

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук
Основная образовательная программа
Экономика и Анализ Данных

ОБЗОР ЛИТЕРАТУРЫ
НА ТЕМУ
"АЛГОРИТМИЧЕСКИЙ ТРЕЙДИНГ"

Выполнил студент группы 223, 2 курса,
Шевченко Артём Эдуардович

Преподаватель НИС:
Степанченко Дмитрий Сергеевич

Москва 2023

Содержание

1	Введение	3
2	Терминология	4
2.1	Финансовые термины	4
2.2	Термины машинного обучения	6
3	Обзор Литературы	10
3.1	Application of deep reinforcement learning on automated stock trading	10
3.2	Financial trading as a game: A deep reinforcement learning approach	12
3.3	Adaptive stock trading strategies with deep reinforcement learning methods	13
3.4	Deep reinforcement learning for trading	14
3.5	Deep direct reinforcement learning for financial signal representation and trading	15
3.6	Quantitative trading on stock market based on deep reinforcement learning	17
3.7	Deep reinforcement learning for automated stock trading: An ensemble strategy	18
3.8	Practical deep reinforcement learning approach for stock trading . .	20
3.9	Stock trading bot using deep reinforcement learning	21
3.10	Adaptive stock trading strategies with deep reinforcement learning methods	22
3.11	Deep reinforcement learning for trading	23
3.12	Fairness in Financial Markets: The Case of High Frequency Trading	23
4	Заключение	26

1 Введение

В последние десятилетия алгоритмический трейдинг стал неотъемлемой частью финансовых рынков, привлекая внимание исследователей и практиков. Эта динамично развивающаяся область финансовых технологий обозначает собой применение компьютерных алгоритмов для принятия решений по торговле финансовыми инструментами. В рамках данного обзора литературы мы сосредоточимся исключительно на методах машинного обучения, поскольку их роль в алгоритмическом трейдинге становится [\[1\] \(все более существенной\)](#). Исследования в области машинного обучения в контексте алгоритмического трейдинга предоставляют богатый арсенал инструментов для анализа рынка, прогнозирования цен, и оптимизации стратегий торговли. В данном обзоре мы рассмотрим различные подходы машинного обучения, проведем анализ результатов, полученных в ходе применения этих методов на финансовых рынках. Особое внимание уделено глубокому обучению с подкреплением – методу, который позволяет алгоритмам обучаться на основе взаимодействия с окружающей средой и полученных откликов. Этот подход демонстрирует выдающиеся результаты в области алгоритмического трейдинга, благодаря способности алгоритмов принимать сложные решения в условиях неопределенности на финансовых рынках. Помимо технических аспектов, наш обзор также включит анализ литературы, посвященной моральным и этическим вопросам, связанным с алгоритмической торговлей. Рассмотрение этих аспектов становится все более актуальным в контексте роста автоматизации торговых процессов и влияния алгоритмического трейдинга на стабильность и эффективность финансовых рынков.

2 Терминология

2.1 Финансовые термины

- 1 Трейдинг — это процесс купли и продажи финансовых инструментов, таких как акции, валюта, фьючерсы и другие, с целью получения прибыли.
- 2 Акции представляют собой доли в собственности компании, которые продаются на фондовой бирже.
- 3 Фьючерсы — это контракты, обязывающие купить или продать актив (например, сырье, валюту) по определенной цене в будущем.
- 4 Игрок в финансовом контексте — это участник рынка, совершающий операции купли-продажи для получения прибыли.
- 5 Биржа — это организованный рынок, где происходит купля-продажа финансовых инструментов.
- 6 Рынок — место встречи покупателей и продавцов, где осуществляется торговля активами.
- 7 Волатильность измеряет степень изменчивости цен на рынке за определенный период времени. Высокая волатильность указывает на большие колебания цен.
- 8 Отрицательная волатильность означает низкую изменчивость цен на рынке.
- 9 OHLCV (Open, High, Low, Close, Volume) — это формат представления ценовых данных, включающий открытие, максимум, минимум, закрытие и объем торгов.
- 10 SR (Support and Resistance) — уровни поддержки и сопротивления на графиках цен, где они обычно меняют свое направление.

- 11 Turtle - термин, связанный с торговой стратегией, разработанной Трейтилами (Turtles) — группой трейдеров под руководством Ричарда Денниса.
- 12 Индикаторы — это математические выражения, используемые для анализа рынка и предсказания его будущего направления.
- 13 ETF (Exchange-Traded Fund) — это фонд, торгующийся на бирже, который отслеживает индекс, сектор или другой актив.
- 14 S&P 500 — это индекс, представляющий 500 крупнейших компаний, котирующихся на фондовых биржах США.
- 15 Forex (Foreign Exchange) — рынок обмена валют, где происходит торговля национальными валютами.
- 16 MACD (Moving Average Convergence Divergence) — это технический индикатор, используемый для определения направления и силы тренда.
- 17 MA (Moving Average) — это скользящее среднее, статистический индикатор, используемый для сглаживания ценовых данных и выявления трендов.
- 18 EMA (Exponential Moving Average) — это экспоненциальное скользящее среднее, которое придает больший вес более свежим данным.
- 19 OBV (On-Balance Volume) — это индикатор, измеряющий объем торговли в зависимости от направления цены.
- 20 RSI (Relative Strength Index) — индикатор, измеряющий скорость и изменчивость цен для оценки перекупленности или перепроданности актива.
- 21 CCI (Commodity Channel Index) — это индикатор, измеряющий степень изменчивости цен относительно их среднего значения.

