乏直观的可解释性。文献提出了一种新颖的方法，旨在最大化长期风险调整后的投资回报，同时保持模型输出的可解释性。采用成熟的基于规则的策略，通过模仿学习生成神经网络模拟模型，从而传递专家知识。这些“导师”模型随后通过结合执行-评估模型（Soft Actor-Critic, SAC）和深度确定性策略梯度（Deep Deterministic Policy Gradient, DDPG）的混合强化学习算法进行增强，目标是创建出表现优于其导师的“学生”模型。回到国内市场，该类基于规则的强化学习方法可解释性更强，在资产配置领域或能发挥较大的作用。

● 导师-学生模型

“导师”模型基于 Keller 和 Keuning 提出的动态资产配置模型，其核心思想倡导一种更稳健且潜在收益更高的投资方法：采取与资产动量方向相反的交易策略。具体而言，当作为诊断工具的指示性资产显示出动量减弱迹象时，该策略建议从高风险资产（主要是股票类 ETF）中撤资，转而增持安全资产（以债券/国债类 ETF 为主）。

“学生”模型继承自上述基于规则（rule-based）的资产配置模型，同时引入深度强化学习（RL）算法进行精细化调优，在 DDPG 框架基础上引入 SAC 特性（如双评估减少偏差、灵活调整决策-评估模型更新频率），形成混合扩展模型。

● 实证结果

在一个近 40 年的价格数据集上，对美国股票、债券、美国国债、大宗商品及其杠杆等价物等广泛的资产类别进行模拟，实证验证了这一策略的有效性。新模型的测试集中，夏普比率提升了高达 39.70%，索提诺比率提升了高达 47.07%。这表明，将成熟策略与先进强化学习相结合，在资产管理领域的潜力。

● 风险提示

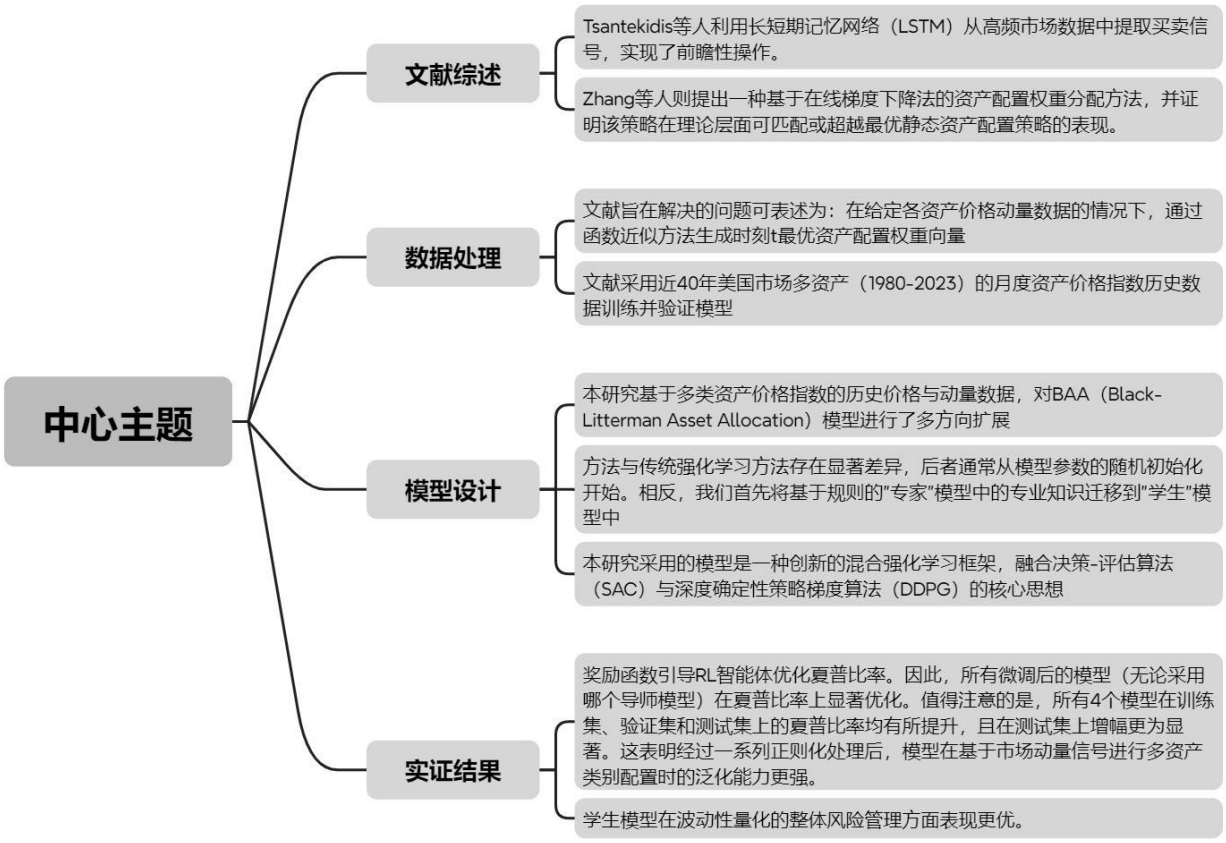
文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

正文目录

图表目录

1 引言

图表 1 文章框架



资料来源：华安证券研究所整理

随着机器学习（尤其是深度学习）在解决复杂现实问题（如机器翻译和图像分类）中展现出的卓越性能，各类创新方法正被逐步融入精密投资策略的构建中。这些计算技术已广泛应用于投资的多个维度，涵盖市场信号处理、动态资产配置、价格预测及金融情绪分析等领域。Tsantekidis 等人利用长短期记忆网络（LSTM）从高频市场数据中提取买卖信号，实现了前瞻性操作。该研究证实，相较于支持向量机（SVM）和多层感知机（MLP），LSTM 模型在信号预测方面具有更优表现。Zhang 等人则提出一种基于在线梯度下降法的资产配置权重分配方法，并证明该策略在理论层面可匹配或超越最优静态资产配置策略的表现。

然而，现有投资组合优化研究中，尽管已整合多种机器学习方法，但大量研究仍难以证明策略的持续长期有效性。此类研究的一个共性特征是依赖时间跨度有限的数据集，通常不足五年：例如 1 年、2 年，或 3-4 年。尽管这些工作有效验证了机器学习模型在投资组合管理中的短期价值，但缺乏证据表明此类策略能在重大金融动荡期（如 1997 年亚洲金融危机、2008 年次贷危机及 2021 年新冠疫情冲击）中保持性能稳定性。这种时间维度的局限性，可能难以充分建立个人投资者的信心——尤其对于该技术熟悉度较低的群体，从而潜在阻碍由智能投顾推荐的投资管理技术的广泛采纳。

此外，部分研究在构建投资组合选择模型时，仅聚焦于特定资产类别子集（通常限于某类股票），以此定义可选资产池。尽管这些研究在限定资产范围内取得了良好

投资表现，但此类狭窄的资产选择范围可能难以确保投资组合管理的稳定性——尤其是在市场条件预示或已导致原有利好趋势逆转时。再者，现代智能投顾方法的不透明性，常因其算法缺乏直观可解释性而阻碍个人投资者与机构决策者的信任。

为应对上述挑战，文献提出一种融合改进的透明化研究框架，该框架继承并优化了现有的基于规则（rule-based）的资产配置模型，同时引入深度强化学习（RL）算法进行精细化调优。本研究采用的规则模型包括进攻性资产配置（Bold Asset Allocation, BAA）和防御性资产配置（Defensive Asset Allocation, DAA）。此类模型以战略性调整著称，即当资产动量（通过比较当前价格与历史移动平均价及短期历史收益率生成信号）显示风险上升时，主动撤离风险市场。通过借鉴现有模型在可信度与可解释性方面的优势，本研究提出的方法进一步利用先进深度强化学习算法，显著提升了所选策略在投资组合中的执行效能。

