Глубокое обучение с подкреплением для автоматизированной торговли акциями: Ансамблевая стратегия

Хоньян Ян[1] , Сяо-Ян Лю[2] , Шань Чжун[2] , Анвар Валид [31] Кафедра статистики, Колумбийский университет
[2]Кафедра электротехники, Колумбийский университет[3] Отдел исследования математических систем, Nokia-Bell Labs Email: {HY2500, XL2427, SZ2495}@columbia.edu, anwar.walid@nokia-bell-labs.com

АннотацияСтратегии торговли *акциями* играют важную роль в инвестировании. Однако разработать прибыльную стратегию в условиях сложного и динамичного фондового рынка довольно сложно. В этой статье мы предлагаем ансамблевую стратегию, которая использует схемы глубокого подкрепления для обучения стратегии торговли акциями, максимизируя доходность инвестиций. Мы обучаем агента глубокого обучения с подкреплением и получаем ансамблевую торговую стратегию, используя три алгоритма, основанные на критических оценках агентов: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C) и Deep Deterministic Policy Gradient (DDPG). Ансамблевая стратегия наследует и объединяет лучшие черты трех алгоритмов, тем самым надежно приспосабливаясь к различным рыночным ситуациям. Чтобы избежать больших затрат памяти при обучении сетей с непрерывным пространством действий, мы используем технику загрузки по требованию для обработки очень больших данных. Мы тестируем наши алгоритмы на 30 акциях Dow Jones, которые имеют достаточную ликвидность. Эффективность торгового агента с различными алгоритмами обучения с подкреплением оценивается и сравнивается как с индексом Dow Jones Industrial Average, так и с традиционной стратегией распределения портфеля с минимальной дисперсией. Показано, что предложенная стратегия глубокого ансамбля превосходит три отдельных алгоритма и две базовые стратегии по доходности с поправкой на риск, измеряемой коэффициентом Шарпа.

***Индексные термины - глубокое* обучение с подкреплением, марковский процесс принятия решений, автоматическая торговля акциями, ансамблевая стратегия, система критики акторов**

I. Введение

Прибыльная автоматизированная стратегия торговли акциями жизненно важна для инвестиционных компаний и хедж-фондов. Она применяется для оптимизации распределения капитала и максимизации эффективности инвестиций, например ожидаемой доходности. Максимизация прибыли может быть основана на оценке потенциальной доходности и риска. Однако аналитикам сложно учесть все значимые факторы на сложном и динамичном фондовом рынке [1], [2], [3].

Существующие работы не являются удовлетворительными. Традиционный подход, состоящий из двух шагов, был описан в [4]. Сначала рассчитывается ожидаемая доходность акций и ковариационная матрица цен акций. Затем можно получить оптимальную стратегию распределения портфеля, либо максимизируя доходность при заданном соотношении рисков, либо минимизируя риск при заданной доходности. Однако этот подход сложен и дорог в реализации, поскольку управляющие портфелем могут захотеть пересмотреть решения на каждом временном шаге и учесть другие факторы, такие как стоимость сделки. Другой подход для

максимизировать доходность.

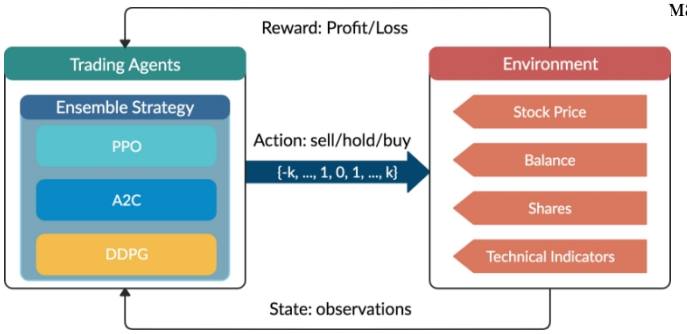

Рис. 1. Обзор стратегии торговли акциями на основе обучения с подкреплением.

Фондовая торговля моделируется как марковский процесс принятия решений (MDP) и с помощью динамического программирования выводится оптимальная стратегия [5], [6], [7], [8]. Однако масштабируемость этой модели ограничена из-за большого пространства состояний при работе с фондовым рынком.

В последние годы алгоритмы машинного обучения и глубокого обучения широко применяются для построения моделей прогнозирования и классификации на финансовом рынке. Фундаментальные данные (отчет о прибылях) и альтернативные данные (новости рынка, данные академических графиков, транзакции по кредитным картам, трафик GPS и т. д.) комбинируются с алгоритмами машинного обучения для извлечения новых инвестиционных альфа-сигналов или прогнозирования будущих показателей компании [9], [10], [11], [12]. Таким образом, генерируется прогностический альфа-сигнал для выбора акций. Однако эти подходы направлены только на выбор высокоэффективных акций, а не на распределение торговых позиций или долей между выбранными акциями. Другими словами, модели машинного обучения не обучены моделированию позиций. В этой статье мы предлагаем новую ансамблевую стратегию, которая объединяет три алгоритма глубокого обучения с подкреплением и находит оптимальную торговую стратегию на сложном и динамичном фондовом рынке. Три алгоритма акторной критики [13], Proximal Policy Optimization (PPO) [14], [15], Advantage Actor Critic (A2C) [16], [17] и Deep Deter- ministic Policy Gradient (DDPG) [18], [15], [19]. Наш подход к глубокому обучению с подкреплением описан на рисунке 1. Применяя ансамблевую стратегию, мы делаем торговую стратегию более надежной и устойчивой. Наша стратегия может приспосабливаться к различным рыночным ситуациям и

с ограничением риска. Во-первых, мы создаем среду и определяем пространство действий, пространство состояний и функцию вознаграждения. Во-вторых, мы обучаем три алгоритма, которые выполняют действия в среде. В-третьих, мы объединяем три агента в ансамбль, используя коэффициент Шарпа, который измеряет доходность, скорректированную на риск. Эффективность ансамблевой стратегии подтверждается более высоким коэффициентом Шарпа, чем у стратегии распределения портфеля с минимальной дисперсией и промышленного индекса Доу-Джонса[1] (DJIA).

Остальная часть данной работы организована следующим образом. В разделе 2 представлены смежные работы. В разделе 3 приводится описание нашей задачи биржевой торговли. В разделе 4 мы создаем среду для торговли акциями. В разделе 5 мы описываем и уточняем три алгоритма, основанные на акторной критике, и нашу стратегию ансамбля. В разделе 6 описана предварительная обработка биржевых данных и наша экспериментальная установка, а также представлена оценка эффективности предложенной ансамблевой стратегии. В разделе 7 мы подводим итоги данной работы.

II. Похожие работы

Недавние приложения глубокого обучения с подкреплением на финансовых рынках рассматривают дискретные или непрерывные пространства состояний и действий и используют один из этих подходов к обучению: подход, основанный только на критике, подход, основанный только на акторах, или подход, основанный на критике акторов [20]. Модели обучения с непрерывным пространством действий обеспечивают более тонкие возможности управления, чем модели с дискретным пространством действий.