- 22 ADX (Average Directional Index) — индикатор, показывающий силу тренда.
- 23 PPO (Percentage Price Oscillator) — это процентный осциллятор цены, используемый для анализа трендов.
- 24 Backtest (далее - бэктест) - применение торговой стратегии к историческим данным для оценки ее прибыльности и возможных рисков
- 25 Коэффициент Шарпа - показатель, измеряющий какую доходность приносят инвестиции на каждую единицу риска. Вычисляется он как отношение средней премии за риск к среднему отклонению портфеля.
- 26 Коэффициент Сортино — показатель, позволяющий оценить доходность и риск инвестиционного инструмента, портфеля или стратегии. Коэффициент Сортино рассчитывается аналогично коэффициенту Шарпа, однако вместо волатильности портфеля используется так называемая «волатильность вниз».

2.2 Термины машинного обучения

- 1 Машинное обучение — это область искусственного интеллекта, где компьютерные системы обучаются автоматически улучшать свою производительность на основе опыта.
- 2 Нейронная сеть - это компьютерная модель, созданная вдохновленной биологической нейросетью, предназначенная для обработки информации подобно тому, как это делает человеческий мозг. Нейронные сети используются в машинном обучении для автоматического извлечения закономерностей из данных, обучения на опыте и принятия решений в различных задачах. Основные компоненты нейронной сети включают слои нейронов, веса, функции активации и алгоритм обучения. Нейроны объединяются в слои, входные данные передаются через сеть, и

в процессе обучения алгоритмы оптимизации корректируют веса для достижения желаемого результата.

- 3 DDR (Deep Direct Reinforcement) — это подход в машинном обучении, использующий алгоритмы обратного распространения ошибки.
- 4 A2C (Advantage Actor-Critic) — это алгоритм глубокого обучения, используемый в области машинного обучения.
- 5 DDPG (Deep Deterministic Policy Gradients)— это алгоритм обучения с подкреплением с непрерывными действиями.
- 6 RCNN (Region-based Convolutional Neural Network) — это тип нейронной сети, используемый для обнаружения объектов в изображениях.
- 7 Q-обучение — это метод обучения с подкреплением, который направлен на обучение агента (например, робота, компьютерной программы и т. д.) принимать решения в окружающей среде для максимизации общей награды. Основной идеей является обучение функции Q (Q-функции), которая оценивает ожидаемую общую награду для выполнения определенного действия в конкретном состоянии.
- 8 Reward-функция - функция, максимизируемая Q-обученными алгоритмами
- 9 GDQN (Generalized Deep Q Network) — это обобщенная версия алгоритма глубокого обучения Q-сетей.
- 10 Политика - в нейронных сетях набор правил, стратегий или функций, определяющих поведение алгоритма обучения или функционирование нейронной сети.
 - Политика обучения (Learning Policy) - это стратегия или набор правил, определяющих, как алгоритм машинного обучения принимает решения на основе предоставленных данных. Например,

в усилении (reinforcement) обучении, политика определяет, какие действия должны предприниматься в конкретных состояниях для достижения определенных целей.

- Политика обновления весов (Weight Update Policy) - в контексте нейронных сетей, особенно при обучении, политика обновления весов определяет, каким образом веса связей между нейронами изменяются в ответ на ошибку в предсказаниях модели.
- Политика активации (Activation Policy) - это относится к правилам или функциям, определяющим активацию нейронов в слоях нейронной сети. Примерами могут быть сигмоидные, гиперболические тангенс, или ReLU (Rectified Linear Unit) функции активации.
- Политика регуляризации (Regularization Policy) - некоторых случаях, политика может определять, какие методы регуляризации (например, L1, L2 регуляризация) применяются для предотвращения переобучения модели.

11 GDPG (Generative Adversarial Networks for Policy Generation) — это алгоритм машинного обучения, использующий генеративные состязательные сети для генерации политики.

12 PG (Policy Gradient) — это метод обучения с подкреплением, основанный на прямой оптимизации политики.

13 БТТ (Backtesting Trading Strategies) — это процесс проверки эффективности торговых стратегий на исторических данных.

14 БТТ (Backpropagation Through Time) - метод обучения рекуррентных нейронных сетей (RNN). RNN - это тип нейронных сетей, способных обрабатывать последовательности данных и сохранять внутреннее состояние или "память" о предыдущих входах.

15 LSTM (Long Short-Term Memory) — это тип рекуррентной нейронной

сети, эффективно работающей с последовательными данными.

- 16 Gated Recurrent Units (GRU) — это тип рекуррентных нейронных сетей, подобных LSTM.
- 17 Deep Q Network (DQN)— это тип нейронной сети, используемой в обучении с подкреплением для оценки долгосрочных стратегий действий.
- 18 Deep Recurrent Q-Network — это комбинация рекуррентной нейронной сети и Q-сети в области обучения с подкреплением.
- 19 Epsilon Greedy Exploration — это стратегия в обучении с подкреплением, при которой с определенной вероятностью выбирается случайное действие для исследования.
- 20 ANNs (Artificial Neural Networks) — искусственные нейронные сети, моделирующие работу человеческого мозга для обработки информации.