文献主要贡献如下：

- **提出高性能的长期动态投资组合选择模型**

创新性地构建了现有规则型投资组合选择策略的仿效模型框架，随后通过深度强化学习（RL）混合扩展优化其可信度与可解释性。这是首次实现从成熟规则型投资组合选择模型的知识迁移，并集成深度强化学习先进扩展进行性能增强的研究。

- **提出 SAC-DDPG 混合 RL 模型的新型扩展**

本在 DDPG 框架基础上引入 SAC 特性（如双评估减少偏差、灵活调整决策-评估模型更新频率），形成混合扩展模型。同时创新性地加入高斯噪声注入器和引导噪声注入网络，显著区别于现有模型。

- **跨广泛资产类别与海量数据集验证模型性能**

为验证模型有效性，研究采用涵盖股票、债券、商品及多种 ETF 的近 40 年综合数据集进行测试。广泛的验证结果凸显了该模型在提供长期稳健风险调整后收益方面的能力，并证明其适用于多样化的金融工具。

本研究应用所提出的模型，引入了多种投资策略，这些策略整合了广泛的资产类别，旨在满足具有不同需求投资者的不同风险承受能力。因此，本研究旨在建立一个稳健可靠的投资组合优化框架，为投资管理贡献一种稳健且有条理的方法。

2 背景

2.1 采用基于规则的资产配置模型

基于 Keller 和 Keuning 提出的动态资产配置模型，并结合后续扩展研究（如 Bold 资产配置策略）展开论述。其核心思想倡导一种更稳健且潜在收益更高的投资方法：采取与资产动量方向相反的交易策略。具体而言，当作为诊断工具的指示性资产显示出动量减弱迹象时，该策略建议从高风险资产（主要是股票类 ETF）中撤资，转而增持安全资产（以债券/国债类 ETF 为主）。

为验证这一策略的有效性，作者进行了双重实证检验：首先采用 1926 至 1970 年的历史数据进行回溯测试，随后进一步在 1970 至 2018 年 3 月的数据集上检验其稳健性。研究结果显示，代表性模型 VAA-G12 取得了 12.7%的复合年化增长率，同时将风险控制在月度最大回撤率 8.0%以内。该策略通过动量反转机制，在控制下行风险的同时实现了长期收益增强，展现了量化投资框架下风险收益平衡的新范式。

2.2 成熟的强化学习算法

强化学习（RL）是本研究改进现有规则型投资组合选择模型的核心方法。在 RL

框架中，智能体通过反复试错迭代优化动作序列，使累积奖励最大化，每次动作执行后获得的奖励值构成反馈信号。该方法天然具备探索新动作的机制以提升学习效率，在“探索-利用”权衡中寻求最优平衡。

标准 RL 算法的核心数学符号体系包含：

- 状态集(S)：智能体可能遭遇的所有情境或环境配置的集合，反映当前环境特征
- 动作集(A)：智能体在任意时刻可采取的所有环境干预选项
- 策略(π)：决定智能体决策的映射规则，建立状态到动作的映射关系
- 奖励函数(R)：定义智能体在特定状态执行动作后获得的即时数值反馈
- 值函数(V)：评估在给定策略下，智能体从特定状态出发可获得的预期累积奖励
- Q 函数(Q)：量化在特定状态下采取特定动作后，遵循给定策略能获得的预期累积奖励

后续章节将系统阐述在 RL 领域取得显著成功的经典算法，这些算法构成了解决动态投资组合选择问题的理论基石。通过结合马尔可夫决策过程（MDP）框架，本研究将展示 RL 如何革新传统投资范式，在复杂市场环境中实现动态最优配置。

2.3.1 SAC 算法

柔性决策-评估算法（SAC 算法）是一种针对高维连续动作空间问题的强化学习（RL）方法。与传统仅关注最大化预期收益的策略优化方法不同，SAC 在追求预期收益最大化的同时，还鼓励策略保持适度的熵值，从而在探索新动作与利用已知有效动作之间建立平衡。

为实现这一目标，SAC 引入了最大熵目标函数。该策略（由参数ϕ参数化）的训练目标不仅是最大化预期收益，还需最大化策略的熵值。其数学表达式为：

$$J(\pi_\phi) = \mathbb{E}_{s\sim\rho_\pi, a\sim\pi_\phi}[r(s,a) + \alpha H(\pi_\phi(s))] \quad (1)$$

其中：

- $J(\pi_\phi)$ 表示策略 π_ϕ 的预期收益
- s 表示状态，$s\sim\rho\pi$ 表示状态服从策略 π 下的状态分布
- $a\sim\pi_\phi(s)$ 表示在状态 s 下从策略采样得到的动作
- r(s,a) 表示在状态 s 执行动作 a 后获得的即时奖励
- $H(\pi_\phi(s))$ 表示策略 π_ϕ 在状态 s 下的熵值，衡量动作选择的随机性
- α 为温度系数，用于调节预期收益与熵值之间的权衡，较大的 α 值鼓励更多探索行为

为计算策略梯度，SAC 采用离线学习策略，使用类似 DDPG 的回放缓冲区。通过回放缓冲区中的转移样本估计由参数 θ 参数化的 Q 函数 $Q_\theta(s,a)$，并利用该 Q 函数通过微分目标函数 $J(\pi_\phi)$ 来更新策略 π_ϕ：

$$\nabla_\phi J(\pi_\phi) = \mathbb{E}_{s\sim\rho_\pi, a\sim\pi_\phi}[\nabla_\phi \log \pi_\phi(a|s)(Q_\theta(s,a) - \alpha \log \pi_\phi(a|s))] \quad (2)$$

其中：

- $\nabla_\phi J(\pi_\phi)$ 表示关于策略参数 ϕ 的预期收益梯度
- $Q_\theta(s,a)$ 为动作-值函数，表示在状态 s 执行动作 a 并遵循策略 π 的预期收益
- $\nabla_\phi \log\pi_\phi(a|s)$ 表示策略 π_ϕ 输出动作 a 的对数概率关于策略参数 ϕ 的梯度

SAC 算法通过高效学习策略参数 ϕ、Q 函数参数 θ 和自适应温度参数 α 来管理策略熵。该算法具有以下特性：

稳定性：通过熵正则化增强策略鲁棒性

样本效率：离线学习策略提升数据利用率

自适应调节：自动平衡探索与利用

双 Q 函数机制：采用双胞胎 Q 网络缓解过估计偏差

这些特性使 SAC 在复杂环境中表现出色，对超参数敏感性较低，有效平衡了环境探索与策略优化。凭借其优异的实证性能、鲁棒性和效率，SAC 在连续控制任务中引起了广泛关注，成为强化学习领域的基准算法之一。

2.3.2 DPG 和 DDPG 算法

确定性策略梯度（DPG）算法专门解决连续动作空间中的强化学习问题。与传统技术（如 Q-learning）不同，后者依赖随机策略且容易引发高方差和训练不稳定，DPG 采用确定性策略。该策略通过函数 $\mu_\theta(s)$将状态 s 直接映射到动作 a，无需动作采样，从而显著降低更新方差。其目标是通过优化策略参数 θ，使期望回报 J(μ_θ)最大化:

$$J(\mu_\theta) = \mathbb{E}_{s\sim\rho_\mu}\left[r(s,\mu_\theta(s))\right] \tag{3}$$

其中：

- $J(\mu_\theta)$表示策略 μ_θ 的期望回报;
- s 为状态，$s\sim\rho_\mu$ 表示状态服从策略 μ_θ 下的状态分布;
- $\mu_\theta(s)$为确定性策略在状态 s 下选择的动作;
- $r(s, \mu_\theta(s))$为在状态 s 执行动作 $\mu_\theta(s)$后获得的奖励。

确定性策略梯度定理给出了期望回报关于策略参数 θ 的梯度表达式∇_θJ(μ_θ)，该梯度指导参数 θ 的更新方向以提高期望回报:

$$\nabla_\theta J(\mu_\theta) = \mathbb{E}_{s\sim\rho_\mu}\left[\nabla_\theta \mu_\theta(s)\nabla_a Q^\mu(s,a)\Big|_{a=\mu_\theta(s)}\right] \tag{4}$$

