李彬, 潘乔, 阎希平. 基于深度强化学习的投资组合构建方法[J]. 智能计算机与应用,2024,14(8):85-90. DOI:10.20169/j.issn.2095-2163.240814

基于深度强化学习的投资组合构建方法

李 彬[1], 潘 乔[1], 阎希平[2]
(1 东华大学 计算机科学与技术学院, 上海 201620; 2 上海兆前投资有限公司, 上海 201107)

摘 要: 传统基于数据分析的投资组合构建方法使用简单的统计学模型,不仅难以发现市场规律,且在处理大量数据时效率不高。而深度强化学习算法具备强大的数据处理和分析能力,能够通过学习自适应调整策略,从海量金融数据中提取出有效信息,处理复杂多变市场环境并为投资决策提供科学建议。针对金融资产价格具有非平稳特点和各资产间具有相互依赖性的问题,本文基于深度强化学习中的深度确定性策略梯度 DDPG 算法,设计了一种并行投资组合特征提取网络 PPFNet 作为策略网络用于构建投资组合。实验结果表明,PPFNet 相较于其他主流投资组合构建方法,取得了最优的收益效益,且表现出良好的稳定性。

关键词: 投资组合; 深度强化学习; DDPG; PPFNet
中图分类号: TP301 **文献标志码**: A **文章编号**: 2095-2163(2024)08-0085-06

Portfolio construction method based on deep reinforcement learning

LI Bin[1], PAN Qiao[1], YAN Xiping[2]
(1 College of Computer Science and Technology, Donghua University, Shanghai 201620, China;
2 Shanghai Zhaoqian Investment Co. LTD., Shanghai 201107, China)

Abstract: Traditional portfolio construction methods based on data analysis often rely on simple statistical models that are unable to discover market patterns and can be inefficient when processing large amounts of data. In contrast, deep reinforcement learning algorithms possess powerful data processing and analysis capabilities, allowing them to extract valuable information from massive financial data, handle complex and changing market environments, and provide scientific advice for investment decisions by adapting strategies through self-learning. To address the issue of non-stationary asset price characteristics and interdependence between assets in the financial market, this paper proposes a parallel feature extraction network, PPFNet, based on the deep deterministic policy gradient (DDPG) algorithm in deep reinforcement learning as a policy network for constructing investment portfolios. Experimental results demonstrate that PPFNet outperforms other mainstream portfolio construction methods in terms of profit efficiency and exhibits excellent stability.

Key words: portfolio; deep reinforcement learning; DDPG; PPFNet

0 引 言

随着信息科技进步,深度学习在金融科技领域广泛应用。大多数金融市场的深度学习方法,将预测价格的波动曲线或运动趋势作为目标[1-2],将历史价格指标作为输入,模型可输出对应预测值,据此做出相应投资行为。但这种做法的性能很大程度上是依赖于预测的准确性,且预测价格本身并不是市场行为,将其转化为市场中的投资行动需要额外的逻辑转换。如果这种转换是通过人工手动编码实现,那么结果不具有稳定性和扩展性[3]。

近年来,在电子竞技[4]和棋类游戏[5]中,深度强化学习表现突出。如果把金融市场看作是一个不断进化的环境,那么强化学习的方法可以很好地应用在量化交易这一领域[6],可以提供一个与金融市场更为贴合的决策方案[7]。

齐岳等[8]在中证 100 指数中随机选取 16 只股票作为数据集,将深度强化学习中的 DDPG 算法构建的投资组合与等权重投资组合进行收益价值对比。结果表明,在各阶段 DDPG 算法构建组合表现

作者简介: 李 彬(1999-),男,硕士研究生,主要研究方向:智慧金融,人工智能; 阎希平(1978-),女,硕士,工程师,主要研究方向:金融大数据。
通讯作者: 潘 乔(1977-),男,博士,副教授,主要研究方向:大数据,人工智能。Email:panqiao@ dhu.edu.cn