3 Обзор Литературы

3.1 Application of deep reinforcement learning on automated stock trading

В рассмотренной статье (2) предпринята попытка реализовать торгового бота, рассматривающего биржу как игру, в которой цель - максимизация reward-функции, то есть, по сути, прибыли. Эта идея изначально была использована в научной среде в некоторых статьях для прохождения игр известной компании Atari, а также игры в покер моделями машинного обучения - глубокого обучения с подкреплением. Авторы статьи используют вариант системы обучения DQN - Deep Recurrent Q-Network (DRQN).

Для обучения и бэктеста модели авторы используют набор исторических данных S&P 500 ETF. Ими тестируются стратегия покупки и удержания, которая используется как опытными инвесторами, торгующими в лонг, так и среднестатистическими игроками. Датасет получен через Yahoo Finance и содержит цены закрытия по бумагам за каждый день на протяжении 19 лет. Данные за первые 5 лет используются для обучения, а остальные для бэктеста.

Авторы положили, что в любое время модель должна уметь оптимально делать выбор меж тремя опциями: покупкой бумаги, её продажей или пропуском итерации и удержанием и текущих позиций. Reward рассчитывается как разница между скорректированной ценой закрытия на следующий день и ценой закрытия в текущий день в зависимости от того, какое действие было совершено.

Авторы достигли успеха в обучении модели - она превосходит все базовые модели, торгующие лишь по различным индикаторам. Годовая прибыль от работы этой модели составляет около 22 — 23%. Это не плохо, но далеко от некоторых других рассмотренных в этом обзоре моделей, которые смогли достичь более 60% годовой прибыли.

Говоря о недостатках работы, существуют явные ограничения в отношении результатов данного исследования. Во-первых, для обучения и бэк-тестирования стратегии используется лишь одна бумага. Это (вообще говоря) неверно и модель может иметь совершенно другую точность для других активов, что ставит под сомнение общий характер подхода к более крупному портфелю. Во-вторых, целевая функция не учитывает и даже не пытается устранить риск, как это нередко делается в других работах по разработке алгоритмов и моделей для алгоритмической торговли. Ровно по этой причине, позволить работать этой модели в реальном времени на реальных данных без человеческого контроля очень рискованно, поскольку поведение модели при необходимости принятия высокорискованных решений относительно непредсказуемо. Одним из быстрых решений было бы использование коэффициента Шарпа в качестве меры поощрения. В-третьих, пространство действий для модели крайне узко и базово - очевидно, покупка и продажа может производить в разных соотношениях и в разном объеме в разных рыночных ситуациях. Вдобавок, представление признаков далеко не оптимальное, необработанные данные о ценах содержат много шума, что решается применением различных аналитических и технических методов обработки биржевых датасетов. Небольшая проблема есть и в выбранном временном промежутке для обучения и тестирования модели (2006-2018), что включает в себя несколько крупных кризисов (например, кризис 2008 года), что могло негативно отразиться на решениях модели, поскольку любой кризис включает в себя факторы, которые невозможно отразить лишь за счёт различных котировок и индикаторов (ведь в рассмотрение, должны включаться риски, о которых я упоминал ранее, а также новостной фон).

Тем не менее, данная работа - отличный эксперимент в этой области, который, хоть и не с лучшей возможной точностью и с учетом множества факторов, уже доказывают практическую применимость глубинного обучения с подкреплением для алгоритмической торговли.

3.2 Financial trading as a game: A deep reinforcement learning approach

С описанной выше работой тесно связано [настоящее исследование \(3\)](#), авторы которого также используют модель глубинного обучения с подкреплением. Тем не менее, ключевое отличие в подходах в обучении - автор использует в общей сложности 16 признаков, включая необработанные данные о ценах (открытие, максимум, минимум, закрытие, объем торгов), а также различные индикаторы и технические показатели.

По-прежнему присутствуют значимые недостатки в настройке модели: функция reward (определяемая здесь как отчет потрфельного лога) лучше, но всё ещё не учитывает риски, а пространство действий также состоит лишь из покупки, продажи и удержания бумаги.

Алгоритм применяется на бирже Forex (FX) с 2012 по 2017 год для 12 валютных пар с 15-минутным таймфреймом. В итоге, алгоритм смог достичь получения 60% годовой прибыли на некоторые бумаги, в среднем около 10% на каждую. Авторы не тестировали алгоритм на фондовом рынке, однако, учитывая то, что рассмотренный алгоритм представляет собой улучшенную версию обсуждавшегося ранее, можно смело предположить, что он с хорошей вероятностью может достаточно успешно работать и на фондовом рынке.

В контексте практической применимости алгоритма полезно попробовать сравнить результаты его работы с действиями профессионального трейдера. К сожалению, не существует единого стандарта/показателя, отражающего средний заработок профессионального трейдера, поскольку это зависит от множества индивидуальных факторов. Тем не менее мы можем сравнить доходность от работы модели с доходностью при лонг-инвестировании в бумаги различных стабильных индексов, например S&P500, что является очень распространенной стратегией для многих игроков. Так, годовой процент прибыли от S&P500 составит порядка 7%. Эту цифру я буду использовать далее для сравнения полученных авторами различных статей результатами с рядом ба-

зовых стратегий, включая и торговлю по S&P. Итак, учитывая стабильность алгоритма, последовательность его действий, отсутствие человеческого фактора, а также непрерывное реинвестирование в выбранные активы, модель вполне можно сравнить с трейдером-профессионалом.