其中：

- $\nabla_\theta J(\mu_\theta)$为期望回报对策略参数的梯度;
- $\nabla_\theta \mu_\theta(s)$为策略输出动作对参数的梯度;
- $\nabla_a Q^\mu(s,a)$为动作价值函数对动作的梯度;
- $Q^\mu(s,a)$为确定性策略 μ 的动作价值函数，表示在状态 s 执行动作 a 后遵循策略 μ 的期望回报。

DPG 具有以下优势，能有效训练强化学习智能体:

高效处理连续动作空间：与 DQN 等针对离散动作的算法不同，DPG 直接处理连续动作，避免了动作空间离散化带来的复杂度爆炸问题。

离线学习（Off-policy Learning）：利用历史经验优化当前策略，提升数据利用率。

噪声注入探索（Noisy Execution）：在动作输出中添加采样噪声，确保环境探索的充分性。

增量式无模型学习：通过随机梯度更新逐步优化策略，适用于未知环境且计算资源受限的场景。

深度确定性策略梯度（DDPG）作为 DPG 的扩展，通过引入经验回放池和目标网络进一步提升学习稳定性和效率。经验回放池存储历史经验元组(state,action.reward,nextstate)，通过小批量采样打破数据的时间相关性。目标网络缓慢更新目标网络参数，稳定训练过程；在执行者网络输出动作时添加噪声，平衡探索与利用。

3 方法论

3.1 问题定义

本研究旨在解决的问题可表述为：在给定各资产价格动量数据的情况下，通过函数近似方法生成时刻 t 最优资产配置权重向量。该函数以表征时刻 t 资产价格动量的特征矩阵为输入，理想情况下应能在每个时刻通过投资组合管理持续产生收益，从而自然实现时间窗口 F 内累积收益的最大化。其数学形式化表达如下：

$$\operatorname{argmax} f \sum_{t=1}^{T} \mathbf{w_t}^{\top} \mathbf{R}_t \tag{5}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} w_{t,i} = 1 \tag{6}$$

$$0 \le w_{t,i} \le 1, ; \forall i = 1, 2, \ldots, N, \forall t = 1, 2, \ldots, T \tag{7}$$

$$f(\mathbf{m}_t, \mathbf{R}t)^{\top} \mathbf{R}t \ge r\min, ; r\min \ge 0 \tag{8}$$

$$\mathbf{m}_t \in \mathbf{M} \tag{9}$$

$$\mathbf{w}_t = f(\mathbf{m}_t), \forall t = 1, 2, \ldots, T \tag{10}$$

其中$w_t^T R_t$表示 t 时刻投资组合的预期收益，约束要求各资产权重和为 1，且$m_t \in M$为 t 时刻的资产动量特征矩阵。

但在现实场景中，要求每个时刻收益均超过特定正阈值几乎不可能实现，因此累积收益更可能成为核心目标。然而单纯追求累积收益可能忽视投资组合价值的随机性下跌风险。为平衡收益与风险管理，研究引入风险调整后的收益指标，基于夏普比率和索提诺比率对目标函数进行改进：

$$\operatorname{argmax} f : \left(w_1 \cdot \frac{\frac{1}{T}\sum_{t=1}^{T} \alpha_t}{\sigma(\alpha_t)} + w_2 \cdot \frac{\frac{1}{T}\sum_{t=1}^{T} \alpha_t}{\sigma(\alpha_t, |, t \in S)} \right) \cdot \frac{12}{\sqrt{12}} \tag{11}$$

其中：

- $\alpha_t = w_t^T R_t - r_f$表示投资组合在时刻 t 的超额收益（相对于无风险利率 r_f，本研究采用美国 3 月期国债利率作为代理变量）
- $\sigma(\{R_t - r_f | t \in S\})$计算负收益时段内$R_t - r_f$超额收益的标准差
- 夏普比率（加号前项）和索提诺比率（加号后项）均按月数据计算，为转换为年化值，分子乘以 12，分母乘以$\sqrt{12}$

实现方法上，函数 f 通过预训练的多层神经网络近似基于规则的模型行为。首先使用梯度提升决策树（GBDT）算法训练神经网络，最小化生成动作 $\widehat{w_t}$与规则模型动作w_t间的均方误差。随后，在强化学习框架中，加载预训练权重的“导师模型”进一步训练“学生模型”，以超越导师模型的表现，具体奖励函数和训练方法详见 3.3.3 和 3.3.4 节。

3.2 数据采集与预处理

为确保投资组合管理策略的历史稳健性，本研究采用近 40 年（1980-2023）的月度资产价格指数历史数据训练并验证模型，针对每项资产，计算以下动量指标作为训练数据：

- 1/3/6/12 个月相对收益率

- 上述四个相对收益率的加权和
- 与 1/3/6/12 个月简单移动平均线的相对收益率
- 与 1/3/6/12 个月指数移动平均线的相对收益率
- 与 1/3/6/12 个月 Hull 移动平均线的相对收益率

尽管市场存在更多可用指标，但为保持训练数据时间跨度的完整性，本研究仅选择计算窗口不超过 12 个时间步长（1 年）的指标。需更大时间窗口的指标（如 MACD 需 26 个时间步长、Awesome Oscillator 需 34 个时间步长）未纳入本次训练。未来研究可探索高频日度数据及更大时间窗口的动量特征，以提升高频交易场景下的策略优化能力。

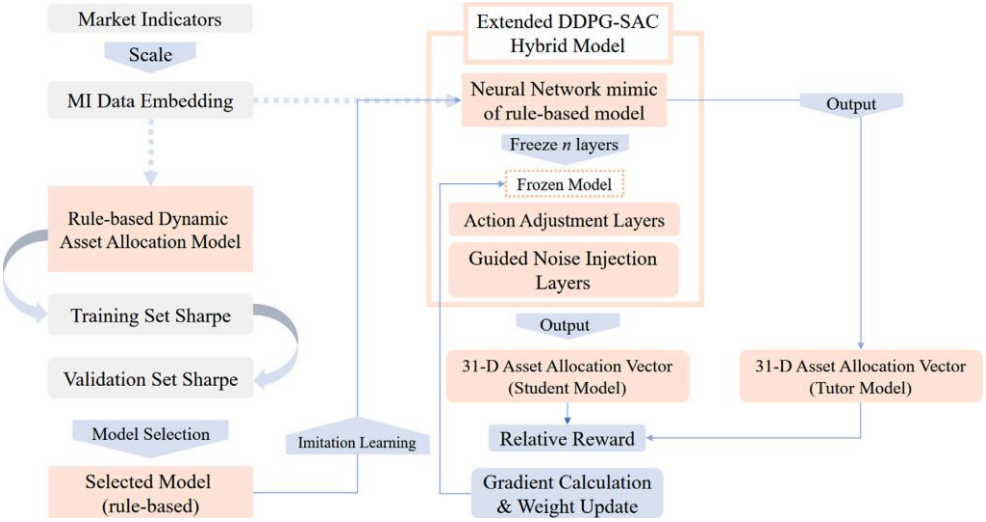
为缓解过拟合风险并提升模型效能，研究采用高斯噪声进行数据增强。原始市场指标数据集包含 472 个数据点（对应 39.3 年月度数据），直接训练可能不足。因此，通过注入零均值、方差可变的高斯噪声扩展数据集规模。受限于计算资源，最终通过该噪声方法将原始数据集扩展 30 倍。

数据分割方面，将完整数据集划分为训练集（70%）、验证集（15%）和样本外测试集（15%）。训练集覆盖 1984 年 2 月至 2011 年 5 月的 40 年市场指标数据；验证集为 2011 年 5 月至 2017 年 4 月；测试集为 2017 年 5 月至 2023 年 5 月。