Наиболее распространенный подход, основанный на обучении только на критике, решает дискретную задачу пространства действий, используя, например, Deep Q-learning (DQN) и его усовершенствования, и обучает агента на одной акции или активе [21], [22], [23]. Идея подхода, основанного только на критике, заключается в использовании функции Q-значения для обучения оптимальной политике выбора действий, которая максимизирует ожидаемое будущее вознаграждение с учетом текущего состояния. Вместо того чтобы вычислять таблицу значений состояния и действия, DQN минимизирует ошибку между оцененной Q-ценностью и целевой Q-ценностью на переходе и использует нейронную сеть для выполнения аппроксимации функции. Основное ограничение подхода, основанного только на критике, заключается в том, что он работает только с дискретными и конечных пространств состояний и действий, что нецелесообразно для большой портфель акций, поскольку цены, конечно, непрерывны.

Подход, основанный на использовании только агентов, применялся в [24], [25], [26]. Идея заключается в том, что агент непосредственно сам обучается оптимальной политике. Вместо того чтобы нейронная сеть учила Q-значение, нейронная сеть учит политику. Политика - это распределение вероятностей, которое, по сути, является стратегией для данного состояния, а именно вероятностью

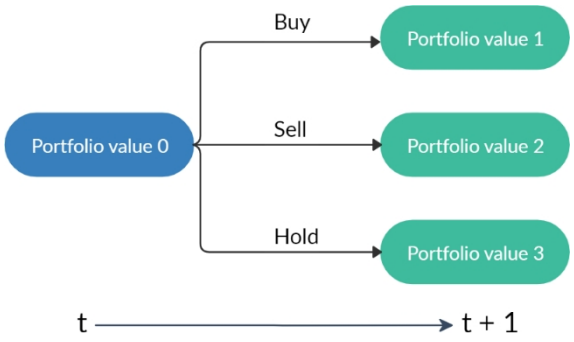

Рис. 2. Начальное значение портфеля при трех действиях приводит к трем возможным портфелям. Обратите внимание, что действие "держать" может привести к различным значениям портфеля из-за изменения цен на акции.

обновляют сеть акторов, представляющую политику, и сеть критиков, представляющую функцию ценности. Критик оценивает функцию ценности, а актор обновляет распределение вероятностей политики, руководствуясь критикой и градиентами политики. Со временем актор учится предпринимать более эффективные действия, а критик лучше оценивает эти действия. Подход "актер-критик" доказал свою способность обучаться и адаптироваться к большим и сложным средам, и был использован для популярных видеоигр, таких как Doom [29]. Таким образом, подход "актор-критик" перспективен для торговли с большим портфелем акций.

III. Описание проблемы

Мы моделируем торговлю акциями как марковский процесс принятия решений (MDP) и формулируем нашу торговую цель как максимизацию ожидаемой доходности [30].

A. MDP-модель для торговли акциями

Для моделирования стохастической природы динамичного фондового рынка мы используем марковский процесс принятия решений (MDP):

- Состояние $\boldsymbol{s} = [\boldsymbol{p}, \boldsymbol{h}, b]$: вектор, включающий цены на акции $\boldsymbol{p} \in \mathbb{R}^D$, акции $\boldsymbol{h} \in \mathbb{Z}^D$, а остальные

предпринять разрешенное действие. Рекуррентное обучение с подкреплением было введено, чтобы избежать проклятия размерности и повысить эффективность торговли в [24]. Подход, основанный только на акторах, может работать с непрерывными средами пространства действий.

Акторно-критический подход был недавно применен в финансовой сфере [27], [28], [17], [19]. Идея заключается в том, чтобы одновременно

1Промышленный индекс Доу-Джонса - это индекс фондового рынка, который показывает, как торговались акции 30 крупных публичных компаний, расположенных в США, в течение стандартной торговой сессии на фондовом рынке.

баланс $b \in \mathbf{R}_+$, где D обозначает количество акций, а $\mathbf{Z}_+$ - неотрицательные целые числа.

- Действие $\boldsymbol{a}$: вектор действий над D акциями. Сайт Разрешенные действия с каждой акцией включают *продажу*, *покупку* или *удержание*, которые приводят к уменьшению, увеличению и отсутствию изменений в акциях $\boldsymbol{h}$, соответственно.
- Вознаграждение $r(s, a, s')$: прямое вознаграждение за выполнение действия a в состоянии s и прибытие в новое состояние s'.
- Политика $\pi(s)$: торговая стратегия в состоянии s, которая представляет собой вероятностное распределение действий в состоянии s.
- Q-значение $Q_\pi(s, a)$: ожидаемое вознаграждение от выполнения действия a в состоянии s, следуя политике π.

Переход состояний в процессе торговли акциями показан на рисунке 2. В каждом состоянии с акцией d ($d = 1, ..., D$) в портфеле совершается одно из трех возможных действий.

- Продажа акций $\boldsymbol{k}[d] \in [1, \boldsymbol{h}[d]]$ приводит к тому, что $\boldsymbol{h}_{t+1}[d] = \boldsymbol{h}_t[d] - \boldsymbol{k}[d]$, где $\boldsymbol{k}[d] \in \mathbf{Z}_+$ и $d = 1, ..., D$.

- Холдинг, $h_{t+1}[d] = h_t[d]$.
- Покупка акций $k[d]$ приводит к тому, что $h_{t+1}[d] = h_t[d] + k[d]$.

В момент времени t происходит действие, и цены на акции обновляются в момент $t+1$, соответственно, стоимость портфеля может измениться с "стоимости портфеля 0" на "стоимость портфеля 1", "стоимость портфеля
2" или "стоимость портфеля 3", соответственно, как показано н а рисунке 2. Обратите внимание, что стоимость портфеля равна $p^T h + b$.

B. Учет ограничений на торговлю акциями

Следующие допущения и ограничения отражают практические рекомендации: транзакционные издержки, ликвидность рынка, риск.
отвращение и т.д.

- Ликвидность рынка: ордера могут быть быстро исполнены по цене закрытия. Мы предполагаем, что фондовый рынок не будет подвержен влиянию нашего торгового агента с подкреплением.
- Неотрицательный баланс $b \geq 0$: разрешенные действия не должны приводить к отрицательному балансу. Исходя из действий в момент времени t, запасы делятся на наборы для продажи S, покупки B и удержания H, где $S \cup B \cup H = \{1, \ldots, D\}$ и они не пересекаются. Пусть $p^B = [p^i : i \in B]$ и $k_t^B = [k_t^i : i \in B]$ - векторы цены и количества покупаемых акций для акций в S S набор покупателей. Аналогичным образом мы можем определить p_t и k_t для
для продающих акций, и p^H и k_t^H для держателей. запасы. Следовательно, ограничение на неотрицательный баланс может быть выражено как

$$b_{t+1} = b_t + (p_t)^{ST} k_t^S - (p_t)^{BT} k_t^B \geq 0. \quad (1)$$

- Транзакционные издержки: транзакционные издержки возникают при каждой сделке. Существует множество видов транзакционных издержек, таких как биржевые сборы, сборы за исполнение и сборы SEC. У разных брокеров разные комиссионные сборы. Несмотря на эти различия в комиссионных, мы предполагаем, что наши транзакционные издержки составляют 0,1 % от стоимости каждой сделки (покупки или продажи), как в [9]:

$$c_t = p^T k_t \times 0{,}1\%. \quad (2)$$

- Неприятие риска краха рынка: существуют внезапные события, которые могут вызвать крах фондового рынка, такие как войны, схлопывание пузырей на фондовом рынке, дефолт по суверенному долгу и финансовый кризис. Чтобы контролировать риск при наихудшем сценарии, таком как мировой финансовый кризис 2008 года,

