

Федеральное государственное бюджетное образовательное учреждение
высшего образования «Московский государственный университет имени
М.В.Ломоносова»

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра Математической теории интеллектуальных систем

КУРСОВАЯ РАБОТА

TODO

Выполнил:

студент 431 группы

Зенин В. О.

Научный руководитель:

к.ф.-м.н., н.с Половников В. С.

Москва - 2023

Оглавление

1.	Введение	3
2.	Данные и извлечение признаков.....	4
2.1.	Информация из сети.....	4
2.2.	Подготовка данных.....	4
2.3.	Формирование обучающего множества	4

1. Введение

Различные нейросетевые подходы и архитектуры могут использоваться для работы с изменяющимися во времени данными. В процессе своего обучения они способны извлекать сложные нелинейные зависимости из данных и генерировать своё предсказание, основываясь на этом.

Существует много временных рядов, связанных с финансами, например, цены различного рода активов. Также можно найти описание паттернов движения цены, полученные путём анализа исторических данных биржевых котировок. Многие игроки используют их как основание для своих стратегий. Правила, образующиеся в результате найденных закономерностей, достаточно примитивны, как и сами паттерны. Если предположить, что кем-то найдена выгодная стратегия, то подобную способны найти и многие другие игроки, сводя на нет любую потенциальную выгоду. Вызывает интерес: способны ли нейронные сети находить паттерны и, тем самым, определять приносящие доход стратегии торговли, скрытые от большинства игроков.

Информация о классических финансовых инструментах во многом скрыта от игроков и хранится на биржах. Финансовые транзакции также скрыты за межбанковским обменом и не поддаются анализу. Однако существуют набирающую популярность криптофинансовые активы, информация о которых, по своей природе, намного более открыта и может быть использована для анализа движения цены.

Цель данной работы – Исследование доступной публично информации о криптовалютах, построение нескольких архитектур нейронных сетей для анализа исторических данных и построение прогноза изменения будущей цены актива.

Основными задачами курсовой являются:

- Изучение существующих данных в блокчейне Bitcoin.
- Практическая реализация моделей на базе рекуррентных нейронных сетей и архитектуре трансформера.
- Постановка задач предсказания движения цены как задачи регрессии и классификации.
- Сравнение полученных результатов между собой и определение перспектив подобных исследований.

2. Данные и извлечение признаков

2.1. Информация из сети

В данной работе использованы дневные наблюдения о состоянии блокчейн сети Bitcoin с 10 мая 2020 года по 8 мая 2023 года, полученные с blockchain.com. Некоторые базовые признаки также вычислены заранее поставщиком данных. Их описания собраны в таблице 1.¹

2.2. Подготовка данных

При работе с ценой актива часто используется логарифм цены,

$$\ln\left(\frac{x_t}{x_{t-1}}\right)$$

позволяющий перейти от абсолютных значений к относительным. Смысл данного преобразования заключается в том, что успешная стратегия приносит доход в результате изменения цен, умноженных на вложенный капитал и именно доход имеет ключевое значение.

Входные данные для нейронных сетей следует скалировать. Однако некоторые признаки в наших данных имеют количественную природу, что выражается в почти линейном росте. Например, абсолютное значение добытых на момент времени t монет BTC. Большой смысл имеет изменение в добыче, так как оно потенциально способно дать сигнал о будущих движениях цены. Поэтому в нашем случае подобное преобразование уместно применить ко всем признакам.

2.3. Формирование обучающего множества

До логарифмирования имелось 1094 векторов, размерности 27 каждый. В результате преобразование наблюдение за первый день вырождается и остается 1093 вектора значений.

Для обучения использовались значения до 15 июня 2022 года. Для валидации – с 15 июня 2022 года по 20 января 2023 года. Для теста – с 20 января 2023 года по 8 мая 2023 года. Данные временные диапазоны выбраны чтобы обеспечить соотношение 70:20:10.

Целевой признак – market-price.

Сформируем из данных следующие пары:

$$(X_{[m;t]}, Y_t),$$

где $X_{[m;t]} = (x_{t-m}, x_{t-m-1}, \dots, x_{t-2}, x_{t-1})$ – подпоследовательность длины m , x_t – вектор признаков для момента времени t , Y_t – значение целевого признака. Для задачи регрессии $Y_t = y_t$,

где y_t логарифм цены. Для задачи классификации $Y_t = \begin{cases} 1, y_t > 0 \\ 0, y_t \leq 0 \end{cases}$

¹Признаки, отмеченные (*) имеют не более 3 пропущенных значений, которые восстановлены линейной интерполяцией.

Таблица 1: Признаки из блокчейн сети Bitcoin

Признак	Описание
total-bitcoins (*)	Количество добытых монет
market-price	Средняя цена в USD на крупнейших обменниках
trade-volume	Объем обменянных BTC (USD)
blocks-size	Размер сети блокчейна (MB)
avg-block-size	Средний размер блока (MB)
n-transactions-total	Количество транзакций
n-transactions-per-block	Среднее число транзакций на блок
n-payments-per-block	Среднее число наград за валидированный блок
median-confirmation-time	Медианное время, за которое обработанная транзакция добавляется к сети
avg-confirmation-time	Среднее время, за которое обработанная транзакция добавляется к сети
hash-rate	Мощность сети
difficulty	Относительная мера сложности сети – насколько трудно валидировать очередной блок
transaction-fees	Выплаченные BTC за валидацию блоков
transaction-fees-usd	Выплаченные USD за валидацию блоков
fees-usd-per-transaction	Средняя выплата в USD за валидированную транзакцию
cost-per-transaction	Общий доход майнеров, разделённый на количество транзакций
n-unique-addresses (*)	Количество уникальных адресов, используемых в сети
n-transactions	Количество подтверждённых транзакций за день
n-payments	Количество подтверждённых выплат за день
mempool-count	Количество неподтверждённых транзакций
mempool-growth	Рост хранилища неподтверждённых транзакций
mempool-size	Размер хранилища неподтверждённых транзакций
n-transactions-excluding-popular	Количество транзакций, за исключением 100 самых популярных адресов
estimated-transaction-volume (*)	Оценочная стоимость транзакций (BTC)
estimated-transaction-volume-usd (*)	Оценочная стоимость транзакций (USD)