3.3 模型设计

图表 2 以可视化形式展示了该系统的整体架构设计，直观呈现了模型的关键组件与数据流逻辑。

图表 2 强化学习框架中基于规则模型扩展的选择与训练架构可视化



资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

3.3.1 基于规则的模型

基于各种资产价格指数的历史价格和动量数据，对 BAA 模型进行了多次扩展。这些模型主要服务于两个目标:

1、比较其风险调整后的投资组合表现与机器学习模型的表现;

2、训练一个模仿模型，以便使用先进的强化学习算法进行进一步训练。

探索的 BAA 模型扩展版本包括安全预警资产和风险预警资产，以增强市场监控能力。与仅依赖预警资产来指示风险资产动量下降的原始 BAA 模型不同，我们的增

强方法还跟踪“安全”资产（如国债或黄金）的动量增加。该算法旨在基于双重动量评估生成“逃逸信号”：风险资产的下降和安全资产的上升。当指定数量的风险资产的综合动量低于某个阈值，且 n 个安全资产的综合动量同时高于预定义水平时，该信号被激活，这表明市场可能出现下跌，并发出减少风险投资敞口的策略信号。图表 3 中提出了扩展 BAA 的算法实现。

在扩展 BAA 模型基础上，在资产配置算法中融入了受“五月卖出离场”规律启发的策略。这一策略基于布曼和雅各布森的实证发现，在 37 个发达和新兴市场中，股票市场在 11 月至 4 月期间与 5 月至 10 月期间的回报存在显著差异，后者始终表现出较低的回报。针对这些发现，扩展模型在投资组合管理方法中融入了市场逃逸信号触发器（如图表 4 中所述），提示算法在 5 月至 10 月期间清算投资组合中的大量风险资产。

在基于规则的动态资产配置模型中，资产被分配到四个类别：安全预警资产、贪婪预警资产、贪婪资产和安全资产。分配设计灵活，允许根据实证分析将任意数量（从 1 到 n）的资产放入每个类别。

图表 3 扩展版本 1：扩展的 BBA 再平衡策略

Input: Asset universes (non-leveraged only):

- G (greedy)
- S (safe)
- GC (greedy_canary)
- SC (safe_canary)

Hyperparameters

- top_n_greedy (Top n greedy assets)
- top_n_safe (Top n safe assets)
- $safe\_pct\_at\_esc$ (Portion of safe assets at escape signal)
- $mmt\_check\_threshold$ (Threshold for judging current market situation)

Output: List $G + S$, each element representing portfolio allocation
Ensure: Sum (G + S) = 1
for each portfolio rebalance interval **do**
  for each asset c in GC and SC **do**
    Calculate 1, 3, 6, 12 months' return: r_1, r_3, r_6, r_{12}
    Calculate momentum score:
    $m\_score = 1 \cdot r_1 + 2 \cdot r_3 + 4 \cdot r_6 + 12 \cdot r_{12}$
  end for
  if m_score of all c in $GC < 0$ **OR**
    m_score of any c in $SC > 0$ **then**
    rebalance holdings in G according to $safe\_pct\_at\_esc$
    Allocate remaining resources uniformly across S according to top_n_safe
  else
    rebalance holdings uniformly across G according to top_n_greedy
  end if
end for

资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

图表 4 扩展版本 2：Halloween 策略支持的扩展 BBA

Input: Asset universes (both leveraged & non-leveraged):

- G (greedy)
- S (safe)
- GC (greedy_canary)
- SC (safe_canary)
- M **(Current Month)**

Hyperparameters

- top_n_greedy (Top n greedy assets)
- top_n_safe (Top n safe assets)
- $safe\_pct\_at\_esc$ (Portion of safe assets at escape signal)
- $mmt\_check\_threshold$ (Threshold for judging current market situation)

Output: List $G + S$, each element representing portfolio allocation
Ensure: Sum (G + S) = 1
for each portfolio rebalance interval **do**
  for each asset c in GC **do**
    Calculate 1, 3, 6, 12 months' return: r_1, r_3, r_6, r_{12}
    Calculate momentum score:
    $m\_score = 1 \cdot r_1 + 2 \cdot r_3 + 4 \cdot r_6 + 12 \cdot r_{12}$
  end for
  for each asset c in SC **do**
    Calculate 1, 3, 6, 12 months' relative return against exponential moving average: r_1, r_3, r_6, r_{12}
    Calculate momentum score:
    $m\_score = 1 \cdot r_1 + 2 \cdot r_3 + 4 \cdot r_6 + 12 \cdot r_{12}$
  end for
  if m_score of all c in $GC < 0$ **OR**
    m_score of any c in $SC > 0$ **OR**
    M is in between 5 and 10 **then**
    rebalance holdings in G according to $safe\_pct\_at\_esc$
    Allocate remaining resources uniformly across S according to top_n_safe
  else
    rebalance holdings uniformly across G according to top_n_greedy
  end if
end for

资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

图表 5 展示了各种基于规则的动态资产配置模型的业绩指标。该表包括了图表 3 中 BAA 策略扩展及其变体的实现，以及图表 4 中详述的另一种 BAA 扩展及其特定变体。我们为每个模型在三个独立的数据集上计算了全面的性能评估指标。

图表 5 基于规则的资产配置扩展策略的业绩比较

Model	Sharpe			Sortino			CAGR			MDD		
	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test
Ext 1	0.677	0.662	0.720	1.209	1.178	1.297	10.74%	5.24%	6.44%	17.03%	14.40%	11.22%
Ext 1A	0.689	0.707	0.649	1.174	1.181	1.181	12.67%	6.75%	7.30%	30.01%	17.91%	9.63%
Ext 2	1.183	1.217	0.675	2.450	2.454	1.264	24.69%	17.48%	10.06%	21.62%	19.10%	13.73%
Ext 2B	1.047	0.955	0.669	2.099	1.651	1.371	15.88%	8.98%	7.53%	19.39%	10.19%	7.74%

资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

如上所述，Ext 1 模型和 Ext 2 模型之间的主要区别在于是否采用了 Halloween 抛售策略。带有字母后缀的模型表示相应变体的扩展版本，它们在结构上没有变化，但在四个类别中的资产分布不同，以及在预警资产指示从风险较高的市场转向较安全资产时（算法 1 和 2 中的 safe_pct_at_esc）的最大抛售速率有所不同。

基于图表 4 的模型往往表现出更高的夏普比率和索提诺比率，这表明其收益风险特征有所改善，实施了 Halloween 抛售策略并包含杠杆资产的模型（Ext 2 和 Ext 2B）在所有评估指标上均表现出增强：复合年增长率（CAGR）、索提诺比率和夏普比率更高，最大回撤（MDD）降低，特别是在验证集和测试集中。

在这些基于规则的模型中观察到的一个显著局限性是，尽管验证性能通常与训练性能一致（有时超过训练准确性，可能是由于训练模型的数据分布有利），但测试集性能出现了明显下降。第 4 节将进一步分析，以确定在强化学习（RL）框架内扩展训练是否能解决这种差异，并在训练和验证期间保持或提高性能。

3.3.2 模仿学习

我们的方法与传统强化学习（RL）方法存在显著差异，后者通常从模型参数的随机初始化开始。而我们首先将基于规则的“专家”模型中的专业知识迁移到“学生”模型中。这一初始阶段的目标不仅是提升模型的智能性，还要增强其在训练过程中的可解释性和决策一致性。通过以基于规则的模型为基础，提供了对模型决策过程更高的透明度，从而在可信度方面相比从零开始训练的模型更具优势。

知识迁移过程涉及训练学生模型以模仿基于规则模型的行为，这与 Hinton 等人提出的知识蒸馏方法类似。在此过程中，重点不在于蒸馏或模型压缩，而在于精准复现初始导师模型的动作。基于规则模型生成的状态-动作对被用作专家动作，用于训练全连接神经网络模型。