C. Максимизация прибыли как цель торговли

Мы определяем нашу функцию вознаграждения как изменение стоимости портфеля при выполнении действия a в состоянии s и приходе в новое состояние s' . Цель состоит в том, чтобы разработать торговую стратегию, которая максимизирует изменение стоимости портфеля:

$$r(s_t, a_t, s_{t+1}) = (b_{t+1} + p_{t+1}^T h_{t+1}) - (b_t + p^T h_t) - c_t, \quad (4)$$

где первый и второй члены обозначают стоимость портфеля на момент $t+1$ и t, соответственно. Для дальнейшей декомпозиции доходности мы определяем переход акций h_t как

$$h_{t+1} = h_t - k_t^S + k_t^B, \quad (5)$$

и переход равновесия b_t определяется в (1). Тогда (4) можно переписать как

$$r(s_t, a_t, s_{t+1}) = r_H - r_S + r_B - c_t, \quad (6)$$

где

$$r_H = (p_{t+1}^H - p_t^H)^T h_t^H, \quad (7)$$

$$r_S = (p_t^S - p_{t+1}^S)^T h_t^S, \quad (8)$$

$$r_B = (p_{t+1}^B - p_t^B)^T h_t^B, \quad (9)$$

где r_H , r_S , и r_B обозначают изменение портфеля.

мы используем индекс финансовой *турбулентности*$_t$, который измеряет экстремальные движения цен на активы [31]:

$$\text{турбулентность}_t = (y_t - \mu)\, \Sigma^{-1}\, (y_t - \mu)' \in \mathbb{R}, \quad (3)$$

стоимость от владения, продажи и покупки акций при переходе от момента t к $t + 1$, соответственно. Уравнение (6) показывает, что мы должны максимизировать положительное изменение стоимости портфеля, покупая и удерживая акции, цена которых будет расти на следующем временном шаге, и минимизировать отрицательное изменение стоимости портфеля, продавая акции, цена которых будет снижаться на следующем временном шаге.

Индекс *турбулентности*$_t$ включается в функцию вознаграждения, чтобы учесть наше неприятие риска краха рынка. Когда индекс в (3) превышает пороговое значение,

где $\boldsymbol{y_t} \in \mathbf{R}^D$ обозначает доходность акций за текущий период t, $\boldsymbol{\mu} \in \mathbf{R}^D$ - среднее значение исторической доходности, а $\boldsymbol{\Sigma} \in \mathbf{R}^{D\times D}$ - ковариация исторической доходности. Когда *турбулентность*$_t$ превышает пороговое значение, что указывает на экстремальные рыночные условия, мы просто прекращаем покупки, и торговый агент продает все акции. Мы возобновляем торговлю, как только индекс турбулентности возвращается ниже порогового значения.

Уравнение (8) становится

$$r_{sell} = (\boldsymbol{p_{t+1}} - \boldsymbol{p})_t^T \boldsymbol{k_t}, \tag{10}$$

что говорит о том, что мы хотим минимизировать отрицательное изменение стоимости портфеля, продав все имеющиеся акции, поскольку цены всех акций упадут.

Модель инициализируется следующим образом. p_0 устанавливается на цены акций в момент времени 0, а b_0 - размер начального фонда. h и $Q_\pi(s, a)$ равны 0, а $\pi(s)$ равномерно распределено между всеми действиями для каждого состояния. Затем, $Q_\pi(s_t, a_t)$ обновляется через взаимодействие с фондовым рынком. Оптимальная стратегия задается уравнением Беллмана таким образом, что ожидаемое вознаграждение от принятия действия a_t в состоянии s_t

это ожидание суммы прямого вознаграждения $r(s_t, a_t, s_{t+1})$ и будущего вознаграждения в следующем состоянии s_{t+1} . Пусть для сходимости будущие вознаграждения дисконтируются с коэффициентом $0 < \gamma < 1$, тогда мы имеем

$$Q_\pi(s_t, a_t) = \mathbf{E}_{s\,t+1}[r(s_t, a_t, s_{t+1}) + \gamma \mathbf{E}_{a\,t+1 \sim \pi(\mathrm{st+1})}[Q_\pi(s_{t+1}, a_{t+1})]]. \tag{11}$$

Цель - разработать торговую стратегию, которая максимизирует положительное кумулятивное изменение портфеля.

значение $r(s_t, a_t, s_{t+1})$ в динамической среде, и мы используем метод глубокого обучения с подкреплением для решения этой задачи.

IV. Состояние фондового рынка

Перед обучением торгового агента с глубоким подкреплением мы тщательно создаем среду для симуляции реальной торговли, которая позволяет агенту осуществлять взаимодействие и обучение. В практической торговле необходимо учитывать различную информацию, например, исторические цены на акции, текущее владение акциями, технические индикаторы и т. д. Нашему торговому агенту необходимо получать такую информацию из окружающей среды и выполнять действия, описанные в предыдущем разделе. Для реализации среды и обучения агента мы используем тренажерный зал OpenAI [32], [33], [34].

A. Среда для нескольких акций

Мы используем непрерывное пространство действий для моделирования торговли несколькими акциями. Мы предполагаем, что в нашем портфеле всего 30 акций.

1) Пространство состояний: Мы используем 181-мерный вектор, состоящий из семи частей информации, для представления пространства состояний в среде торговли несколькими акциями: $[b_t, \boldsymbol{p}_t, \boldsymbol{h}_t, \boldsymbol{M}_t, \boldsymbol{R}_t, \boldsymbol{C}_t, \boldsymbol{X}_t]$. Каждый компонент определяется следующим образом:

- $b_t \in \mathbb{R}_+$: доступный баланс на текущем шаге времени t.
- $\boldsymbol{p}_t \in \mathbb{R}_+^{30}$: скорректированная цена закрытия каждой акции.
- $\boldsymbol{h}_t \in \mathbb{Z}_+^{30}$: акции, принадлежащие каждой акции.
- $\boldsymbol{M}_t \in \mathbb{R}^{30}$: Moving Average Convergence Divergence (MACD) рассчитывается по цене закрытия. MACD - один из наиболее часто используемых индикаторов импульса, который определяет скользящие средние [35].
- $\boldsymbol{R}_t \in \mathbb{R}^{30}$: Индекс относительной силы (RSI) рассчитывается по цене закрытия. RSI оценивает степень недавнего изменения цены. Если цена движется вокруг линии поддержки, это указывает на перепроданность акций, и мы можем совершить покупку. Если цена движется вокруг линии сопротивления, это указывает на перекупленность акций, и мы можем совершать действия на продажу". [35].
- $\boldsymbol{C}_t \in \mathbb{R}^{30}$: Индекс товарного канала (CCI) рассчитывается на основе цены максимума, минимума и закрытия. CCI сравнивает текущую цену со средней ценой за временной промежуток, чтобы указать на действия по покупке или продаже [36].
- $\boldsymbol{X}_t \in \mathbb{R}^{30}$: Индекс средней направленности (ADX) рассчитывается с помощью цены максимума, минимума и закрытия. ADX определяет силу тренда, количественно оценивая объем движения цены [37].

