

Применение методов машинного обучения для построения прогнозов на финансовых рынках

*Волоцкой Дмитрий Владимирович, магистрант,
Национальный исследовательский университет ИТМО, г. Санкт-Петербург*

Аннотация

Фондовый рынок — публичная торговая площадка для покупки и продажи ценных бумаг различных компаний. Предсказание стоимости акций той или иной компании является важной аналитической задачей. Успех в точном предсказании позволяет компаниям и инвесторам преумножать свой капитал.

Алгоритмическая торговля в последние десятилетия активно развивается благодаря различным факторам: развитию методов машинного обучения, развитию технологий обработки и анализа данных, развитию систем хранения данных. Помимо этого, конкуренция на финансовых рынках приводит к поиску все новых возможностей для увеличения точности прогнозов. В связи с этими тенденциями исследование применимости методов машинного обучения к задачам алгоритмической торговли является актуальной задачей.

В данной работе рассматривается возможность применения методов машинного обучения для построения прогнозов на финансовых рынках, а также разработана и протестирована модель для построения краткосрочных прогнозов. Была проведена оценка ее эффективности.

Ключевые слова: машинное обучение, финансовые рынки, LSTM, прогнозирование котировок, анализ

Application of machine learning methods for building forecasts in financial markets

Annotation

The stock market is a public trading platform for buying and selling shares of various companies. Predicting the value of a company's shares is an important analytical task. Success in accurate predictions allows companies and investors to multiply their capital.

Algorithmic trading has been actively developing in recent decades due to various factors: the development of machine learning methods, the development of data processing and analysis technologies, and the development of data storage systems. In addition, competition in the financial markets leads to the search for new opportunities to increase the accuracy of forecasts. Due to these trends, the study of the applicability of machine learning methods to the problems of algorithmic trading is an important task.

This paper discusses the possibility of using machine learning methods to build forecast models for financial markets. The model for building short-term forecasts was developed and tested. Its effectiveness has been evaluated.

Keywords: machine learning, financial markets, LSTM, quote forecasting, analysis

Введение

С самого зарождения фондовых рынков люди искали и развивали способы максимально точно предсказывать котировки ценных бумаг. В итоге сформировались три основных подхода: фундаментальный анализ, технический анализ и технологические методы. Несмотря на порой взаимоисключающие результаты использования этих подходов в рамках одной задачи (одни предсказывают будущее значение, другие — направление движения для заданного периода времени), все они используются для торговли на фондовых рынках. При этом задача повышения точности прогноза остается актуальной для всех компаний, работающих на фондовых рынках [1].

Машинное обучение, как раздел науки, занимается изучением алгоритмов и статистических моделей. Это разновидность искусственного интеллекта, которая дает компьютерам возможность учиться на данных и выполнять задачи без явных пошаговых инструкций, опираясь только на данные, которые модель использует для обучения. Машинное обучение используется в различных приложениях, как правило, там, где сложно создать обычный алгоритм решения задачи.

В связи с развитием компьютерных технологий, расширением сфер применения машинного обучения и повышения производительности моделей, исследование применимости различных алгоритмов машинного обучения к задачам алгоритмической торговли является актуальной задачей. Интерес к подобным исследованиям проявляют не только организации, занимающиеся инвестициями, но и научное сообщество, поскольку применение алгоритмов машинного обучения к рассматриваемой области способно привнести новые знания в развитие машинного обучения как раздела компьютерных наук, которые могут быть применены к другим предметным областям.

Обзор применения методов машинного обучения в предметной области

В результате развития технологий прогнозирование цен на акции переместились в технологическое пространство. Некоторые из лучших методов прогнозирования включают использование рекуррентных нейронных сетей (RNN) и искусственных нейронных сетей (ANN), которые также подпадают под категорию машинного обучения [2].

В результате исследования предметной области для обучения была выбрана нейронная сеть LSTM (Long Short-term memory). Её достоинствами являются работа с историческими данными и более высокая точность прогноза [1, 2, 3, 4, 5]. Кроме того, данная модель показывает свое превосходство в точности предсказаний в сравнении с другими моделями, например моделью

ARIMA. Значение ошибки RMSE модели LSTM меньше соответствующей ошибки для модели ARIMA в среднем на 65% [1]. LSTM, созданные на основе Recurrent Neural Network (RNN), может решать проблему долгосрочных зависимостей и может применять историческую информацию для решения текущей задачи [4,5].

LSTM нейронные сети хорошо подходят для классификации, обработки и построения прогнозов на основе временных рядов, где взаимосвязанные явления могут происходить с неопределенным временным лагом [1, 4, 5]. Этот временной лаг приводит к затруднениям в использовании классических нейронных сетей в решении данных проблем из-за затухания градиента, в то время как LSTM сети нечувствительны к величине временного лага. В LSTM сети существуют 3 типа слоев [1]:

1. Слой «забывания» (Forget gate) — на выход подается число от 0 до 1, где 1 обозначает необходимость полного запоминания, а 0 полностью стирает из памяти.

2. Слой «запоминания» (Memory gate) выбирает, какие данные необходимо сохранить. В первую очередь с помощью сигмоидного слоя выбираются значения, которые затем запоминаются.