使用的损失函数简洁明了：计算导师模型动作向量a与学生模型生成的动作向量$\hat{a}$之间的欧氏范数，即$\|a-\hat{a}\|_2$。这些初始模型充当“导师”模型，为每个“学生”网络提供基础架构。一旦模型权重被初始化为与导师模型对齐，学生网络会进一步训练以超越其对应的导师模型。为防止过拟合，采用了高斯噪声数据增强、权重衰减和引入 Dropout 层等标准方法。

3.3.3 导师-学生模型

鉴于学生模型的权重由其对应的导师模型初始化以传递基于规则模型的专业知识，本文 3.3.4 节详述的混合强化学习框架（Hybrid RL Framework）会进一步训练该模型以最大化累积收益，这与典型的强化学习算法目标一致。然而，本研究与先前研究的不同之处在于，我们重点引导模型提升“相对”累积收益。目标是在预定的投资期限内（例如通过持续执行月度再平衡的序贯投资策略，在 20 年即 240 个月的投资周期结束时）超越基准（导师）模型，实现风险调整后的组合价值。强化学习框架中用于训练学生模型的奖励函数 R 基于 3.1 节建立的前提，具体公式如下：

$$R = w_1 \cdot f(ALL) + w_2 \cdot f(RCT) + w_3 \cdot g(ALL) + w_4 \cdot g(RCT) \quad (12)$$

其中：

- f 和 g 分别表示基于月度组合收益率计算的夏普比率（Sharpe Ratio）和索提诺比率（Sortino Ratio）
- ω_t为权重系数且总和为 1（初始值均设为 0.25，后续实验测试不同取值）

• "ALL"和"RCT"分别代表训练后的智能体从完整训练数据集和仅最近三分之一训练数据中获得的月度组合收益率

• 特别设定 $\omega_1=\omega_2=0.5$，$\omega_3=\omega_4=0$

如 2.1 节所述，作为基础的导师模型来源的传统资产配置框架在历史表现上存在显著偏态。具体而言，在 40 年回测周期的最后 10-15 年，其表现明显低于前 25-30 年。为应对这一问题，奖励计算中对模型近期表现赋予更大权重。经过探索性实验后，最终将ω_1和ω_2均设为 0.5，ω_3和ω_4设为 0。这一调整表明奖励评估仅考虑夏普比率，且特别强调模型近期表现以抵消收益递减效应。选择夏普比率作为唯一评价指标，旨在集中展示通过知识迁移神经网络模型的进一步训练所能实现的最优性能提升。这与 Eling 和 Schuhmacher 的研究发现一致，即大多数成熟的组合评估指标高度相关，不同指标下的排名可能不会产生显著差异。

3.3.4 DDPG-SAC 混合模型

本研究中用于训练学生模型的框架是一种先进的混合框架，它融合了柔性决策-评估（Soft Actor-Critic，SAC）和深度确定性策略梯度（Deep Deterministic Policy Gradient，DDPG）算法的思想。该模型的核心基于 DDPG 架构，采用确定性方法来提高可靠性和可重复性。这一设计选择对于金融模型特别合适，因为它能确保行动的一致性，从而建立可信度。

在我们的框架中，行动者模型以确定性的方式产生行动，采用独特的导师-学生范式。模型首先通过模仿一个预训练的“导师”模型开始，然后随着时间的推移不断精炼其策略。此外，在保持 DDPG 确定性核心的同时，模型还融入了 SAC 的双评估方法来减轻评估偏差，这是决策-评估模型中常见的一个问题。

SAC 和 DDPG 算法都以软策略更新为特征。我们的混合模型利用了这一特性，使得主网络的权重能够缓慢地集成到目标网络中，从而稳定训练过程，解决了深度强化学习通常具有的波动性。此外，决策和评估的更新频率不同，评估按照 SAC 的方法接受更频繁的更新。

然而，与标准的 SAC 模型不同，混合模型并没有直接将熵项集成到损失函数中。相反，它引入了两个不同的促进熵的组件：a. 一个高斯噪声加法器，它直接干预行动向量，鼓励探索行为；b. 一个带有可训练参数的“引导”噪声注入网络。这个网络是一个关键的区别点，它动态地产生噪声，旨在最大化累积奖励。

3.3.5 动作调整模块

该过程旨在在指定的数据时间框架内最大化经风险调整后的回报。调整网络由多个全连接层组成，且隐藏神经元的数量有限；关键的是，这些层不包含任何非线性激活函数。该网络的主要功能是调整执行者模型（actor model）的输出，而不是独立地生成动作。网络参数是可通过梯度学习的，并通过梯度下降法进行优化，以在预定的投资窗口内提高累计回报。整合这个辅助网络被认为是有利的，因为它为学生模型提供了额外的自由度，使其能够从市场指标数据中识别出超出导师模型所传授的基础知识的模式。

3.3.6 引导噪声注入网络

为了在动作空间中鼓励探索并防止模型在训练过程中陷入局部最小值，并没有

直接向投资组合优化模型生成的动作添加随机高斯噪声，而是采用了一个专门的神经网络模块来生成一种更有结构的噪声。这种噪声不是随机的，而是与投资组合优化的主要任务一起学习的。称这个辅助神经网络为“引导噪声注入网络”，它与主要投资组合模型一起优化可训练参数。与随机噪声不同，这种方法允许模型以一种使其产生的噪声与主模型输入特征或动作相结合的方式，引导投资组合优化模型朝着更有效、更具鲁棒性的数据表示方向调整。

例如，在市场波动较高的时期，引导噪声注入网络可能会学习到产生噪声模式，促使主要投资组合模型采取更保守的仓位，这作为一种学习到的风险管理方式。相反，在市场更稳定的条件下，生成的噪声可能会推动投资组合模型利用更积极的市场机会。这与传统的随机噪声注入方法形成对比，在传统方法中，噪声并未根据数据的特定特征进行适应，也没有通过迭代优化来提高投资组合的表现。

图表 6 资产池在训练阶段由模型选择

Idx	Asset name	Asset type	Idx	Asset name	Asset type
1	Nasdaq Index Composite	Stock	17	Goldman Sachs Commodities Index	Commodities
2	S&P 500 Index	Stock	18	Consumer Price Index	Inflation
3	Dow Jones Index	Stock	19	Gold Spot	Precious metal
4	US Total Stocks Index	Stock	20	Nasdaq Index Composite 3X	Lvg'd stock
5	MSCI Pacific Stocks Ex Japan Index	Stock	21	S&P 500 Index 3X	Lvg'd stock
6	US Small-Cap Growth Stock Index	Stock	22	Dow Jones Index 2X	Lvg'd stock
7	US Small-Cap Value Stock Index	Stock	23	Dow Jones Index 3X	Lvg'd stock
8	US Large-Cap Growth Stock Index	Stock	24	US Total Stocks Index 3X	Lvg'd stock
9	US Large-Cap Value Stock Index	Stock	25	US Small-Cap Value Stock Index 3X	Lvg'd stock
10	Cash Proxy (3 Month T-Bill)	Treasury	26	Long Term Treasury 3X	Lvg'd treasury
11	Long Term Treasury	Treasury	27	Intermediate Term 3X	Lvg'd treasury
12	Intermediate Term Treasury	Treasury	28	Gold 3X	Lvg'd precious metal
13	Short Term Treasury	Treasury	29	Nasdaq 100 Index Short	Stock short
14	Investment Grade Short Term CB	Bond	30	S&P 500 Index Short	Stock short
15	Long Term CB	Bond	31	Dow Jones Index Short	Stock short
16	High Yield CB	Bond			

资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

4 实证结果

在强化学习（RL）中，标准做法是保持其他条件一致的情况下使用不同随机种子进行多次训练，最后按预定方式聚合结果。这种方法的合理性源于 RL 的内在特性——它强调探索，尤其在智能体面对复杂非稳态状态（如金融数据的典型特征）时更为关键。基于这一认识，我们训练了 4 个智能体，各进行 20 个训练周期，每个智能体使用不同初始种子但保持相同的设置和超参数。随后，根据验证集的夏普比率及与训练数据表现的一致性，选择表现最好的 3 个模型计算其平均性能。