收稿日期: 2023-05-03

均更优。王康等[9]利用 TD3 和 PG 两种深度学习算法进行投资组合管理,在国内 5 只成交量最大的股票上进行实验,其测试集的结果显示,年收益率分别达到了 84.71%和 55.06%。Jang 等[10]将深度学习技术和传统的金融理论结合,使其投资组合可以根据市场行情动态调整投资权重的同时,在夏普比率、年化收益率和最大回撤方面上的表现优于其他主流方法。Yue 等[11]提出了一种基于深度强化学习的抗风险投资组合交易方法,通过大量实验结果表明,该方法优于道琼斯工业平均指数和主流最先进的方法。

尽管相关工作已使得模型策略大致符合真实市场行情,但仍有部分问题需要解决。其一,资产价格序列的非平稳使得特征学习表示困难,平稳化会阻碍模型预测能力,不同序列产生不可区分时间注意力。金融时序数据具有噪声、跳跃和振荡,时变均值、方差和协方差[12],深度学习模型学习特征表示难度大。手工特征方法[13]泛化能力差,且平稳化限制模型预测能力,无法捕捉事件时间依赖性[14]。其二,金融资产间是相互依赖的,同一行业间的股票往往会一起波动[15]。若忽略这一点,可能会导致构建投资组合过程中面临高风险的问题。

为此,本文基于深度强化学习中的深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法,设计了一种并行投资组合特征提取网络(Parallel Portfolio Feature Extraction Network, PPFNet)作为策略网络,用于构建投资组合。在 PPFNet 中,一方面引入了非平稳 Transformers(Non-stationary Transformers)解决金融时序序列非平稳性导致特征难以表示问题;另一方面使用图卷积网络(Graph Convolutional Network, GCN)提取资产间依赖性特征,避免投资组合中出现高风险状况,最终将两部分特征进行融合做出决策。

1 问题描述

投资组合构建指的是根据当前市场条件,将资金按照一定比例分配到一系列金融资产的过程,在最大化收益的同时约束资产组合的整体风险。这种决策过程可以通过马尔科夫决策过程进行描述,即 $M = < S, A, P, R >$,其中 S 为状态空间,A 为动作空间。其执行过程为:发生某个交易动作 $a_t \in A$,则当前状态 s_t 按照以下转移分布发生变化:

$$s_{t+1} \sim P(s_{t+1} \mid s_t, a_t) \tag{1}$$

其中,$P(\cdot)$ 为状态概率分布函数。

得到新状态 s_{t+1} 后,计算得到当前的收益值:

$$r_t = R(s_{t+1}, s_t, a_t) \tag{2}$$

其中,$R(\cdot)$ 为奖励函数,通常以组合收益进行定义。

若投资组合中包含 n 种资产,在 t 时刻,投资组合向量 $\boldsymbol{w}_t$ 可描述为

$$\boldsymbol{w}_t = [w_{1,t}, w_{2,t}, \cdots, w_{n,t}]^{\mathrm{T}} \in \mathbb{R}^{n\times 1}, \ \sum_{i=1}^{n} w_{i,t} = 1 \tag{3}$$

假定使用 $close_t \in \mathbb{R}^{1\times m}$ 表示资产的收盘价格序列,m 表示序列窗口大小。对应收盘价变化量 p_t 可以表示为

