

УДК 004.032.26

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ НЕЙРОННЫХ СЕТЕЙ, ОБУЧЕННЫХ НА СОБРАННЫХ ПРОГРАММНЫМ ОБРАЗОМ ДАННЫХ, ДЛЯ ПРОГНОЗИРОВАНИЯ ЗАДАННЫХ КОТИРОВОК АКЦИЙ

© Раздьяконов А.А., Савельев Д.А.

Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация

e-mail: artem.razd_nov@mail.ru

Финансы и экономика – неотъемлемый атрибут человеческих отношений в настоящее время. За 2020 год в России было открыто физическими лицами почти 5 миллионов новых брокерских счетов на Московской бирже. Общее число брокерских счетов достигло 8,8 миллиона. Таким образом, в прошлом году на Московскую биржу пришло частных инвесторов больше, чем суммарно за все предыдущие годы [1].

Использование машинного обучения – эффективная стратегия, которую для увеличения доходов используют многие инвестиционные фонды. Ведь изменения в сфере финансов происходят нелинейно, и иногда может показаться, что цены на акции формируются совершенно случайным образом [2]. Для анализа котировок необходимо предварительно собрать большое количество данных, желательно в их исторической ретроспективе [3]. Много данных в удобной форме представлено на сайте Финам [4]. Их можно загружать с ресурса в виде, например, файла формата csv. Однако лучше автоматизировать этот процесс для случаев, когда предстоит анализировать датасеты большого объема. Разработанный для решения данной задачи скрипт на языке программирования Python позволяет формировать запрос к сервису для получения файла с необходимыми для изучения котировками акций. Главной задачей скрипта, изображенного на рисунке, является формирование строки определенного вида, которая направляется на сайт Финам, получает там данные с указанными параметрами и возвращает их в виде файла csv.

Полученная таблица данных содержит в себе цену открытия на акции Сбербанка, цену закрытия, максимальную и минимальные цены, а также объем торгов. Однако для обучения нейронных моделей сперва к датасету добавлены колонки курса доллара и нефти за соответствующие дни. Это сделано с целью увеличения точности предсказаний, так как между ценой рассматриваемых акций и двумя вышеупомянутыми экономическими единицами существует корреляция.

Для создания, обучения и дальнейших прогнозов используется функционал пакета sklearn. В исследовании применяются две нейронные сети: MLPRegressor, предсказывающий непрерывные значения цены акции на момент закрытия торгов на день, и MLPClassifier, определяющий то, вырастет ли цена закрытия торгов относительно цены открытия. Для проведения сравнения качества полученных моделей используется метрика MAE («средняя абсолютная ошибка», англ. Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t|,$$

где $e_t = original_t - predict_t$, $original_t$ – реальное значение целевой переменной на объекте t , $predict_t$ – предсказанное значение целевой переменной на объекте t , n – количество объектов тестовой выборки.

```
# упаковка параметров в единую структуру
params = urlencode([
    ('market', market),
    ('em', tickers[ticker]),
    ('code', ticker),
    ('apply', 0),
    ('df', start_date.day),
    ('mf', start_date.month - 1),
    ('yf', start_date.year),
    ('from', start_date),
    ('dt', end_date.day),
    ('mt', end_date.month - 1),
    ('yt', end_date.year),

    ('to', end_date),
    ('p', period),
    ('f', ticker + "_" + start_date_rev + "_" + end_date_rev), #Имя сформированного файла
    ('e', ".csv"), #Расширение сформированного файла
    ('cn', ticker), # тикер акции
    ('dtf', 1), #В каком формате брать даты.
    ('tmf', 1), #В каком формате брать время.
    ('MSOR', 0), #Время свечи
    ('mstime', "on"), #Московское время
    ('mstimever', 1), #Коррекция часового пояса
```

Рис. Скрипт для скачивания данных с Финам

Закономерным образом качество метрики в случае классификатора выше фактически вдвое, так как предсказать, к какой группе относится та или иная запись в таблице, проще, чем спрогнозировать конкретное число – цену акции на момент закрытия торгов. Также был проведен эксперимент, результаты которого представлены в таблице: из исходного датасета убраны колонки цены курса доллара и нефти, которые были добавлены для увеличения точности прогнозов. Однако эксперимент показал, что в случае данного исследования метрика качества моделей без этих колонок значительно лучше, то есть наша гипотеза о повышении точности работы нейронных сетей в случае наличия дополнительных столбцов опровергнута. Эти столбцы увеличивают меру хаоса в данных, поэтому в погоне за точностью в данном случае необходимо отказаться от них.

Таблица. Точность прогноза для рассматриваемых нейронных сетей при разных исходных данных

Нейронная сеть	Данные с Финам	Нефть	Доллар	Доллар+Нефть
MLPRegressor	0.487	0.477	0.647	0.822
MLPClassifier	0.493	0.494	0.495	0.494

Библиографический список

1. Количество инвесторов-физлиц Мосбиржи в 2020 году выросло на 5 млн. URL: <https://tass.ru/ekonomika/10453365> (дата обращения: 11.02.2021).
2. Абаньшин П.А., Барышникова М.Ю. Сравнительный анализ методов автоматизированного прогнозирования изменения цен акций на фондовом рынке // Вестник науки и образования. 2020. № 3 (81). Ч. 1. С. 24–27.
3. Вайсман Ричард. Механические торговые системы. Психология трейдинга и технический анализ. М.: Альпина Пабlishер, 2019. 535 с.
4. Финам. URL: <https://www.finam.ru> (дата обращения: 04.12.2020).