图表 7 模型性能提升结果

Model	Sharpe			Sortino			CAGR			MDD		
	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test
Ext 1 T	0.677	0.662	0.720	1.209	1.178	1.297	10.74%	5.24%	6.44%	17.03%	14.40%	11.22%
Ext 1 S	0.681	0.697	0.753	1.228	1.267	1.397	11.26%	5.64%	7.24%	19.67%	13.83%	11.45%
Boost	0.59%	5.29%	**4.58%**	1.57%	7.55%	**7.71%**	4.84%	7.63%	**12.42%**	+15.50%	−3.96%	**+2.50%**
Ext 1A T	0.689	0.707	0.649	1.174	1.181	1.181	12.67%	6.75%	7.30%	30.01%	17.91%	9.63%
Ext 1A S	0.702	0.794	0.711	1.237	1.330	1.300	12.95%	7.68%	7.98%	26.13%	16.04%	10.18%
Boost	1.89%	12.31%	**9.55%**	5.37%	12.62%	**10.08%**	2.21%	13.78%	**9.32%**	−12.93%	−10.44%	**+5.71%**
Ext 2 T	1.183	1.217	0.675	2.450	2.454	1.264	24.69%	17.48%	10.06%	21.62%	19.10%	13.73%
Ext 2 S	1.300	1.371	0.943	2.749	2.788	1.893	21.16%	15.10%	10.80%	17.54%	13.38%	8.09%
Boost	9.89%	12.65%	**39.70%**	12.20%	13.61%	**47.07%**	−14.30%	−13.62%	**7.36%**	−18.87%	−29.95%	**−41.08%**
Ext 2B T	1.047	0.955	0.669	2.099	1.651	1.371	15.88%	8.98%	7.53%	19.39%	10.19%	7.74%
Ext 2B S	1.141	0.982	0.857	2.226	1.754	1.638	15.72%	8.89%	8.69%	21.60%	10.30%	6.61%
Boost	9.98%	2.83%	**28.10%**	6.05%	6.24%	**19.47%**	−1.00%	−1.00%	**15.41%**	+11.40%	+1.08%	**−14.60%**

资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

图表 7 展示了第 3.3.1 节介绍的基于规则的基准模型在经历 RL 训练后，其神经网络模仿模型（学生模型）的性能提升幅度。"Boost"表示各学生模型（标记为 S）相对于对应导师模型（标记为 T）在训练集、验证集和测试集各分区上的百分比性能提升。如前所述，CAGR（复合年化增长率）和 MDD（月末最大回撤）指标在训练集、验证集和测试集上均进行了评估。

根据上文所述，奖励函数引导 RL 智能体优化夏普比率。因此，预计所有微调后的模型（无论采用哪个导师模型）将在夏普比率上显著优化，这一点得到了验证。值得注意的是，所有 4 个模型在训练集、验证集和测试集上的夏普比率均有所提升，且在测试集上增幅更为显著。这表明经过一系列正则化处理后，模型在基于市场动量信号进行多资产类别配置时的泛化能力更强。在索提诺比率方面，学生模型在训练集和验证集上表现出显著提升，测试集的索提诺比率提升幅度甚至更为突出。

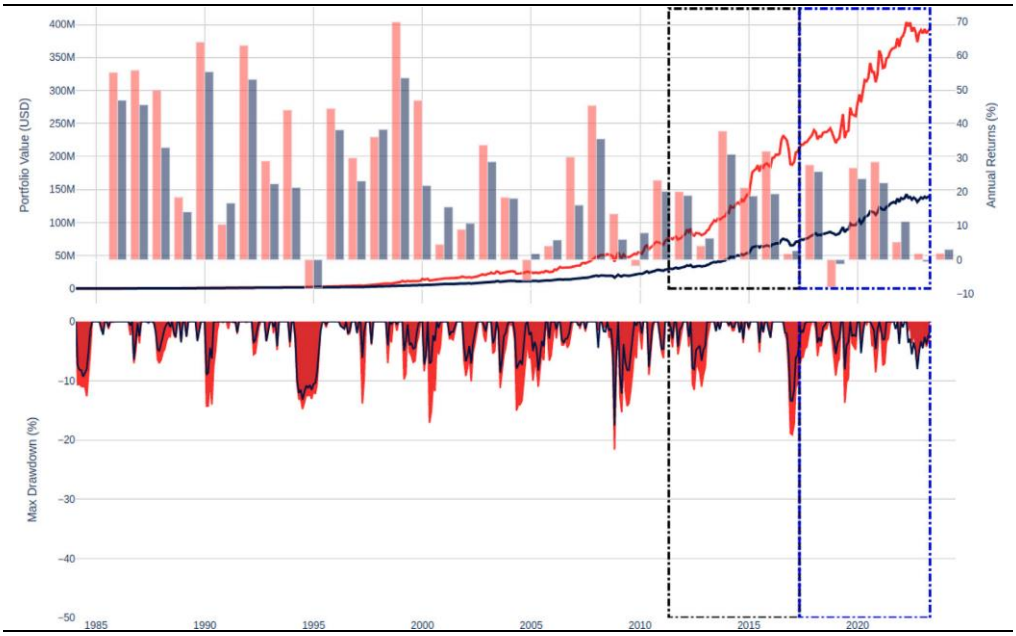
关于 CAGR 和 MDD 指标，虽然两个指标均有小幅改善，但在验证集或测试集上也存在个别性能下降的情况。部分模型出现 CAGR 小幅下降或 MDD 上升，这源于优化夏普比率时的权衡。例如，Ext 2B S 模型相比其导师模型，测试集夏普比率大幅提升 28.10%，但训练集和验证集的 MDD 分别上升 11.40%和 1.08%。尽管如此，我们认为夏普比率的显著提升足以补偿 MDD 的轻微恶化，这种妥协是可接受的。

观察到复杂度更高的模型（Ext 2 S 和 Ext 2B S，测试集夏普比率分别提升 39.70%和 28.10%）相比简单模型（Ext 1 S 和 Ext 1A S，测试集夏普比率分别提升 4.58%和 9.55%）性能提升更为显著。在 RL 训练阶段，Ext 2 学生模型及其 2B 变体可从 31 个资产类别中自由选择。尽管这种灵活性可能引入不稳定性（尤其是纳入杠杆资产时），但也增强了其适应性。相比之下，Ext 1 学生模型及其 1A 版本仅限于非杠杆资产。影响模型组间性能差异的另一因素是"Halloween 策略"的采用：Ext 2 模型组包含该策略，而 Ext 1 组不包含。目前尚不确定该策略是否独立提升了 Ext 2 学生模型的表现，还是与杠杆资产选择产生了协同效应。尽管探索这一问题颇具价值，但本文的核心目标并非重新评估"Halloween 策略"的有效性。

图表 8 通过三个指标（累计投资组合价值、年化收益率和月末最大回撤）比较了 Ext 2 导师模型与学生模型的表现。图中红色始终代表 Ext 2 导师模型，深色代表 Ext 2 学生模型。上半部分的折线图展示了从 10 万美元初始假设投资开始的累计投资组合价值（美元）。顶部相邻的柱状图以百分比形式量化了年化投资回报率。下半部分的折线图则解释了各投资组合的月末最大回撤。两个子图均用黑色和蓝色边框