$$p_t = \frac{close_t}{close_{t-1}} \tag{4}$$

经过 T 个周期后,投资组合的资产总价值可以表示为

$$W_n = W_0 \prod_{t=1}^{T} \boldsymbol{w}_t p_t (1 - c_t) \tag{5}$$

其中,W_0 为初始资产总价值,c_t 为交易费比例。

2 模型架构

本文基于深度强化学习中的 DDPG 算法,设计了一种并行特征提取网络 PPFNet 作为策略网络用于构建投资组合,模型的整体架构如图 1 所示。

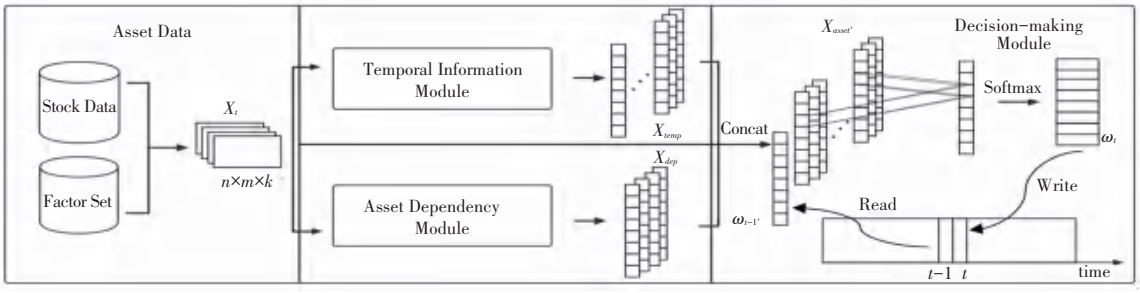

图 1 投资组合构建模型整体架构

Fig. 1 Overall architecture of portfolio construction model

首先，根据资产原始数据 Stock Data 和因子集合 Factor Set，构建一个大小为 $n \times m \times k$ 的张量作为 PPFNet 模型输入数据。其中 n 表示资产数量，m 表示窗口大小，k 为通道数，表示共有 k 种因子。输入价格序列数据 X_t，分别通过时间序列信息模块和依赖性信息模块进行特征提取，得到 X_{temp} 和 X_{dep}，两者进行合并生成 X_{asset}。在投资决策模块中，将 X_{asset} 和上一轮迭代的投资组合权重结果 w_{t-1} 共同用于计算本次迭代的投资组合向量 $\boldsymbol{w}_t$。

2.1　时间序列信息提取模块

本文引入非平稳 Transformers（Non－stationary Transformers，NsT）作为资产时序信息提取框架，与传统的 Transformer 相比，其在架构上加入了序列标准化和反标准化模块，使数据在标准化后，仍可以恢复原始数据的非平稳时间依赖关系，具体架构见图 2。

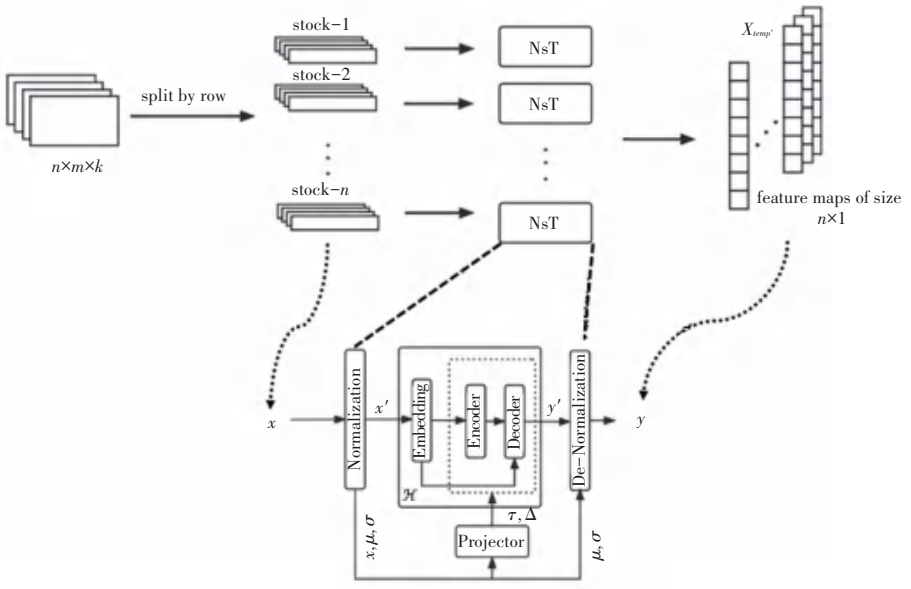

图 2　资产时间序列信息提取模块架构图

Fig. 2　Architecture of time series information extraction module

对于每个资产时序数据 $x = [x_1, x_2, \cdots, x_m]^{\mathrm{T}} \in \mathbb{R}^{m \times k}$，标准化后的数据 $x' = [x_1', x_2', \cdots, x_m']^{\mathrm{T}} \in \mathbb{R}^{m \times k}$ 可以表示为

$$x_i' = \frac{1}{\sigma} \odot (x_i - \mu) \tag{6}$$

其中，$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i$，$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu)^2$ 分别表示原始资产价格序列的均值和方差，$\odot$ 表示哈达玛积。