3.3 Adaptive stock trading strategies with deep reinforcement learning methods

Существует несколько исследований, в которых методы глубокого подкрепления используются для сравнения с другими методами машинного обучения для алгоритмической торговли. В [настоящей статье \(4\)](#), Gated Recurrent Units (GRUs) определяются и используются как метод глубинного обучения для автоматического извлечения признаков с фондового рынка для работы алгоритма Deep Reinforcement Q-Learning (GDQN). Стимулом для внедрения такого подхода стала попытка выявить новые закономерности в динамике рынка. Авторами вводится система, включающая функции update gate и reset gate для обучения нейронной сети. На каждой итерации модели reset gate действуют как фильтр для датасета полученного на предыдущей итерации. Для это он сохраняет часть даты и update gate решает, будет ли hidden state обновлено текущей информацией или нет.

Вклад рассматриваемой статьи заключается в использовании новой функции вознаграждения, которая учитывает риски и определенным образом обрабатывает их, чего не делалось в прежде рассмотренных работах. Делается это с помощью [коэффициента Сортино \(СР\) \(5\)](#). Авторы утверждают, что эта reward-функция будет более справедливой на рынке высокой волатильности, потому что она умеет измерять отрицательную волатильность.

Датасет состоит из классического OHLCV, но также включает и некоторые популярные технические индикаторы, такие как MACD, MA, EMA, OBV. Набор действий не отличается от предыдущих работ - вновь модель может работать лишь с 1 бумагой за 1 транзакцию, которую она может купить,

продать, или удерживать в портфеле.

Используемый таймлайн является лучшим среди доселе рассмотренных мной - данные взяты с начала 2008 года до конца 2018 года с таймфреймом в день. Данные за первые восемь лет используются для обучения, и они включают 15 американских, британских и китайских бумаг. Модель тестировалась на каждой из этих бумаг отдельно.

Лучший результат работы модели был зарегистрирован для бумаги на китайском рынке: 171% доходности и 1,79 SR. Также был зарегистрирован SR 3,49 и доходность 96,6% для другой китайской акции. Были и акции с отрицательной доходностью - их оказалось 2. Авторы сравнивают свою модель с прочими стратегиями и заявляют, что она работает лучше, чем базовая [Turtle Strategy \(6\)](#) и более стабильно, чем [state-of-the-art direct reinforcement learning trading strategy \(7\)](#).

Таким образом, в работе показано, что GDPG более стабилен, чем GDQN, и обеспечивает несколько лучшие результаты. Недостатки этого исследования фактически те же, что и недостатки предыдущих - набор возможных действий моделей с бумагой всё такой же маленький. Хотя reward-функция для обучения и была значительно улучшена в сравнении с прежде рассмотренными работами, она всё ещё не идеальна в вопросе учёта рисков и дополнительных издержек (например, вопрос легальной математической минимизации налогового вычета с портфеля хоть и не благороден в формате научного вопроса, однако имеет прямое отношение к задаче, решаемой физическими и юридическими лицами при реальной торговле).

3.4 Deep reinforcement learning for trading

Совсем кратко упомяну статью [Deep Reinforcement Learning for Trading \(7\)](#), поскольку исследование сосредоточено на фьючерсах, а не на акциях. Обучаемая модель очень похожа на рассмотренную [в статье из второго раздела \(3\)](#). В работе осуществляется внедрение LSTM-based Q-network с исполь-

зованием технических индикаторов, таких как RSI и MACD. Особое внимание уделяется рассмотрению волатильности как важного фактора риска на финансовых рынках. Авторы предлагают уникальный подход к формированию reward-функции, в которой учитывается общая волатильность рынка. При низкой волатильности модель автоматически увеличивает объем торговых позиций, снижая их в периоды повышенной волатильности. Это стратегическое решение значительно снижает риски при торговле. Авторы демонстрируют, что использование алгоритмов DQN превосходит базовые модели, такие как покупка и удержание, а также стратегия на основе сигнала MACD (8). Основываясь на проведенных исследованиях, представленных в работе, подтверждается, что разработанный алгоритм обеспечивает устойчивую прибыль даже при высоких операционных издержках. Дополнительно, для полноты сравнения, в работе проводится анализ эффективности алгоритмов Policy Gradient (PG) и Advantage Actor-Critic (A2C) по сравнению с базовыми торговыми стратегиями, основанными на индикаторах, и с использованием DQN.

3.5 Deep direct reinforcement learning for financial signal representation and trading

В статье [Deep Direct Reinforcement Learning for Financial Signal Representation and Trading \(9\)](#) 2017 года предпринята конкретная попытка ответить на вопрос о том, как победить опытных трейдеров с помощью моделей углубленного обучения, основанного на алгоритме-трейдере. Это исследование является особенным, поскольку авторы утверждают, что (в то время) оно впервые использовало Deep Learning для финансовых операций в режиме реального времени. Они перестраивают структуру обычной модели Recurrent Neural Network для принятия повторяющихся решений для работы в онлайн-среде.