2) Пространство действий: Для одной акции пространство действий определяется как $\{-k, \ldots, -1, 0, 1, \ldots, k\}$, где k и $-k$ - количество акций, которые мы можем купить и продать, и $k \leq h_{max}$, а h_{max} - предопределенный параметр, который задает максимальное количество акций для каждого действия покупки. Таким образом, размер всего пространства действий составляет $(2k + 1)^{30}$. Затем пространство действий нормируется на $[-1, 1]$, поскольку RL-алгоритмы A2C и PPO определяют политику непосредственно на гауссовом распределении, которое должно быть нормированным и симметричным [34].

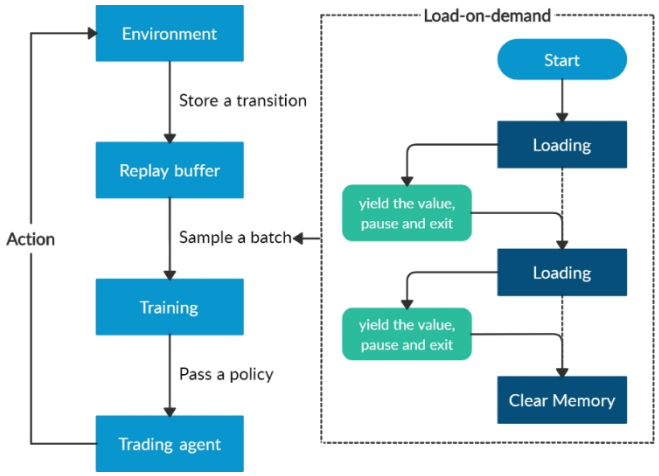

Рис. 3. Обзор метода "нагрузка по требованию".

B. Управление памятью

Потребление памяти для обучения может расти в геометрической прогрессии с увеличением количества запасов, типов данных, особенностей пространства состояний, количества слоев и нейронов в нейронных сетях, а также размера партии. Чтобы решить проблему нехватки памяти, мы используем технику загрузки по требованию для эффективного использования памяти. Как показано на рисунке 3, метод загрузки по требованию не хранит все результаты в памяти, а генерирует их по запросу. Память используется только тогда, когда результат запрашивается, поэтому потребление памяти снижается.

V. Торговый агент на основе глубокого усиления Обучение

Для реализации торгового агента мы используем три алгоритма, основанных на акторной критике. Эти три алгоритма - A2C, DDPG и PPO, соответственно. Предлагается стратегия ансамбля, объединяющая три агента для создания надежной торговой стратегии.

A. Advantage Actor Critic (A2C)

A2C [16] - типичный алгоритм акторной критики, и мы используем его в качестве компонента ансамблевой стратегии. A2C введен для улучшения обновления градиента политики. A2C использует функцию преимущества для уменьшения дисперсии градиента политики. Вместо того чтобы оценивать только функцию ценности, критическая сеть оценивает функцию преимущества. Таким образом, оценка действия зависит не только от того, насколько оно хорошо, но и от того, насколько лучше оно может быть. Это уменьшает высокую дисперсию сети политики и делает модель более устойчивой.

A2C использует копии одного и того же агента для обновления градиентов с разными выборками данных. Каждый агент работает независимо и взаимодействует с одной и той же средой. На каждой итерации, после того как все агенты завершают вычисление своих градиентов, A2C использует координатор для передачи средних градиентов по всем агентам в глобальную сеть. Таким образом, глобальная сеть может обновить агента и сеть критиков. Сайт

Наличие глобальной сети увеличивает разнообразие обучающих данных. Синхронизированное обновление градиента более экономично, быстрее и лучше работает с большими объемами партий. A2C - отличная модель для биржевой торговли, поскольку она

стабильность.

Целевая функция для A2C имеет вид:

$$\nabla J_\theta(\vartheta) = \mathbf{E}[\sum_{t=1} \nabla_\theta \log \pi_\theta(a_t | s_t) A(s_t, a_t)], \quad (12)$$

где $\pi_\theta(a_t | s_t)$ - сеть политики, $A(s_t, a_t)$ - функция преимущества, которая может быть записана как:

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t), \quad (13)$$

или

$$A(s_t, a_t) = r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1}) - V(s_t). \quad (14)$$

B. Глубокая детерминированная политика градиента (DDPG)

DDPG [18] используется для стимулирования максимальной отдачи от инвестиций. DDPG сочетает в себе основы Q-обучения
[38] и градиент политики [39], а также использует нейронные сети в качестве аппроксиматоров функций. В отличие от DQN, которая обучается косвенно через таблицы Q-значений и страдает от проблемы проклятия размерности [40], DDPG обучается непосредственно из наблюдений через градиент политики. Предлагается детерминированно отображать состояния на действия, чтобы лучше соответствовать непрерывной среде пространства действий.

На каждом временном шаге агент DDPG выполняет действие $_t$ на s_t, получает вознаграждение r_t и прибывает на s_{t+1}.

переходы (s_t, a_t, s_{t+1}, r_t) сохраняются в буфере воспроизведения.

R. Из R берется партия из N переходов, и Q-значение y_i обновляется как:

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} | \theta^{\mu}), \vartheta^{Q}), i = 1, \ldots, N. \quad (15)$$

Затем критическая сеть обновляется путем минимизации функции потерь $L(\vartheta^Q)$, которая представляет собой ожидаемую разницу между выходами целевой критической сети Q' и критической сети Q, т.е,

$$L(\theta^Q) = E_{s_t, a_t, r_t, s_{t+1} \sim \text{буфер}}[(y_i - Q(s_t, a_t | \vartheta^Q))^2]. \quad (16)$$

DDPG эффективно работает с непрерывным пространством действий, поэтому он подходит для биржевой торговли.

Обрезанная суррогатная целевая функция PPO имеет вид:

$$J^{\text{CLIP}}(\vartheta) = E\hat{}_t[\min(r_t(\vartheta)\hat{A}(s_t, a_t), \text{clip}(r_t(\vartheta), 1 - \epsilon, 1 + \epsilon)\hat{A}(s_t, a_t))], \quad (18)$$

где $r_t(\vartheta)\hat{A}(s_t, a_t)$ - нормальный градиентный объектив политики, а $\hat{A}(s_t, a_t)$ - оценочная функция преимущества. Функция $clip(r_t(\vartheta), 1 - \epsilon, 1 + \epsilon)$ ограничивает отношение $r_t(\vartheta)$, чтобы оно находилось в пределах $[1 - \epsilon, 1 + \epsilon]$. В качестве объективной функции PPO берется минимум обрезанной и нормальной цели. PPO препятствует большим изменениям политики, выходящим за пределы обрезанного интервала. Таким образом, PPO повышает стабильность обучения сети политик, ограничивая обновление политики на каждом шаге обучения. Мы выбрали PPO для биржевой торговли, потому что он стабилен, быстр, прост в реализации и настройке.