经过标准化后，使得模型不受到各资产原始数据的统计差异影响。当传统模型 $\mathcal{H}$ 输出结果 $\boldsymbol{y}' = [y_1', y_2', \cdots, y_t']^{\mathrm{T}} \in \mathbb{R}^{m \times k}$ 后，将其与初始的 μ 和 σ 进行反标准化操作，得到最终结果 $\boldsymbol{y} = [y_1, y_2, \cdots, y_t]^{\mathrm{T}} \in \mathbb{R}^{m \times k}$，其中 t 表示预测长度。具体公式表示如下：

$$\begin{cases} y' = \mathcal{H}(x') \\ y_i = \sigma \odot (y_i' + \mu) \end{cases} \tag{7}$$

除增添额外两个标准化模块外，NsT 对于传统的自注意力计算方式进行了改进，以获取到资产数据的时间依赖性。原始的自注意力计算公式为

$$\mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{8}$$

其中，$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{m \times d_k}$ 分别是长度为 m，维度为 d_k 的查询、键和值。假设嵌入层和前馈层 $f(\cdot)$ 保持线性特性，且每个资产序列 x 的方差相同均为 σ。经过标准化后，对于 $\boldsymbol{Q} = [q_1, q_2, \cdots, q_m]^{\mathrm{T}}$，根据

$$x' = \frac{x - E\mu_x^{\mathrm{T}}}{\sigma}$$

$$q_i = f(x_i) \tag{9}$$

其中，$\boldsymbol{E} \in \mathbb{R}^{m \times 1}$ 为全 1 向量。可以得到标准化后的 $\boldsymbol{Q}'$：

$$\boldsymbol{Q}' = \frac{\boldsymbol{Q} - \boldsymbol{E}\mu_Q^{\mathrm{T}}}{\sigma} \tag{10}$$

同理可得，$\boldsymbol{K}' = \frac{K - \boldsymbol{E}\mu_K^{\mathrm{T}}}{\sigma}$，$V' = \frac{V - \boldsymbol{E}\mu_V^{\mathrm{T}}}{\sigma}$。因此标准化后可以得到如下等式：

$$\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right) = \mathrm{Softmax}\left(\frac{\sigma^2\boldsymbol{Q}'\boldsymbol{K}'^{\mathrm{T}} + \boldsymbol{E}\mu_Q^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{11}$$

从等式(11)中可以看出，计算注意力分数时需要用到原始资产序列中的方差 σ、$\boldsymbol{Q}$ 和 $\boldsymbol{K}$。令去稳定因子 $\tau = \sigma^2$，$\Delta = \boldsymbol{K}\mu_Q$，则改进后的注意力计算公式为

$$\text{Attention}(\boldsymbol{Q}',\boldsymbol{K}',\boldsymbol{V}',\tau,\Delta)=\text{Softmax}\left(\frac{\boldsymbol{Q}'\boldsymbol{K}^{\mathrm{T}}+\Delta^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}' \quad (12)$$

其中：$\log\tau=\text{MLP}(\sigma,x)$，$\Delta=\text{MLP}(\mu_x,x)$，MLP(·)为多层感知机模型。通过将原始序列中具有非平稳表现的数学特征融合到反标准模块中，可以恢复对原始序列的非平稳特性。对于每个资产分别通过各自的 NsT 模型进行特征提取，再将得到的各资产特征值进行合并得到 $\boldsymbol{X}_{\text{temp}}$。

2.2 资产间依赖信息提取模块

对于单个资产而言，其价格序列具有一定的波动性。但从市场整体来看，资产行业的整体表现更能够反映一定的经济形式和未来的走势情况。为了提取资产间的依赖性信息，通过 GCN 对输入的资产价格序列进行建模，进行资产间相互作用关系学习。