Итак, авторы используют Recurrent Neural Network для части RL, однако для извлечения признаков они используют Deep Learning в сочетании с [fuzzy](#)

[learning concepts \(10\)](#) для хорошего рандомизирования входных данных. Далее, авторы используют [fuzzy representations \(11\)](#) в рамках этапа обработки данных; а затем - углубленное обучение для анализа и обработки признаков в полученных данных. По прохождении этого процесса, данные подаются модели для непосредственно торговли.

Замечу, что предлагаемая авторами модель, несмотря на действительно впечатляющую систему обработки данных и генерации признаков, вышеупомянутую систему для обработки первичных данных и функции обучения, чтобы делать обоснованные предположения через систему RL для максимизации прибыли, совершенно не использует технический анализ, различные индексы и индикаторы в качестве признаков, тогда как они могли еще сильнее повысить эффективность и точность результата. Нерассмотрение этой возможности показалось мне немотивированным - даже если в данном случае работа с индикаторами и оказалась бесполезной, странно не включать сравнение обучения модели с ними и без в исследование.

Экспериментальная часть является очень приличной с точки зрения сравнения между устоявшимися стратегиями в литературе и предлагаемыми подходами. Тем не менее, к сожалению, портфель, рассмотренный в работе, является очень своеобразным - он включает два товара - сахар и серебро, и один фондовый индекс Future IF. Тем не менее, в разделе, посвященном будущим направлениям работы авторов, рассматривается вопрос об адаптации модели для управления более широким портфелем. Сравнения авторов как по биржевому индексу, так и по товарным фьючерсам показывают, что предлагаемая DDR и FDDR генерируют значительно большую общую прибыль в различных рыночных условиях, модели, использующие лишь пространство действий Buy + Hold, [SCOT \(12\)](#) и DDR без fuzzy representation для обработки данных. Коэффициент Шарпа составил около 9, что является очень хорошим результатом и говорит о малых рисках.

Кроме того, авторы также протестировали алгоритмы на биржевых индексах - таких как S&P500. Данные о запасах охватывали период с 1990 по

2015 год, когда первые восемь лет использовались для подготовки кадров, а операционные издержки были установлены на уровне 0,1% от величины индекса. Результатом явилось то, что модель действительно приносит прибыль - хотя и не такую высокую, учитывая продолжительность периода тестирования (около 18 лет). Я не привожу конкретные цифры для данной модели в виду того, что авторы предоставили лишь P&L Statement.

Одним из серьезных недостатков этого исследования является последовательность в ответе на вопрос сравнения работы модели с работой опытных трейдеров - ни в большей части работы, ни в выводах этот вопрос не обсуждается. Конечно, прибыль от работы модели в экспериментах и прибыль успешного профессиональный дневного трейдера несопоставимы (поскольку первое далеко даже от средней годовой заработной платы). Это связано с тем, что авторы не упоминают фактический стартовый капитал, использованный в экспериментах; поэтому мы можем полагаться только на результаты соотношения Шарпа для заключения. Тем не менее, предпосылки к положительному ответу на вопрос сопоставимости работы модели и работы профессионала есть, и не обсуждать это в статье, исследующей конкретно этот вопрос, предельно немотивированно.

Несмотря на недостатки, построенная модель дает положительный результат в долгосрочной перспективе, что само по себе является хорошим достижением. Ведь тогда, если мы сменим исследовательский вопрос на сравнение модели со средним трейдером, подход к глубокому подкреплению, вероятно, окажется успешнее.

3.6 Quantitative trading on stock market based on deep reinforcement learning

Основным вкладом [настоящей статьи \(13\)](#) является всесторонний анализ преимуществ выбора глубокой нейронной сети (LSTM) по сравнению с полнотью подключенной (fully connected one) и влияние некоторых комбинаций

технических показателей на производительность на ежедневных данных китайских рынков ценных бумаг. Это исследование доказывает, что выбор правильных технических показателей действительно имеет значение в вопросе максимизации прибыли от работы торгового алгоритма. Результаты чистой производительности являются смешанными, авторы показывают, что предложенный метод может приносить достойную прибыль в одних акциях, но он также может иметь колебательное (и даже непредсказуемое) поведение в других, поэтому я не буду рассматривать малоинформативные конкретные численные результаты, представленные в работе.

3.7 Deep reinforcement learning for automated stock trading: An ensemble strategy

В [данном исследовании \(14\)](#) была предложена модель (ensemble system), использующая DRL для дальнейшего улучшения результатов работы модели после основного обучения. Достигается это при помощи некоторых экспериментов по торговле акциями. Они представляют модели, что работают на основе 3 алгоритмов DRL: PPO, A2C, DDPG. Авторы утверждают, что стратегия на основе совокупности этих алгоритмов делает торговлю более надежной в различных рыночных ситуациях и может максимизировать прибыль, подверженную рисковым ограничениям. Эксперименты проводятся на основе 30 акций индекса Dow Jones. По словам авторов, особенностями, используемыми для обучения модели, являются скорректированная цена закрытия, а также некоторые технические показатели: MACD, RSI, CCI и ADX.