框出了重点区域，分别表示在验证数据和测试数据上的回测表现。

图表 8 Ext 2 型学生模型（深色）与导师模型（红色）的对比

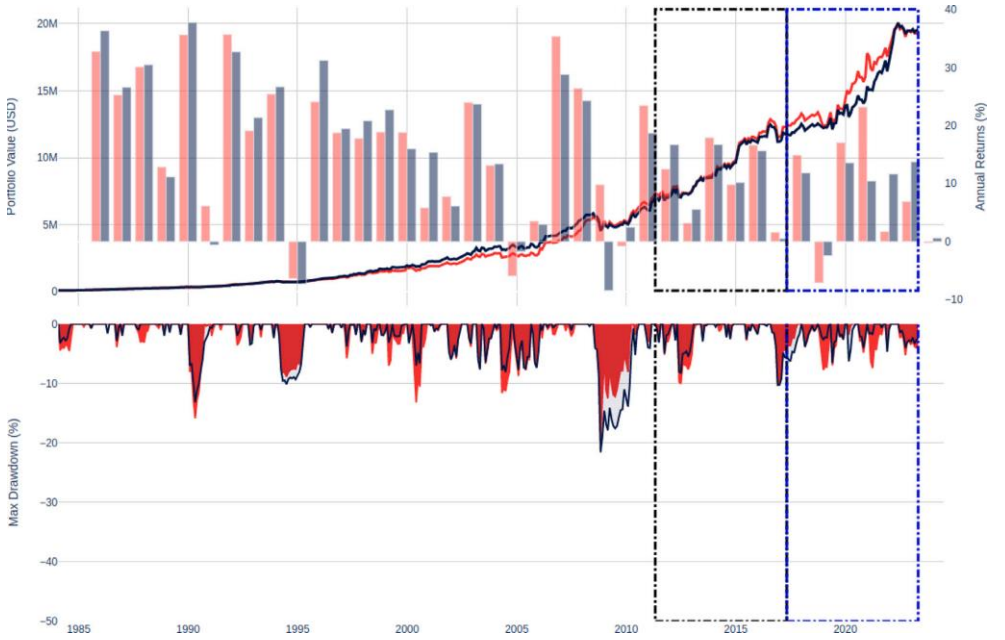


资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

观察到的 3.2%复合年增长率（CAGR）差异（Ext 2 Tutor 模型为 23.45% vs. Ext 2 Student 模型为 20.25%，基于 40 年数据集）导致投资期末的资产组合价值出现显著差距。具体而言，Tutor 模型的最终资产价值比 Student 模型高出约 180%。初看之下，这似乎表明 Tutor 模型具有压倒性优势。然而，进一步分析图表 8 中的回撤曲线发现，由学生模型（深色线）呈现月末回撤幅度更窄且更浅，而导师模型（红色区域）的回撤则更深更广。这表明**学生模型在波动性量化的整体风险管理方面表现更优**。需要强调的是，该模型的设计初衷是优化夏普比率，因此将系统化风险管理功能嵌入资产组合优化框架，而非单纯追求收益最大化。图表 7 的结果进一步印证了这一设计目标：**学生模型在训练集、验证集和测试集上分别将夏普比率提升了 9.89%、12.65%和 39.70%，显著优于导师模型**。

与 Ext 2 Student 模型通过牺牲部分年度增长率以换取更强的风险管理能力和收益稳定性不同，Ext 2B Student 模型在 CAGR 上几乎未做妥协。尽管其回撤曲线差异不如图表 7 所示案例显著，但该模型仍分别在训练集、验证集和测试集上将夏普比率提升了 9.98%、2.83%和 28.10%，其能在几乎不影响资产收益的前提下提升风险调整绩效，堪称重要突破。

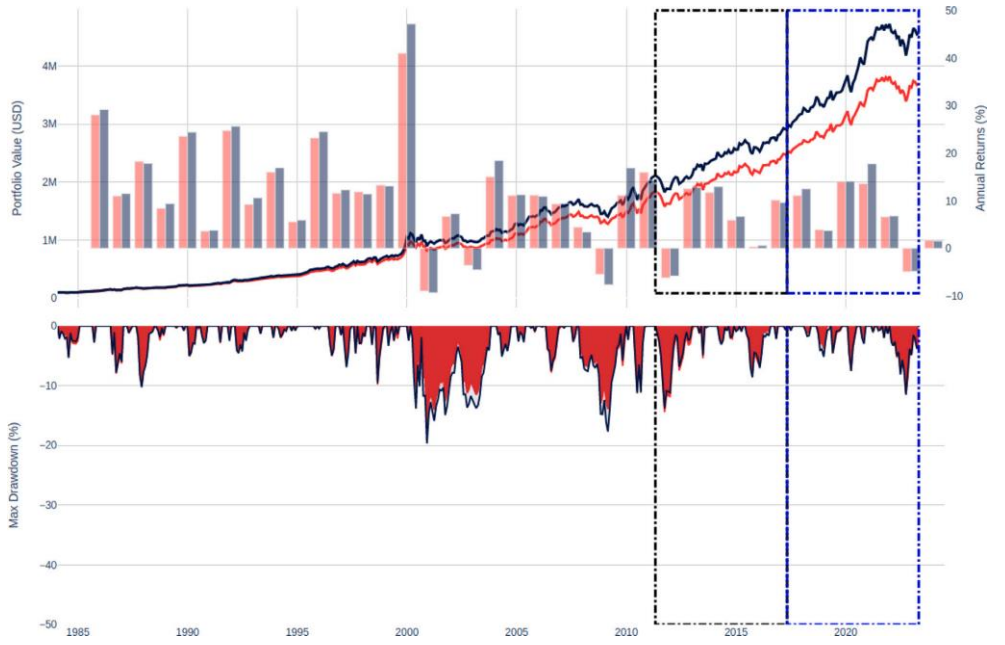
图表 9 Ext 2B 型学生模型（深色）与导师模型（红色）的对比



资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

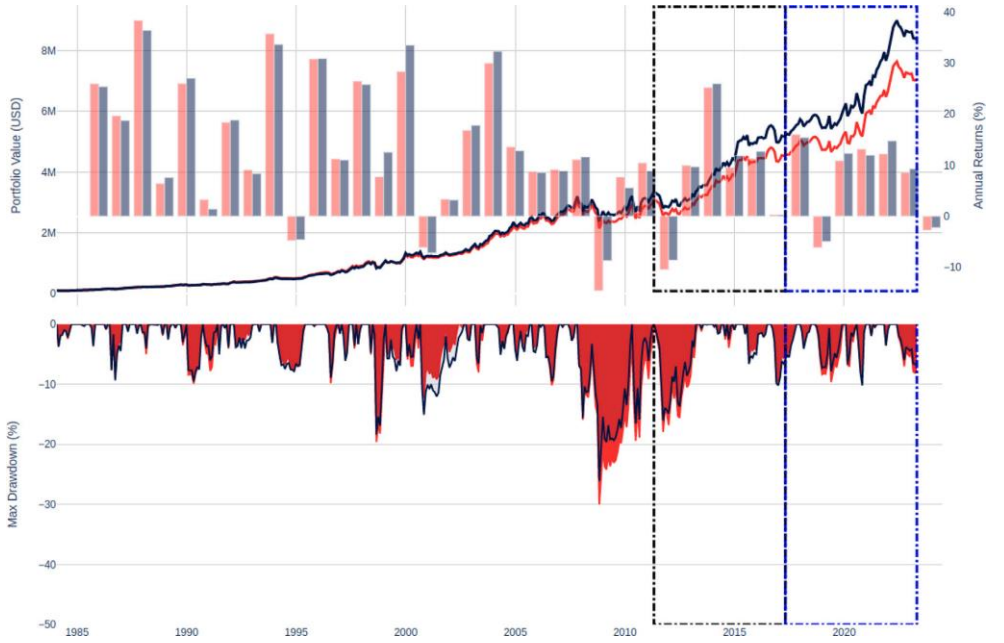
类似地，图表 10 和图表 11 对两组简化版导师模型（Ext 1 及其变种）及其对应学生模型进行了对比评估。这些导师模型资产选择范围有限且未采用 Halloween 策略改进方案。有趣的是，两组学生模型均实现了更高的期末资产价值（Ext 1 提升 23.63%，Ext 1A 提升 19.58%），同时测试集最大回撤仅分别增加 0.23 和 0.55 个百分点。尽管在夏普比率和索提诺比率的提升幅度上不及复杂模型（Ext 2 及 Ext 2B），但改进效果仍清晰可见。