对于输入的 $n\times m\times k$ 大小的资产价格张量数据，按照特征维度将数据进行重构，得到 k 个图。每个图都可以通过一个邻接矩阵 $\boldsymbol{A}$ 进行表示，即

$$\boldsymbol{A}=\begin{bmatrix}a_{1,1} & a_{1,2} & \cdots & a_{1,n}\\ a_{2,1} & a_{2,2} & \cdots & a_{2,n}\\ \vdots & & \ddots & \vdots\\ a_{n,1} & a_{n,2} & \cdots & a_{n,n}\end{bmatrix}\in\mathbb{R}^{n\times n} \quad (13)$$

$$a_{i,j}=1-\text{corr}(X_i,X_j) \quad (14)$$

其中，$\text{corr}(X_i,X_j)$ 表示资产 i 和资产 j 之间的相关性，本模块的网络架构如图 3 所示。

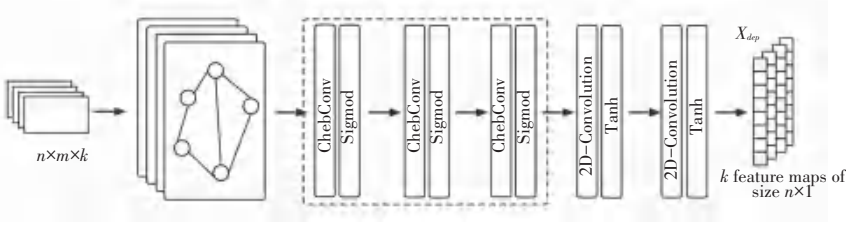

图 3 资产间依赖信息提取模块架构图

Fig. 3 Architecture of asset dependency information extraction module

k 个通道的邻接矩阵 $\boldsymbol{A}$ 首先通过三层图卷积层进行局部领域内资产节点信息的学习，以此来更新当前资产节点的内容。如果将多个图卷积层进行叠加，则可表示为

$$\boldsymbol{H}^{(l+1)}=\sigma(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{A}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}H^{(l)}\boldsymbol{W}^{(l)}) \quad (15)$$

其中，$\boldsymbol{H}^{(l)}$ 表示第 l 层节点的特征矩阵；$\widetilde{A}=\boldsymbol{A}+\boldsymbol{I}$ 表示图的邻接矩阵 $\boldsymbol{A}$ 加上自连接；$\widetilde{\boldsymbol{D}}$ 表示 $\widetilde{A}$ 的度矩阵；$\boldsymbol{W}^{(l)}$ 表示第 l 层的权重矩阵；$\sigma(\cdot)$ 表示激活函数。

经过 GCN 后，再通过两个二维卷积层对各特征维度上的图进行卷积操作，即对每个节点的特征向量进行加权平均，最终得到 k 个资产间依赖信息特征 $\boldsymbol{X}_{\text{dep}}$，具体过程见算法 1。

算法 1 资产间依赖信息提取模块算法

输入 原始资产价格序列 X

输出 资产间依赖信息特征 $X_{\text{dependency}}$

1 calculate the adjacency matrix: A

$a_{i,j}=1-\text{corr}(X_i,X_j)$

2 calculate the degree matrix $\widetilde{\boldsymbol{D}}$ of $\widetilde{A}$, D of $\boldsymbol{A}$

$\widetilde{A}=\boldsymbol{A}+I,\widetilde{D}_{ii}=\sum_j\widetilde{A}_{ij},D_{ii}=\sum_j A_{ij}$

3 calculate the graph Laplacian: $L=D-\boldsymbol{A}$

4 define the weight kernel: $W=\sum_{k=0}^{K}\theta_k\Lambda^k$, where k is the feature diagonal matrix of L

5 graph convolution: $H^{(l+1)}=\sigma(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{A}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}H^{(l)}\boldsymbol{W}^{(l)})$

6 2D-convolution: $X_{dep}Conv2D(H)$

2.3 基于深度强化学习的投资决策模块

为避免本轮决策与上轮差异过大，在决策模块中将上轮决策 w_{t-1} 合并到特征中，可在交易中减小交易费用、降低交易风险。最后对整合后的特征使用 Softmax 函数进行向量归一化，产生本次的投资组合权重向量 $\boldsymbol{w}_t$。每一次的交易过程可由图 4 进行表示。