Выбор пространства действия также подвергся расширению, в сравнении с предыдущими работами: $\{k, \dots, 1, 0, 1, \dots, k\}$, где k и $-k$ представляет количество акций, которые мы можем купить и продать. Таким образом, в общей сложности у модели есть $2k + 1$ выбора для каждой отдельно взятой акции. Для корректной работы основных алгоритмов, авторы позже нормируют это пространство к $[-1, 1]$. Этот подход, очевидно, предоставляет гораздо боль-

шую свободу модели, увеличивает её вычислительную сложность, но благотворно влияет на результирующей прибыли.

Заметим также, что признаки для каждой из 30 рассматриваемых акций в работе объединены, поскольку авторы ставят цель в определении, какие акции купить или продать. Такой подход также отличается от ранее рассмотренных и скорее должен рассматриваться отдельно от основного торгового алгоритма.

Способ работы стратегии заключается в том, что время обучения и время проверки тестирования выбираются непрерывно, после чего выбирается наиболее эффективный временной промежуток. Он же затем используется в течение следующего фиксированного периода времени в качестве единственного агента в модели. Авторы объясняют это тем, что каждый торговый агент чувствителен к различным тенденциям - какие-то, ввиду особенностей обучения - к растущему рынку, а какие-то напротив, а какие-то в целом лучше приспособлены к волатильности. Говоря о последнем, авторы рассматривают и поведение модели при обвале рынка и высокой нестабильности. Они используют индекс финансовой турбулентности, который измеряет экстремальные колебания цен на активы, используя его вместе с фиксированным порогом, так что агент продает все и останавливается для совершения любых сделок, если на рынке происходит обвал. Это очень интересный подход, который несомненно снижает риски, однако стоит заметить, что опытные трейдеры на практике и используют такие обвалы для максимизации прибыли и проведения закупок в больших объемах.

В результате, алгоритм PPO обеспечил наибольшее значение совокупную функции reward (около 15% ежегодно). Однако коэффициент Шарпа, который по сути является функцией reward, нормированной по риску, оказался выше в случае совместной работы алгоритмов: 1,30 против 1,10. Это означает, что модель обеспечила авторам 13% годовой прибыли. Предлагаемая модель также превосходит промышленный индекс Доу-Джонса, который является основным предметом сравнения для выбранного датасета.

С большой вероятностью, единый подход ко всем бумагам оказался бы доходнее, чем индивидуальное рассмотрение каждой бумаги по отдельности, однако модель и так работает очень хорошо. Кроме того, она демонстрирует успех в проведении единовременного портфельного менеджмента и торговле по стратегии, а доходность оказывается намного выше, чем ожидаемая при дефолтной стратегии покупки/продажи (составляющая по S&P500 около 7%). Что немаловажно, алгоритм оказывается достаточно стабильным в экстремальных рыночных условиях (например, в ситуации обвала биржи в марте 2020 года).

3.8 Practical deep reinforcement learning approach for stock trading

В данной статье (15) рассматривается только алгоритм DDPG (из ранее рассмотренной статьи) в тех же условиях и с теми же параметрами оценки, но без каких-либо технических показателей и индикаторов в качестве входных данных. Есть и другие незначительные различия: данные о ценах за 2009-2018 годы для обучения агента и тестирования производительности; данные за период с начала 2009 года по конец 2014 года используются для обучения, а данные за период с 2015 года используются для тестирования.

Модель проверялась по торговым данным, с начала 2016 года до конца 2018 года. Результаты оказались очень хорошими приличными: 25,87% годовой доходности по индексу US 30 (Доу-Джонс), более 15,93% по методу Мин-Вариации (Min-Variance) и 16,40% по средне-индустриальному показателю доходности. Коэффициент Шарпа также оказался достаточно высоким (1.79), что демонстрирует рассмотренную модель как более надежную в вопросе баланса риска и прибыли.

3.9 Stock trading bot using deep reinforcement learning

[Stock Trading Bot Using Deep Reinforcement Learning \(16\)](#) является особенно интересной работой, поскольку она сочетает в себе подход DRL с [психологическим анализом \(17\)](#), используя внешний новостной фон. Хотя наиболее важной частью построенной авторами модели является RL-агент, дополнительный компонент, который пытается предсказать будущие движения на рынках посредством анализа настроений в новостях на финансовую тематику. Это очень актуальная и важная особенность модели DRL, поскольку новости являются важным сигналом для анализа критических ситуаций на бирже.

Предложенный в этой статье подход довольно прост: они используют DDPG для обучения и RCNN для классификации новостей. Изменение цены акции рассматривается как признак в модели наравне с оным, генерируемым путем анализа новостей, после чего модель предсказывает двоичное значение (0/1) - будет ли цена расти (или падать) в течение следующих нескольких дней.

Для эксперимента авторами были отобраны исторические данные о ежедневных котировках лишь для одной акции, однако статья нацелена больше на доказательство валидности концепции анализа новостного фона. RCNN была заранее подготовлена, для чего было использовано почти 96k новостных заголовков для обучения и около 31.5k для тестирования. В совокупности были использованы новости от 3.3k компаний. Количество параметров в модели составило около 26 миллионов.