图表 10 Ext 1 型学生模型（深色）与导师模型（红色）的对比



资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

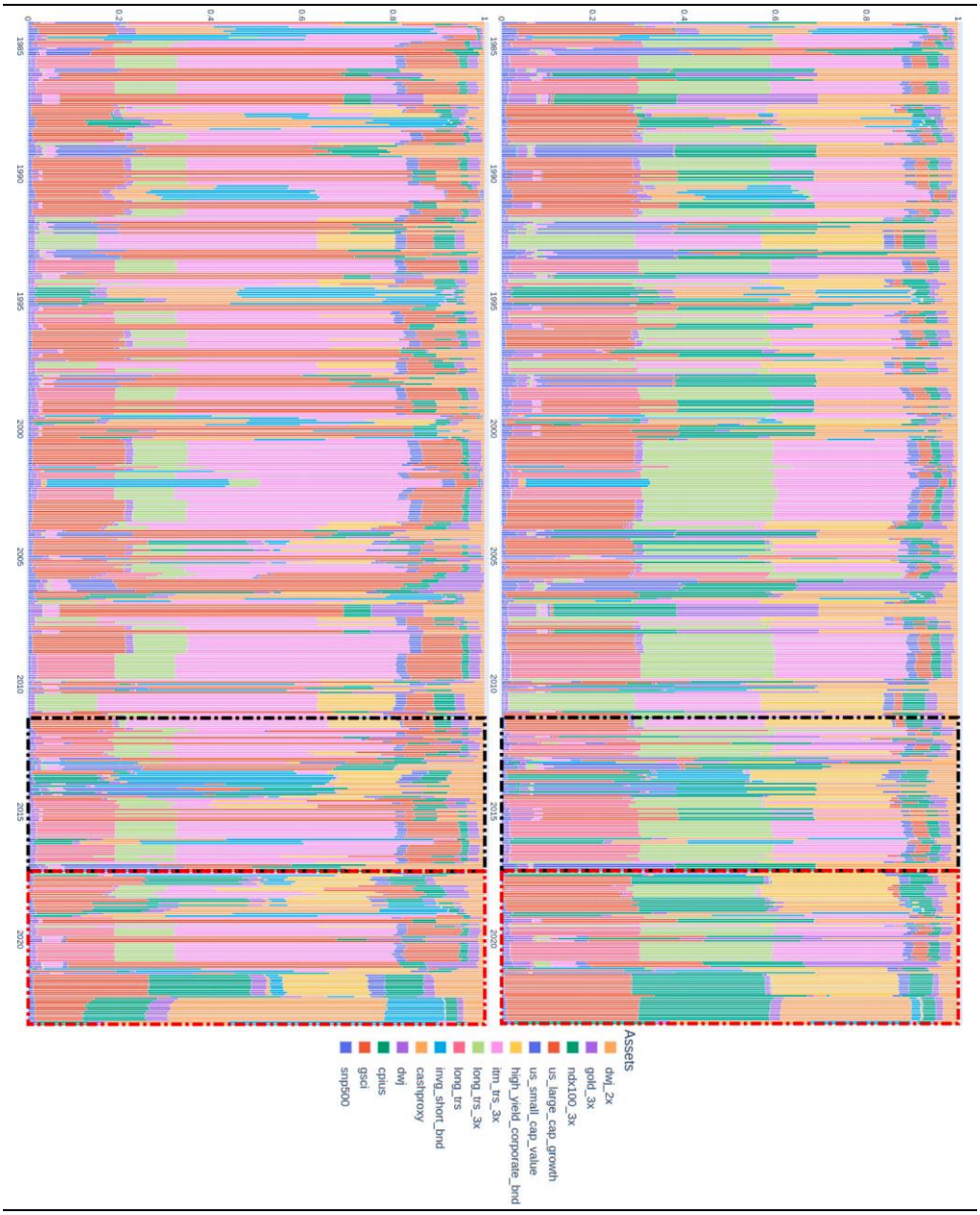
图表 11 Ext 1A 型学生模型（深色）与导师模型（红色）的对比



资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

图表 12 展示了训练前后资产分配的历史变化情况，重点对比了与 Ext 2 导师模型配对的学生模型。图中验证集和测试集的资产分配记录分别以黑色和红色边框框示。为便于理解，图中未展示占比低于资产组合 0.1%的类别及其图例。相较于规则型导师模型的固定资产配比，基于神经网络的学生模型在资产比例决策中引入了一定灵活性，但整体分配模式仍保留了原始模型的显著特征。这种相似性可归因于模型设计时冻结了 30%-50%的隐藏层，并结合极低的训练学习率，确保投资组合管理策略的高层主题保持相对一致。不过，经过微调后，具体配置细节已发生部分调整。根据研究结果显示，神经网络模型所赋予的这种分配灵活性，可能有助于形成更稳健的资产配置策略，从而极大提升风险调整后的收益稳定性。

图表 12 Ext 2 型学生模型（左图）与导师模型（右图）资产配置记录对比



资料来源：《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》，华安证券研究所

5 结论

文献首次将量化投资组合管理中基于规则方法的动态资产配置知识迁移到深度强化学习领域，从而提升了模型性能。通过借鉴进攻性资产配置和防御性资产配置等技术，增强现有基于规则的组合管理策略。为充分发挥成熟模型的优势，这些规则型模型并未被直接采用，而是通过将"安全型"和"进取型" Canary assets 融入现有市场状态诊断框架，对原有规则体系进行了扩展。

该增强型基础模型进一步通过模仿学习、强化学习（RL）和相对导师-学生机制进行精细化调优，展现出提升组合管理性能的潜力。部分规则型模型通过模仿学习被转换为神经网络架构，随后采用融合深度确定性策略梯度（DDPG）和柔性决策-

评估算法（SAC）的混合 RL 算法，并引入多种探索增强机制，实现了性能突破。该方法显著提升了既有规则型模型的表现：测试集的夏普比率最高提升 39.70%（这是 RL 微调的核心目标），另一风险调整收益指标索提诺比率在不同模型扩展中从 7.71% 提升至 47.07%。值得注意的是，RL 微调对复杂模型（如允许杠杆操作并采用 Halloween 策略的 Ext 2 和 2B 模型）的提升效果更为显著，其夏普比率分别提升 39.70%和 28.10%；而较简单的 Ext 1 和 1A 模型测试集表现提升幅度较小，分别为 4.58%和 9.55%。

将成熟的规则型模型与现代强化学习技术相结合，为组合管理提供了新颖有效的解决方案。基础模型的系统性扩展与优化，为传统金融模型与现代计算方法的融合优势提供了实证依据。模型性能的改善不仅验证了所提方法的有效性，也指明了未来探索方向。随着金融市场的日益复杂，对严谨且适应性强的策略需求愈发迫切。本研究充分利用规则型算法与 RL 算法优势的方法论，为量化投资组合管理的后续实证研究树立了基准。

文献来源：

核心内容摘选自 Chanwoo Choi, Juri Kim 在 Knowledge-BasedSystems 上的论文《Outperforming the tutor: Expert-infused deep reinforcement learning for dynamic portfolio selection of diverse assets》。

风险提示：

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

重要声明

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

免责声明

华安证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

投资评级说明

以本报告发布之日起 6 个月内，证券（或行业指数）相对于同期沪深 300 指数的涨跌幅为标准，定义如下：

行业评级体系

增持—未来 6 个月的投资收益率领先沪深 300 指数 5%以上；

中性—未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-5%至 5%；

减持—未来 6 个月的投资收益率落后沪深 300 指数 5%以上；

公司评级体系

买入—未来 6-12 个月的投资收益率领先市场基准指数 15%以上；

增持—未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%；

中性—未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%；

减持—未来 6-12 个月的投资收益率落后市场基准指数 5%至；

卖出—未来 6-12 个月的投资收益率落后市场基准指数 15%以上；

无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。市场基准指数为沪深 300 指数。