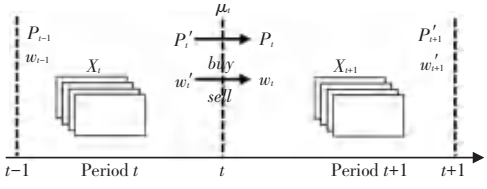

图 4 t 时期到 t+1 时期的交易转化过程

Fig. 4 Trading transformation process from period t to period t+1

在 t 时刻，会对资产进行买入或卖出操作，则对应投资组合权重由 w_t' 变为 w_t。交易过程产生的交易费会导致投资组合价值由原来的 P_t' 变为 P_t，其交易剩余系数为 $\mu_t\in(0,1]$。

2.3.1 环境和智能体

在公式(4)中，令 $close_t=v$，则：

$$\boldsymbol{p}_t := \frac{\boldsymbol{v}_t}{\boldsymbol{v}_{t-1}} = \left(\frac{v_{1,t}}{v_{1,t-1}}, \frac{v_{2,t}}{v_{2,t-1}}, \cdots, \frac{v_{m,t}}{v_{m,t-1}}\right)^{\mathrm{T}} \tag{16}$$

在图 4 中，$\boldsymbol{w}_t'$ 和 $\boldsymbol{w}_{t-1}$ 二者的关系可由下式表示：

$$\boldsymbol{w}_t' = \frac{\boldsymbol{y}_t \odot \boldsymbol{w}_{t-1}}{\boldsymbol{y}_t \cdot \boldsymbol{w}_{t-1}} \tag{17}$$

可理解为，在 t 阶段最初时的投资权重向量为 $\boldsymbol{w}_{t-1}$，由于 t 时段末时市场上资产的价格相较于 t 时刻初时发生了波动，则投资权重会发生相应的转变。

在 $t+1$ 阶段的买卖交易受到上一时段中投资组合权重 $\boldsymbol{w}_t'$ 和 $\boldsymbol{w}_t$ 影响，又因 $\boldsymbol{w}_{t-1}$ 已知，由公式(17)得 $\boldsymbol{w}_t'$ 已知，故 $t+1$ 时的投资只受到 $\boldsymbol{w}_t$ 影响。可将 t 时刻的动作描述为投资权重 $\boldsymbol{w}_t$，即

$$\boldsymbol{a}_t = \boldsymbol{w}_t \tag{18}$$

从图 4 中可以看出，在做出投资决策时，会受到上一阶段的权重 $\boldsymbol{w}_{t-1}$ 影响，本文将其视为环境的一部分对智能体加以约束。因此，在 t 时刻的状态可以通过价格序列 $\boldsymbol{X}_t$ 和上阶段投资权重 $\boldsymbol{w}_{t-1}$ 表示，即

$$\boldsymbol{s}_t = (\boldsymbol{X}_t, \boldsymbol{w}_{t-1}) \tag{19}$$

2.3.2　奖励函数

对于大多数研究，都采用对数收益作为回报函数。假设交易费为 c_t，则奖励函数可以通过对数收益的期望进行表示。

$$R = \mathbb{E}\{\log r_t \times (1-c_t)\} = \frac{1}{T}\sum_{t=1}^{T} \log r_t \times (1-c_t) \tag{20}$$

其中，T 表示交易迭代总数。此奖励函数仅以收益单方面进行约束，忽略了交易后投资组合面临高风险的问题。若令 $r_t' = \log r_t \times (1-c_t)$，将 r_t' 的方差 $\sigma^2(r_t')$ 定义为投资组合风险价值，则此时奖励函数表示为

$$R = \frac{1}{T}\sum_{t=1}^{T} r_t' - \theta\sigma^2(r_t') \tag{21}$$

由图 4 可知，在 t 时刻时会将投资权重由 $\boldsymbol{w}_t'$ 调整为 $\boldsymbol{w}_t$，这会对整体收益产生影响，因此需要将单位交易成本 c_t 加入到奖励函数中。