В результате, авторы получили точность в 96,88% при тестировании модели. Несмотря на то, что эксперименты отнюдь не демонстрируют ошеломляющего увеличения прибыли, они показали, что модель с учётом новостного фона фактически повторяет базовую трейдерскую стратегию: она постоянно покупает и продает, но предпочитает удерживать бумагу в портфеле, когда происходит серьезное падение в цене.

Авторы также проверили несколько подходов к reward-функции: сначала они определили функцию как разницу между совокупной стоимостью активов модели и неизменной стоимостью активов (при бездействии), затем - как разницу между стоимостью, по которой бумаги торгуются в конечной временной точке, и стоимостью, по которой они были куплены, и затем - с использованием двойного вознаграждения (reward), которое говорило модели, являлось ли отдельно взятое действие прибыльным или нет. Первые две функции вознаграждения не сработали должным образом, тогда как последняя показала себя хорошо, несмотря на полное отсутствие учёта рисков.

3.10 Adaptive stock trading strategies with deep reinforcement learning methods

В статье [Adaptive stock trading strategies with deep reinforcement learning methods \(18\)](#), авторы рассмотрели GDPG модель объединяющую Q- Network from the GDQN system с policy network. Они использовали особенные подходы для обработки данных - так, например, для извлечения информации из необработанных данных и технических показателей в целях повышения точности и надежности они предложили использовать GRU (Gated Recurrent Unit). Кроме того, reward-функция учитывала риски при работе торговых стратегий GDQN и GDPG, что позволило показать достойные результаты для нестабильных и волатильных рынков. Модель получила расширенное пространство действий, не ограничивающееся лишь покупкой, продажей и удержанием бумаги в портфеле. В выводах и итогах работы авторы утверждают, что их модель превосходит обычный DRL, обеспечивает более стабильную доходность с поправкой на риск и превосходит базовую стратегию turtle trading, уже упомянутую ранее.

3.11 Deep reinforcement learning for trading

[Исследование \(19\)](#) рассматривает алгоритм A2C. При сравнении его работы с работой алгоритмов MACD, DQN, PG и торговлей в Long, оказывается, что A2C проигрывает лишь DQN. Алгоритм достигает 10-процентной годовой доходности, что является очень достойным результатом. Немаловажно, что алгоритм хорошо ведёт себя на медвежьем (падающем) рынке, а также слабо реагирует на высокую волатильность и продолжает почти непрерывный рост даже при серьезных процентных рыночных обвалах. Невысокая процентная доходность объясняется тем, что A2C генерирует большой оборот, что приводит к меньшей средней доходности на оборот. Эксперименты также показали, что алгоритм A2C может извлекать прибыль из быстро-растущего и быстро-падающего рынка через своевременное удержание существующие позиции.

3.12 Fairness in Financial Markets: The Case of High Frequency Trading

[Последняя статья \(20\)](#), которую я рассмотрю в своём обзоре, ставит научный вопрос как “является ли честной высокочастотная торговля (HFT)?”.

В обзоре мало упоминался термин HFT, поскольку мы больше обсуждали стратегии для инвестирования в лонг, и хоть алгоритмы высокочастотной торговли имеют свою специфику, я попробую со всех сторон проанализировать справедливость алгоритмов для торговли наряду с использованием в ней компьютерных методов.

Автор на базового введения в основные алгоритмы HFT, рассуждает о скорости работы алгоритмов, утверждая, что фирмы могут заполучить информационное и техническое преимущество, получая гораздо более широкий спектр рыночной информации, включая реальную стоимость активов и трудно вычисляемые индикаторы и показатели. Тем не менее, позже он пишет, что большинством инструментов, включая сервера с высокой вычислитель-

ной мощностью для торговли, может овладеть любой желающий трейдер, и это не создает никакой конкурентной несправедливости. Тем не менее, он не учитывает, ежегодно растущие издержки на необходимый для торговли и построения алгоритмов софт даже при наличии знания, как им распоряжаться. Мысля глобально, по сути, любые два отдельно взятых трейдера и так находятся в неравном отношении по возможности улучшать свой инвестиционный UI - даже покупка второго монитора для обычной не роботизированной торговли дает игроку преимуществу. Однако концентрация исключительной технической силы в руках очень крупных игроков всё же может неблагоприятно сказываться на рынке и некрупных игроках, как минимум банально ввиду снижения конкуренции. Тут лишь я позволю себе заметить, что крупные фирмы всегда владели большим спектром технологий для анализа рынка (даже когда о компьютерах не шло и речи), а значит, что алгоритмы если и усугубили проблему, то точно не создали ее.

Автор в своей работе старается структурированно рассмотреть возможные последствия алгоритмической торговли как для отдельно взятых трейдеров, не использующих софт в торговле, так и для всей экономики в целом. Так, он утверждает, что фактически нет смысла рассматривать влияние алгоритмической торговли на глобальное неравенство доходов, поскольку она поддерживает и провоцирует его не больше, чем прочие черты рыночной экономики. Рассматривая микроэкономические эффекты, автор утверждает, что большинство стратегий не вредят прочим игрокам на рынке, а значит не могут считать несправедливыми в бытовом смысле.