D. Стратегия ансамбля

Наша цель - создать высоконадежную торговую стратегию. Поэтому мы используем ансамблевую стратегию для автоматического выбора агента с наилучшими показателями среди PPO, A2C и DDPG для торговли на основе коэффициента Шарпа. Процесс ансамбля описывается следующим образом:

Шаг 1. Мы используем растущее окно в n месяцев для одновременного переобучения трех агентов. В данной работе мы переобучаем трех агентов каждые три месяца.

Шаг 2. Мы проверяем всех трех агентов, используя скользящее окно проверки за 3 месяца после окна обучения, чтобы выбрать агента с наилучшими показателями и наибольшим коэффициентом Шарпа [42]. Коэффициент Шарпа рассчитывается следующим образом:

$$\text{Коэффициент Шарпа} = \frac{\bar{r}_p - r_f}{\sigma_p}, \quad (19)$$

где $\bar{r}_p$ - ожидаемая доходность портфеля, r_f - безрисковая ставка, а σ_p - стандартное отклонение портфеля. Мы
На этапе проверки мы также скорректировали неприятие риска с помощью индекса турбулентности.

Шаг 3. После выбора лучшего агента мы используем его для прогнозирования и торговли на следующий квартал.

Причина такого выбора заключается в том, что каждый торговый агент чувствителен к разным типам трендов. Один агент хорошо работает при бычьем тренде, но плохо - п р и медвежьем.
Другой агент более приспособлен к изменчивому рынку. Сайт

C. Оптимизация проксимальной политики (PPO)

Мы исследуем и используем PPO в качестве компонента

ансамблевого метода. PPO [14] был введен для управления обновлением градиента политики и обеспечения того, чтобы новая политика не слишком отличалась от предыдущей. PPO пытается упростить задачу оптимизации политики доверительной области (TRPO), вводя в объективную функцию член обрезания [41], [14].

Предположим, что отношение вероятностей между старой и новой политикой выражается как:

$$r(\vartheta) = \frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{старый}}(a_t | s_t)} \quad (17)$$

Чем выше коэффициент Шарпа агента, тем выше его доходность по отношению к величине инвестиционного риска, который он принял. Поэтому мы выбираем торгового агента, который может максимизировать доходность с учетом возрастающего риска.

VI. Оценки эффективности

В этом разделе мы представим оценку эффективности предложенной нами схемы. Мы провели бэктестирование для трех отдельных агентов и нашей ансамблевой стратегии. Результаты, представленные в таблице 2, показывают, что наша ансамблевая стратегия достигает более высокого коэффициента Шарпа, чем три агента, промышленный индекс Доу-Джонса и традиционная стратегия распределения портфеля с минимальной дисперсией.

Наши коды доступны на Github[2] .

[2]Ссылка: https://github.com/AI4Finance-LLC/Deep-Reinforcement-Обучение для автоматизированной торговли акциями - ансамблевая стратегия - ICAIF-2020

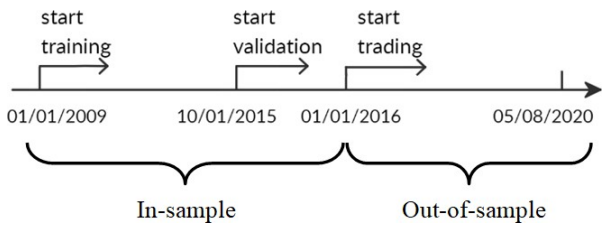

Рис. 4. Разделение данных о запасах.

A. Предварительная обработка данных о запасах

В качестве пула торговых акций мы выбираем акции, входящие в состав Dow Jones 30 (по состоянию на 01/01/2016). **Для** оценки эффективности в наших бэктестах используются исторические ежедневные данные с 01/01/2009 по 05/08/2020. Данные по акциям могут быть загружены из базы данных Compustat через Wharton Research Data Services (WRDS) [43]. Наш набор данных состоит из двух периодов: период в выборке и период вне выборки. Период in-sample содержит данные для этапов обучения и проверки. Период вне выборки содержит данные для этапа торговли. На этапе обучения мы обучаем трех агентов, используя PPO, A2C и DDPG, соответственно. Затем проводится этап проверки, на котором проверяются три агента по коэффициенту Шарпа и настраиваются ключевые параметры, такие как скорость обучения, количество эпизодов и т. д. Наконец, на этапе торговли мы оцениваем прибыльность каждого из алгоритмов.

Весь набор данных разбит на части, как показано на рисунке 4. Данные с 01/01/2009 по 09/30/2015 используются для обучения, а данные с 10/01/2015 по 31/2015 - для проверки и настройки параметров. Наконец, мы проверяем работу нашего агента на торговых данных, которые представляют собой невидимые данные из выборки с 01/01/2016 по 05/08/2020. Чтобы лучше использовать торговые данные, мы продолжаем обучение нашего агента на этапе торговли, поскольку это поможет ему лучше адаптироваться к динамике рынка.

B. Сравнение производительности

1) Выбор агента: Из таблицы 1 мы видим, что PPO имеет наилучший валидационный коэффициент Sharpe 0,06 в период с 2015/10 по 2015/12, поэтому мы используем PPO для торговли в следующем квартале с 2016/01 по 2016/03. У DDPG наилучший валидный коэффициент Sharpe, равный 0,61 за период с 2016/01 по 2016/03, поэтому мы используем DDPG для торговли в следующем квартале с 2016/04 по 2016/06. A2C имеет лучший коэффициент достоверности -0,15 в период с 2020/01 по 2020/03, поэтому мы используем A2C для торговли в следующем квартале с 2020/04 по 2020/05. Для оценки наших результатов используются пять метрик:

- Кумулятивная доходность: рассчитывается путем вычитания конечной стоимости портфеля из его начальной стоимости, а затем деления на начальную стоимость.
- Годовая доходность: средняя геометрическая сумма денег, зарабатываемых агентом каждый год в течение периода времени.
- Годовая волатильность: годовое стандартное отклонение доходности портфеля.

- Коэффициент Шарпа: рассчитывается путем вычитания годовой безрисковой ставки из годовой доходности и деления на годовую волатильность.
- Максимальная просадка: максимальный процент потерь в течение торгового периода.

ТАБЛИЦА I
КОЭФФИЦИЕНТЫ ШАРПА С ТЕЧЕНИЕМ ВРЕМЕНИ.