3. Выходной слой (Output gate) выбирает информацию из каждой «ячейки», в которой произведено запоминание.

Обучение и тестирование модели

Для разработки модуля предсказаний на основе модели LSTM был использован язык программирования Python и библиотеки pandas и keras. Библиотека pandas использовалась для подготовки дата-сетов, а библиотека keras реализует функционал такой библиотеки как TensorFlow.

Обучение и тестирование прототипа было произведено с использованием данных по акциям компаний Yandex, Сбербанк и Лукойл. Кроме того, были выбраны временные рамки для тренировочных данных. Для обучения были выбраны данные с 1 сентября 2021 года по 16 февраля 2022 года в связи со стабильностью финансового рынка на этом временном промежутке.

Для оценки производительности модели и точности ее предсказаний рассмотрим ее математическую оценку (Рисунок 1), сравнение графика предсказанных значений с реальными (Рисунок 2) и отклонение предсказанного значения цены акции на следующий день от её реального значения (Рисунок 3). Рассмотрим показатели производительности модели:

- **Средняя абсолютная ошибка (MAE)** – среднее отклонение предсказанных значений временного ряда от реальных

- **Средняя абсолютная ошибка в процентах (MAPE)** – среднее процентная разница между предсказанием и реальным значением временного ряда [2].

- **Медианная абсолютная ошибка в процентах (MDAPE)** – медианное значение разницы между предсказанным и реальным значением временного ряда.

```
Median Absolute Error (MAE): 3.08
Mean Absolute Percentage Error (MAPE): 6.58 %
Median Absolute Percentage Error (MDAPE): 6.26 %
```

Рисунок 1 – Оценка производительности модели для акций «Yandex»



Рисунок 2 – Сравнение предсказанных и реальных значений

```
The close price for Yandex at 2022-02-16 was 52.4
The predicted close price is 47.84000015258789 (-9.53%)
```

Рисунок 3 – Отличие предсказанной цены от реальной

Проведем дополнительное тестирование модели. Построим предсказания для Сбербанка (Рисунок 4, 5) и Лукойла (Рисунок 6, 7) на том же временном промежутке и оценим точность предсказания прототипа.



Рисунок 4 – Сравнение предсказанной и реальной стоимости акций «Сбербанк»

```
Median Absolute Error (MAE): 0.94
Mean Absolute Percentage Error (MAPE): 7.17 %
Median Absolute Percentage Error (MDAPE): 5.74 %
The close price for Sberbank of Russia at 2022-02-16 was 14.97
The predicted close price is 14.15999984741211 (-5.72%)
```

Рисунок 5 – Оценка производительности модели для предсказаний акций «Сбербанк»



Рисунок 6 – Сравнение реальной и предсказанной стоимости акций «Лукойл»

```
Median Absolute Error (MAE): 3.82
Mean Absolute Percentage Error (MAPE): 4.37 %
Median Absolute Percentage Error (MDAPE): 3.95 %
The close price for PJSC LUKOIL at 2022-02-16 was 92.64
The predicted close price is 89.5 (-3.51%)
```

Рисунок 7 – Оценка производительности модели для акций «Лукойл»

Таким образом, мы получим среднюю точность предсказаний для Яндекса – 93,42%, для Сбербанка – 92,83 %, и Лукойла – 95,63%. Полные значения точности предсказаний модели представлены в таблице 1.

Таблица 1. Оценка точности предсказаний модели.

Набор данных	Средняя точность предсказаний	Точность предсказаний на день вперед
Яндекс	93,42%	91,29%
Сбербанк	92,83%	94,59%
Лукойл	95,63%	96,61%

Заключение

В результате выполнения данной работы был проведен анализ применения методов машинного обучения для построения прогнозов на финансовых рынка. Был разработан и протестирован прототип системы построения краткосрочных прогнозов на основе модели LSTM. Было проведено дальнейшее тестирование модели с дополнительными наборами данных.

Разработанный прототип показал достаточную точность предсказаний. Но система показала себя недостаточно точно в предсказании резких скачков и падений цены на акции. В дальнейшем планируется продолжить разработку прототипа системы с целью увеличить точность предсказания и увеличить временной отрезок, на который система сможет делать предсказания с приемлемой точностью.

Литература

1. Алжеев А.В., Кочкаров Р.А. Сравнительный анализ прогнозных моделей ARIMA и LSTM на примере акций российских компаний // Финансы: теория и практика. - 2020. №1.
2. Emioma C.C., Edeki S.O. Stock price prediction using machine learning on least-squares linear regression basis // Journal of Physics: Conference Series. - 2021, том 1734, № 1.
3. Manurung A.H., Budiharto W., Prabowo H. Algorithm and modeling of stock prices forecasting based on long short-term memory (LSTM) // ICIC Express Letters. 2018, том 12, № 12, стр. 1277-1283
4. Воинов Н. В., Ворошилов М. К., Молодяков С. А., Дробинцев П. Д., Прокофьев О. В., Зайцев И. В. Прогнозирование котировок фьючерсов на индекс РТС на основе машинного обучения // XXIV Международная конференция по мягким вычислениям и измерениям (SCM'2021) – 2021.
5. Sun Z., Zhao S. Machine learning in stock price forecast // E3S Web of Conferences. - 2020, том 214