Автор упоминает, что злоумышленники могут использовать технологии торговли для проведения манипуляций на рынке при наличии больших объемов с целью получения большой прибыли путём обесценивания активов или, наоборот, искусственному завышению их стоимости. Тем не менее, подобные махинации существовали всегда - даже тогда, когда роботизированных торговых алгоритмов не существовало, а значит, что их использование никак не повлияло на существующую проблему. Более того, если отойти от рассмот-

рения алгоритмов к рассмотрению наличия компьютера как объекта на бирже, то использование различных алгоритмов регуляторами точно пошло на пользу среднестатистическому трейдеру, поскольку сильно упростился процесс выявления аномальных сделок на рынке.

Работа показалась мне интересной и полезной для анализа всей концепции алгоритмической торговли в целом. Говоря о недостатках, я не увидел в работе и намека на анализ бирж, где уже сегодня количество роботов преобладает над количеством людей-трейдеров, тогда как рассмотрение этой области могло подтолкнуть автора на новые мысли о том, насколько полезно постоянно растущее присутствие алгоритмов на фондовых биржах.

4 Заключение

Алгоритмическая торговля является многогранной, быстро-растущей и крайне интересной для научного анализа областью. С ней, как и с любой другой, связано много проблем - какие-то из них связаны с философскими и моральными сторонами алго-торговли, какие-то с возможностью контролировать наличие роботов на бирже, а какие-то - с оптимизацией скорости алгоритмов, вычислительной мощностью процессоров и нерешенными математическими задачами. Сказать можно лишь одно - эта область, появившаяся в научном и информационном поле относительно недавно, положила начало новому этапу в торговле и выполнении как стандартизированных торговых действий, так и разработке совершенно новых подходов к трейдингу. Ясно одно - несмотря на все попытки, хорошо это или плохо, но полностью воссоздать психологию, которой руководствуется опытный трейдер при принятии решений - невозможно, поскольку в ней, не говоря уж о сумасшедшем количестве факторов, влияющих на стоимость активов и оцениваемых игроком, всегда, каким бы профессионалом он ни был, присутствуют эмоции. А эмоции - это именно то, что, в конечном итоге, и делает нас отличными от набора алгоритмов в железной коробке.

Список литературы (или источников)

1. Machine learning models are becoming increasingly prevalent in algorithmic trading and investment management.
2. Lin Chen and Qiang Gao. Application of deep reinforcement learning on automated stock trading. In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), pages 29–33. IEEE, 2019.
3. Chien Yi Huang. Financial trading as a game: A deep reinforcement learning approach. arXiv preprint arXiv:1807.02787, 2018.
4. Xing Wu, Haolei Chen, Jianjia Wang, Luigi Troiano, Vincenzo Loia, and Hamido Fujita. Adaptive stock trading strategies with deep reinforcement learning methods. Information Sciences, 2020.
5. Curtis Faith. The original turtle trading rules, 2003.
6. Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. IEEE transactions on neural networks and learning systems, 28(3):653–664, 2016.
7. Zihao Zhang, Stefan Zohren, and Stephen Roberts. Deep reinforcement learning for trading. The Journal of Financial Data Science, 2(2):25–40, 2020.
8. Jamil Baz, Nicolas Granger, Campbell R Harvey, Nicolas Le Roux, and Sandy Rattray. Dissecting investment strategies in the cross section and time series. Available at SSRN 2695101, 2015.
9. Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. IEEE transactions on neural networks and learning systems, 28(3):653–664, 2016.

10. Nikhil R Pal and James C Bezdek. Measuring fuzzy uncertainty. *IEEE Transactions on Fuzzy Systems*, 2(2):107–118, 1994.
11. Chin-Teng Lin, C. S. George Lee, et al. Neural-network-based fuzzy logic control and decision system. *IEEE Transactions on computers*, 40(12):1320–1336, 1991.
12. Yue Deng, Youyong Kong, Feng Bao, and Qionghai Dai. Sparse coding-inspired optimal trading system for hft industry. *IEEE Transactions on Industrial Informatics*, 11(2):467–475, 2015.
13. WU Jia, WANG Chen, Lidong Xiong, and SUN Hongyong. Quantitative trading on stock market based on deep reinforcement learning. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.
14. Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. Available at SSRN, 2020.
15. Zhuoran Xiong, Xiao-Yang Liu, Shan Zhong, Hongyang Yang, and Anwar Walid. Practical deep reinforcement learning approach for stock trading. *arXiv preprint arXiv:1811.07522*, 2018.
16. Akhil Raj Azhikodan, Anvitha GK Bhat, and Mamatha V Jadhav. Stock trading bot using deep reinforcement learning. In *Innovations in Computer Science and Engineering*, pages 41–49. Springer, 2019.
17. Rudy Prabowo and Mike Thelwall. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157, 2009.
18. Xing Wu, Haolei Chen, Jianjia Wang, Luigi Troiano, Vincenzo Loia, and Hamido Fujita. Adaptive stock trading strategies with deep reinforcement learning methods. *Information Sciences*, 2020.

19. Zihao Zhang, Stefan Zohren, and Stephen Roberts. Deep reinforcement learning for trading. *The Journal of Financial Data Science*, 2(2):25–40, 2020.
20. James J. Angel, Douglas McCabe, Fairness in Financial Markets: The Case of High Frequency Trading *Journal of Business Ethics*, Vol. 112, No. 4, Special Issue: The 17th Annual International Vincentian Business Ethics Conference (2010) (February 2013), pp. 585-595 (11 pages)