2.3.3　确定性策略梯度算法

在金融市场中，连续的投资构建可以通过奖励函数获取即时的收益。因此，可以通过梯度下降算法对奖励函数进行优化来获取最优策略。对于采取的策略，通过一组参数 θ 进行描述，对于每次采取的投资策略，如设策略对应描述为 $\pi: S \to A$，则根据当前市场环境可有：

$$\boldsymbol{a}_t = \pi_\theta(s_t) \tag{22}$$

对于策略 π 的性能优劣评价，可以通过相应时间段 $[0, T]$ 内的奖励函数进行比较，即根据投资组合的收益大小来评判策略的好坏，公式描述为

$$F(\pi_\theta) = R \tag{23}$$

因此，若模型的学习率为 η，则参数更新规则为

$$\boldsymbol{\theta} \to \boldsymbol{\theta} + \eta \nabla_\theta F(\pi_\theta) \tag{24}$$

该规则会在每轮迭代中触发，更新 PPFNet 中的各个网络参数。

3　实验

3.1　数据集

本文参照 M6 竞赛[16]给出的资产池，选取实验数据为从 2019-06-10 至 2022-06-10，共有 88 个资产的 699 条数据作为本文的实验数据。其中将 90%的数据用于训练，剩余 10%的数据用于测试。

对于因子集，选取了基本的开盘价、最高价、最低价、收盘价和交易量行情数据。上述所有数据均下载自 Wind 金融终端。

3.2　评价标准

对于投资组合在一段时间内的收益效果，通常采用最直接的累计投资组合价值（*APV*）作为评价标准，其定义如下：

$$APV = W_n = W_0 \Pi_{t=1}^{\mathrm{T}} \boldsymbol{w}_t \boldsymbol{x}_t (1 - c_t) \tag{25}$$

其中，$\boldsymbol{x}_t$ 为资产收盘价的变化量；c_t 为交易费比例；$W_0 = 1$ 为初始投资组合价值。

使用 *APV* 评价仅仅是对于最终收益表现进行了量化，忽略对风险的考虑。对于那些关注交易过程的投资者，投资组合风险调整后的收益是十分重要的[17]。夏普比率（*SR*）正是用来评估这一方面的指标，含义为在承受每一单位风险的情况下，投资组合能带来的多少超额收益。

$$SR = \frac{APV - R_f}{\sigma_p} \tag{26}$$

其中，R_f 为无风险资产的收益率，σ_p 为投资组合日收益率的年化标准差。此外，对于下行波动的评估通常使用最大回撤（*MDD*）进行表示：

$$MDD = \max_{\tau > t} \frac{W_t - W_\tau}{W_t} \tag{27}$$

其中，W_t 为收益最大值，W_τ 为 t 时刻后的收益最小值。在 *APV* 和 *MDD* 的基础上，卡玛比率（*CR*）进一步给出了投资组合收益与最大回撤之间的关系，即：

$$CR = \frac{APV}{MDD} \tag{28}$$

3.3　实验结果

在实验中，将本文模型 PPFNet 与目前的常用方法

进行了比较,包括 EIIE[3]、RMR[18]、WMAMR[19]以及统一买入并持有(UBAH)和平均连续再平衡投资(UCRP)策略。具体实验对比详细结果见表 1。

表 1 不同模型的表现对比结果

Table 1 Comparison results of performance of different models

Model	*APV*	*SR*	*CR*
UBAH	1.43	0.040	1.09
UCRP	1.23	0.038	0.98
RMR	5.36	0.065	9.74
WMAMR	2.61	0.083	17.45
EIIE	20.12	0.075	22.24
PPFNet	**80.35**	**0.101**	**264.19**

在使用到的基准模型当中,RMR 和 WMARM 均为基于均值回归理论研发的模型,UBAH 和 UCRP 则是量化投资中的常见策略,而 EIIE 和 PPFNet 方法则是基于强化学习构建投资组合。

可以看出,基于强化学习的方法在各指标中均优于其他基准模型,本文的方法 PPFNet 优于基准的 EIIE 模型,说明 PPFNet 中的各模块提取的特征信息更利于构建有效的投资组合策略。

除对比实验外,本文还设计了消融实验,以验证各模块的效果,效果见表 2。

表 2 消融实验对比结果

Table 2 Comparison results of ablation experiments

Model	*APV*	*SR*	*CR*
PPFNet-up	66.35	0.072	171.17
PPFNet-down	47.76	0.081	54.93
PPFNet	**80.35**	**0.101**	**264.19**