Торговый квартал	PPO	A2C	DDPG	Выбранная модель
2016/01-2016/03	**0.06**	0.03	0.05	PPO
2016/04-2016/06	0.31	0.53	**0.61**	DDPG
2016/07-2016/09	-0.02	0.01	**0.05**	DDPG
2016/10-2016/12	**0.11**	0.01	0.09	PPO
2017/01-2017/03	**0.53**	0.44	0.13	PPO
2017/04-2017/06	0.29	**0.44**	0.12	A2C
2017/07-2017/09	**0.4**	0.32	0.15	PPO
2017/10-2017/12	-0.05	-0.04	**0.12**	DDPG
2018/01-2018/03	**0.71**	0.63	0.62	PPO
2018/04-2018/06	-0.08	-0.02	**-0.01**	DDPG
2018/07-2018/09	-0.17	**0.21**	-0.03	A2C
2018/10-2018/12	0.30	**0.48**	0.39	A2C
2019/01-2019/03	-0.26	-0.25	**-0.18**	DDPG
2019/04-2019/06	**0.38**	0.29	0.25	PPO
2019/07-2019/09	**0.53**	0.47	0.52	PPO
2019/10-2019/12	-0.22	**0.11**	-0.22	A2C
2020/01-2020/03	-0.36	**-0.13**	-0.22	A2C
2020/04-2020/05	-0.42	**-0.15**	-0.58	A2C

Кумулятивная доходность отражает доходность в конце торгового этапа. Годовая доходность - это доходность портфеля в конце каждого года. Годовая волатильность и максимальная просадка измеряют устойчивость модели. Коэффициент Шарпа - это широко используемая метрика, которая объединяет доходность и риск.

2) Анализ эффективности агента: Из таблицы 2 и рисунка 5 видно, что агент A2C более адаптивен к риску. У него самая низкая годовая волатильность 10,4% и максимальная просадка -10,2% среди всех трех агентов. Таким образом, A2C хорошо справляется с медвежьим рынком. Агент PPO хорошо следует за трендом и генерирует большую доходность, у него самая высокая годовая доходность 15,0% и кумулятивная доходность 83,0% среди всех трех агентов. Поэтому PPO предпочтительнее использовать на бычьем рынке. DDPG демонстрирует схожие, но не такие высокие результаты, как PPO, и может использоваться в качестве дополнительной стратегии к PPO на бычьем рынке. Показатели всех трех агентов превосходят два эталона, промышленный индекс Доу-Джонса и минимальное распределение портфеля DJIA, соответственно.

3) Эффективность при обвале рынка: На рисунке 6 видно, что наша ансамблевая стратегия и три агента показывают хорошие результаты при обвале фондового рынка в 2020 году. Когда индекс турбулентности достигает порогового значения, это указывает на экстремальную ситуацию на рынке. Тогда наши агенты продают все имеющиеся у них акции и ждут, когда рынок вернется в нормальное состояние, чтобы возобновить торговлю. Благодаря учету индекса турбулентности агенты могут сократить потери и успешно пережить обвал фондового рынка в марте 2020 года. Мы можем снизить пороговое значение индекса турбулентности для более высокого неприятия риска.

4) Сравнение с эталонами: На рисунке 5 показано, что наша ансамблевая стратегия значительно превосходит DJIA и распределение портфеля с минимальной дисперсией [9]. Как можно заметить

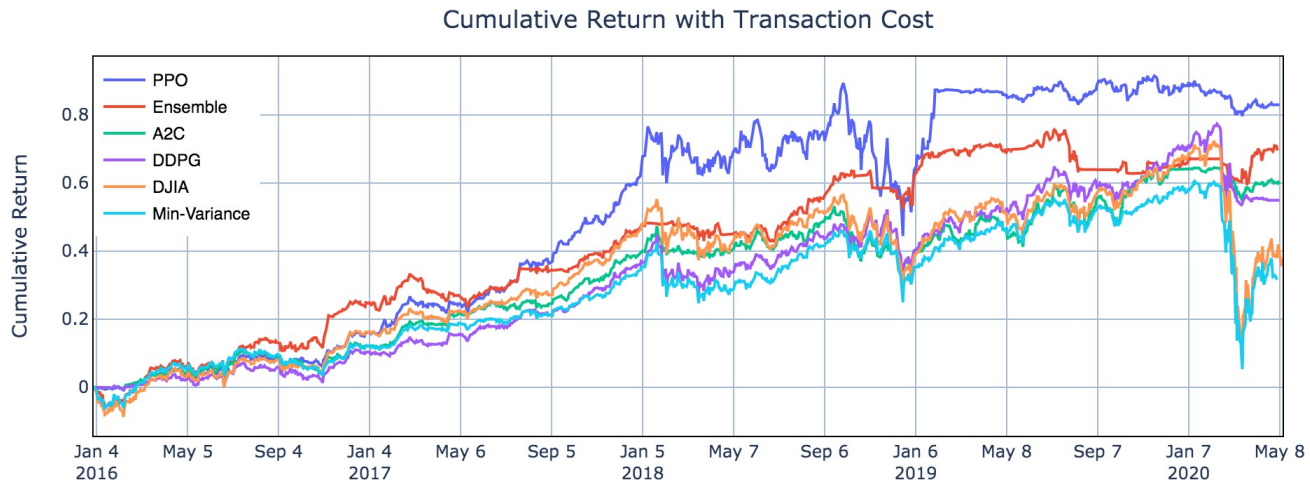

Рис. 5. Кривые кумулятивной доходности нашей ансамблевой стратегии и трех алгоритмов, основанных на акторной критике, стратегии распределения портфеля с минимальной дисперсией и промышленного индекса Доу-Джонса. (Начальная стоимость портфеля $1, 000, 000, с 2016/01/04 по 2020/05/08).

ТАБЛИЦА II
СРАВНЕНИЕ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ.

(2016/01/04-2020/05/08)	Ансамбль (Наш)	PPO	A2C	DDPG	Min-Variance	DJIA
Кумулятивная доходность	70.4%	83.0%	60.0%	54.8%	31.7%	38.6%
Годовая декларация	13.0%	15.0%	11.4%	10.5%	6.5%	7.8%
Годовая волатильность	9.7%	13.6%	10.4%	12.3%	17.8%	20.1%
Коэффициент Шарпа	1.30	1.10	1.12	0.87	0.45	0.47
Максимальная просадка	-9.7%	-23.7%	-10.2%	-14.8%	-34.3%	-37.1%

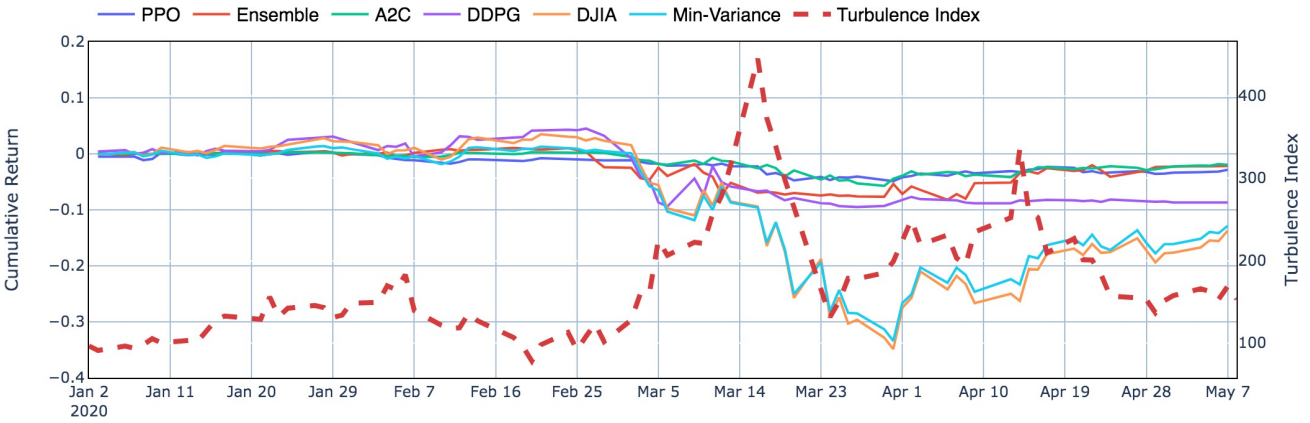

Рис. 6. Показатели во время обвала фондового рынка в первом квартале 2020 года.

Как видно из таблицы 2, ансамблевая стратегия достигает коэффициента Sharpe 1,30, что намного выше, чем коэффициент Sharpe 0,47 для DJIA и 0,45 для портфеля с минимальной вариацией. Годовая доходность ансамблевой стратегии также намного выше, а годовая волатильность намного ниже, что говорит о том, что ансамблевая стратегия превосходит DJIA и портфельное распределение с минимальной волатильностью в балансе между риском и доходностью. Ансамблевая стратегия также превосходит A2C с коэффициентом Sharpe 1,12, PPO с коэффициентом Sharpe 1,10 и DDPG с коэффициентом Sharpe 0,87, соответственно. Таким образом, наши результаты показывают, что предложенная ансамблевая стратегия может эффективно разработать торговую стратегию, которая превосходит три индивидуальных алгоритма и два базовых.

VII. Заключение

В этой статье мы изучили возможности использования алгоритмов на основе критики агентов, таких как Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C) и Deep Deterministic Policy Gradient (DDPG), для обучения стратегии торговли акциями. Чтобы приспособиться к различным рыночным ситуациям, мы используем ансамблевую стратегию для автоматического выбора наиболее эффективного агента для торговли на основе коэффициента Шарпа. Результаты показывают, что наша ансамблевая стратегия превосходит три индивидуальных алгоритма, промышленный индекс Доу-Джонса и метод распределения портфеля с минимальной дисперсией по коэффициенту Шарпа за счет баланса между риском и доходностью в условиях транзакционных издержек.

В будущем будет интересно исследовать более сложную модель [44], решить эмпирические задачи [45], работать с большими данными [46], такими как акции, входящие в S&P 500. Мы также можем исследовать дополнительные возможности пространства состояний, например добавить расширенную модель транзакционных издержек и ликвидности [47], включить в наши наблюдения показатели фундаментального анализа [9], анализ новостей финансового рынка с помощью естественного языка [48] и ESG-функции [12]. Мы заинтересованы в прямом использовании коэффициента Шарпа в качестве функции вознаграждения, но при этом агенты должны наблюдать гораздо больше исторических данных, и пространство состояний будет расти экспоненциально.

Ссылки

[1] Стелиос Д. Бекирос, "Нечеткое адаптивное принятие решений для ограниченно рациональных трейдеров на спекулятивных фондовых рынках", *Европейский журнал операционных исследований*, том 202, № 1, стр. 285-293, апрель 2010.

[2] Юн Чжан и Синьюй Ян, "Онлайн-стратегия выбора портфеля на основе комбинирования советов экспертов", *Вычислительная экономика*, том 50, 05 2016.

[3] Youngmin Kim, Wonbin Ahn, Kyong Joo Oh, and David Enke, "An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms," *Applied Soft Computing*, vol. 55, pp. 127-140, 02 2017.

[4] Гарри Марковиц, "Выбор портфеля", *Финансовый журнал*, том 7, № 1, стр. 77-91, 1952 г.

[5] Димитри Берцекас, *Динамическое программирование и оптимальное управление*, т. 1, 01 1995.

[6] Франческо Бертолуццо и Марко Корацца, "Тестирование различных конфигураций обучения с применением реинформации для финансовой торговли: введение и приложения", *Procedia Economics and Finance*, vol. 3, pp. 68-77, 12 2012.

[7] Ральф Нойнер, "Оптимальное распределение активов с помощью адаптивного динамического программирования", *Конференция по нейронным системам обработки информации, 1995*, 05 1996.

[8] Ральф Нойнейер, "Улучшение q-обучения для оптимального распределения активов", 01 1997.

[9] Hongyang Yang, Xiao-Yang Liu, and Qingwei Wu, "A practical machine learning approach for dynamic stock recommendation," in *IEEE TrustCom/BiDataSE, 2018.*, 08 2018, pp. 1693-1697.

[10] Юньчжэ Фан, Сяо-Ян Лю и Хонгян Ян, "Практический подход машинного обучения для захвата альфа-технологий, основанных на научных данных, в авиационной промышленности", в *2019 году IEEE Международная конференция по большим данным (Big Data) Специальная сессия по интеллектуальному поиску данных*, 12 2019, стр. 2230-2239.

[11] Вэньбин Чжан и Стивен Скиена, "Торговые стратегии для использования настроений в блогах и новостях", *Четвертая международная конференция AAAI по веб-блогам и социальным медиа, 2010*, 01 2010.

[12] Цянь Чэнь и Сяо-Ян Лю, "Количественная оценка альфа-активности esg с использованием больших данных ученых: автоматизированный подход машинного обучения", *ACM International Conference on AI in Finance, ICAIF 2020*, 2020.

[13] Виджай Конда и Джон Цициклис, "Акторно-критические алгоритмы", *Общество промышленной и прикладной математики*, том 42, 04 2001.

[14] Джон Шульман, Филип Вольски, Прафулла Дхаривал, Алек Рэдфорд и Олег Климов, "Алгоритмы оптимизации проксимальной политики", *arXiv:1707.06347*, 07 2017.

[15] Zhipeng Liang, Kangkang Jiang, Hao Chen, Junhao Zhu, and Yanran Li, "Adversarial deep reinforcement learning in portfolio manage- ment", *arXiv: Portfolio Management*, 2018.

[16] Владимир Мних, Адриа Бадиа, Мехди Мирза, Алекс Грейвс, Тимо Ти Лилликрап, Тим Харли, Дэвид Сильвер и Корай Кавуккуоглу, "Асинхронные методы глубокого обучения с подкреплением", *33-я Международная конференция по машинному обучению*, 02 2016.

[17] Zihao Zhang, "Deep reinforcement learning for trading", *ArXiv 2019*, 11 2019.

[18] Тимоти Лилликрап, Джонатан Хант, Александр Притцель, Николас Хесс, Том Эрез, Юваль Тасса, Дэвид Сильвер и Даан Виерстра, "Непрерывное управление с глубоким обучением с подкреплением", *Международная конференция по изучению представлений (ICLR) 2016*, 09 2015.