其中,PPFNet-up 表示只使用时间序列信息模块,PPFNet-down 为只是用资产依赖信息模块。从表 2 中可以看出,在 *APV* 上 PPFNet 比 PPFNet-up 高出约 17%,但 *CR* 却高出了约 35%,说明资产依赖性信息提取模块起到了控制风险的作用。同时,从 PPFNet-down 与 PPFNet 的对比来看,PPFNet 表现出更佳的收益效果。

4 结束语

本文基于深度强化学习中的 DDPG 算法,设计了一种并行特征提取网络 PPFNet 作为策略网络用于构建投资组合。通过与其他投资组合构建方法进行对比分析,证明了本文方法具有良好的收益效果。在接下来的工作中,可以将文本信息对资产价格序列的影响考虑到模型设计当中,使模型的泛化性能更佳。

参考文献

[1] HEATON J B, POLSON N G, WITTE J H. Deep learning for finance: Deep portfolios [J]. Applied Stochastic Models in Business and Industry, 2017, 33(1): 3-12.

[2] TENG X, ZHANG X, LUO Z. Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction [J]. Neurocomputing, 2022, 505: 92-100.

[3] JIANG Z, XU D, LIANG J. A deep reinforcement learning framework for the financial portfolio management problem [J]. arXiv preprint arXiv:1706.10059, 2017.

[4] BERNER C, BROCKMAN G, CHAN B, et al. Dota 2 with large scale deep reinforcement learning[J]. arXiv preprint arXiv:1912.06680, 2019.

[5] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.

[6] MOSAVI A, FAGHAN Y, GHAMISI P, et al. Comprehensive review of deep reinforcement learning methods and applications in economics[J]. Mathematics, 2020, 8(10): 1640.

[7] BACOYANNIS V, GLUKHOV V, JIN T, et al. Idiosyncrasies and challenges of data driven learning in electronic trading [J]. arXiv preprint arXiv:1811.09549, 2018.

[8] 齐岳, 黄硕华. 基于深度强化学习 DDPG 算法的投资组合管理[J]. 计算机与现代化, 2018, 5: 93-99.

[9] 王康, 白迪. 基于深度强化学习的投资组合管理研究[J]. 现代计算机, 2021, 1(3): 11.

[10] JANG J, SEONG N Y. Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory [J]. Expert Systems with Applications, 2023: 119556.

[11] YUE H, LIU J, TIAN D, et al. A Novel anti-risk method for portfolio trading using deep reinforcement learning [J]. Electronics, 2022, 11(9): 1506.

[12] DGHAIS A A A, ISMAIL M T. A study of stationarity in time series by using wavelet transform [J]. American Institute of Physics, 2014, 1605(1): 798-804.

[13] NEELY C J, RAPACH D E, TU J, et al. Forecasting the equity risk premium: The role of technical indicators[J]. Management Science, 2014, 60(7): 1772-1791.

[14] SCHMITT T A, CHETALOVA D, SCHÄFER R, et al. Non-stationarity in financial time series: Generic features and tail behavior[J]. Europhysics Letters, 2013, 103(5): 58003.

[15] ZHU H, LIU S Y, ZHAO P, et al. Forecasting asset dependencies to reduce portfolio risk[C]//Proceedings of the AAAI Conference on Artificial Intelligence. IEEE, 2022, 36(4): 4397-4404.

[16] SOURSOSA. The M6 Financial Forecasting Competition [EB/OL].https://m6competition.com/.

[17] JAFARZADEH H, AKBARI P, ABEDINB. A methodology for project portfolio selection under criteria prioritisation, uncertainty and projects interdependency - combination of fuzzy QFD and DEA[J]. Expert Systems with Applications, 2018, 110: 237-249.

[18] HUANG D, ZHOU J, LI B, et al. Robust median reversion strategy for online portfolio selection[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9): 2480-2493.

[19] GAO L, ZHANG W. Weighted moving average passive aggressive algorithm for online portfolio selection[C]//Proceedings of 2013 5th International Conference on Intelligent Human - Machine Systems and Cybernetics. IEEE, 2013: 327-330.