[19] Чжуорань Сюн, Сяо-Ян Лю, Шань Чжун, Хонгян Ян, и А. Эльвалид, "Практический подход глубокого обучения с подкреплением для биржевой торговли", *NeurIPS Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, 2018*, 2018.

[20] Томас Г. Фишер, "Reinforcement learning in financial markets - a survey", FAU Discussion Papers in Economics 12/2018, Friedrich- Alexander University Erlangen-Nuremberg, Institute for Economics, 2018.

[21] Линь Чэнь и Цян Гао, "Применение глубокого подкрепляющего обучения для автоматизированной торговли акциями", *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, 2019, pp. 29-33.

[22] Куанг-Винь Данг, "Обучение с подкреплением в биржевой торговле", в журнале *Advanced Computational Methods for Knowledge Engineering. ICCSAMA 2019. Advances in Intelligent Systems and Computing, vol 1121. Springer, Cham*, 01 2020.

[23] Гыын Чон и Ха Ким, "Улучшение финансовых торговых решений с помощью глубокого q-обучения: прогнозирование количества акций, стратегии действий и трансфертное обучение", *Expert Systems with Applications*, vol. 117, 09 2018.

[24] Джон Муди и Мэтью Саффелл, "Обучение торговле с помощью прямого подкрепления", *IEEE Transactions on Neural Networks*, vol. 12, pp. 875-89, 07 2001.

[25] Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai, "Deep direct reinforcement learning for financial signal repre- sentation and trading," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 1-12, 02 2016.

[26] Чжэнъяо Цзян и Цзиньцзюнь Лян, "Управление портфелем криптовалют с помощью глубокого обучения с подкреплением", *конференция 2017 Intelligent Systems Conference*, 09 2017.

[27] Стелиос Бекирос, "Гетерогенные торговые стратегии с адаптивным нечетким акторно-критическим обучением с подкреплением: Поведенческий подход", *Журнал экономической динамики и управления*, том 34, стр. 1153-1170, 06 2010.

[28] Jinke Li, Ruonan Rao, and Jun Shi, "Learning to trade with deep actor critic methods", *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 02, pp. 66-71, 2018.

[29] Юксин Ву и Юаньдун Тянь, "Обучающий агент для игры в шутер от первого лица с акторно-критическим обучением", *Международная конференция по обучающим представлениям (ICLR), 2017*, 2017.

[30] А. Илманен, "Ожидаемая доходность: Руководство инвестора по извлечению рыночных выгод", 05 2012.

[31] Марк Крицман и Юаньчжэнь Ли, "Черепа, финансовая турбулентность и управление рисками", *Журнал финансовых аналитиков*, том 66, 10 2010.

[32] Грег Брокман, Вики Чунг, Людвиг Петтерссон, Йонас Шнайдер, Джон Шульман, Цзе Танг и Войцех Заремба, "Openai gym", 2016.

[33] Прафулла Дхаривал, Кристофер Хессе, Олег Климов, Алекс Никол, Маттиас Плапперт, Алек Рэдфорд, Джон Шульман, Шимон Сидор, Юхуай Ву и Петр Жохов, "Базовые линии Openai", https://github.com/openai/baselines, 2017.

[34] Эшли Хилл, Антонин Раффин, Максимилиан Эрнестус, Адам Глив, Ансси Канервисто, Рене Траоре, Прафулла Дхаривал, Кристофер Хессе, Олег Климов, Алекс Никол, Маттиас Плапперт, Алек Рэдфорд, Джон Шульман, Шимон Сидор и Юхуай Ву, "Стабильные базовые уровни", https://github.com/hill-a/stable-baselines, 2018.

[35] Теренс Чонг, Винг-Кам Нг и Венус Лью, "Пересмотр эффективности осцилляторов macd и rsi", *Журнал о рисках и финансовом менеджменте*, том 7, стр. 1-12, 03 2014.

[36] Мансур Майтах, Петр Прочажка, Михал C˘ erma´k и Карел S˘re´dl, "Индекс товарного канала: оценка правил торговли сельскохозяйственными товарами", *Международный журнал экономики и финансовых проблем*, том 6, стр. 176-178, 03 2016.

[37] Ихлаас Гурриб, "Эффективность среднего индекса направленности как инструмента рыночного тайминга для наиболее активно торгуемых валютных пар на базе usd", *Банки и банковские системы*, том 13, стр. 58-70, 08 2018.

[38] Ричард Саттон и Эндрю Барто, "Reinforcement learning: an introduction", *IEEE Transactions on Neural Networks*, vol. 9, pp. 1054, 02 1998.

[39] Ричард Саттон, Дэвид Макалстер, Сатиндер Сингх и Йишай Ман- кис, "Градиентные методы политики для обучения с подкреплением и аппроксимацией ф у н к ц и й ", *Конференция по нейронным системам обработки информации (NeurIPS), 1999*, 02 2000.

[40] Лучиан Бузоню, Тим де Бруин, Домагой Толич, Йенс Кобер и Ивана Палунко, "Обучение с подкреплением для управления: Производительность,

стабильность и глубокие аппроксиматоры", *Ежегодные обзоры по управлению*, 10 2018.

[41] Джон Шульман, Сергей Левин, Филипп Мориц, Майкл Джордан и Питер Аббел, "Оптимизация политики доверительных областей", *31-я Международная конференция по машинному обучению*, 02 2015.

[42] У.Ф. Шарп, "Коэффициент Шарпа", *Журнал "Управление портфелем"*, 01 1994.

[43] Wharton Research Data Service, "Standard & poor's compustat", 2015, данные получены из Wharton Research Data Service,.

[44] Лу Ванг, Вэй Чжан, Сяофэн Хэ и Хонюань Чжа, "Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation," in *Conference on Knowledge Discovery and Data Mining (KDD), 2018*, 07 2018, pp. 2447-2456.

[45] Габриэль Дюлак-Арнольд, Н. Левин, Дэниел Дж. Манковиц, Дж. Ли, Космин Падурару, Свен Говаль и Т. Хестер, "Эмпирическое исследование проблем обучения с подкреплением в реальном мире", *ArXiv*, vol. abs/2003.11881, 2020.

[46] Юрий Бурда, Харрисон Эдвардс, Дипак Патхак, Амос Сторки, Тревор Даррелл и Алексей Эфрос, "Крупномасштабное исследование обучения на основе любопытства", *2019 Seventh International Conference on Learning Representations (ICLR) Poster*, 08 2018.

[47] Вэньхан Бао и Сяо-Ян Лю, "Мультиагентное глубокое обучение с подкреплением для анализа стратегии ликвидации", *ICML Workshop on Applications and Infrastructure for Multi-Agent Learning, 2019*, 06 2019.

[48] Xinyi Li, Yinchuan Li, Hongyang Yang, Liuqing Yang, and Xiao-Yang Liu, "Dp-lstm: Differential privacy-inspired lstm for stock prediction using financial news", *33rd Conference on Neural Information Processing Systems (NeurIPS 2019) Workshop on Robust AI in Financial Services: Data, Fairness, Explainability, Trustworthiness, and Privacy, December 2019*, 12 2019.