



Моделирование изменения цены биржевого инструмента на базе микроструктурных рыночных данных

Н.А. Билев,

Московский государственный университет
им. М.В. Ломоносова,
Москва, Россия
<https://orcid.org/0000-0002-0321-785X>

АННОТАЦИЯ

На современных биржевых площадках, где большинство операций совершается посредством электронных транзакций, появляется возможность анализировать событийные микроструктурные рыночные данные. Они описывают физическое формирование цены на биржевые инструменты и обладают высокой информативностью, позволяя торговой системе находить сложные закономерности в поведении цены. Выявление таких закономерностей вручную является очень трудоемким процессом. Однако современные модели машинного обучения способны решать подобные задачи автоматически, подстраиваясь под постоянно меняющееся поведение рынка. В данной работе разрабатывается торговая система на базе модели машинного обучения и микроструктурных рыночных данных. Для исследования собрана информация с московской биржи ММВБ о событийных изменениях биржевой книги заявок и ленте всех сделок по ликвидному биржевому инструменту. Для моделирования используется модель логистической регрессии и ряд моделей искусственных нейронных сетей. Результаты исследования демонстрируют, что модель логистической регрессии не уступает в качестве прогнозирования более сложным моделям и при этом имеет высокую скорость формирования прогноза, что очень важно при принятии торгового решения на современных торговых площадках. Разработанная торговая система имеет среднюю частоту заключения сделок, что позволяет избежать дорогой инфраструктуры по сравнению с высокочастотной биржевой торговлей, но при этом дает возможность использовать весь потенциал высококачественных микроструктурных рыночных данных. Статья описывает все этапы построения торговой системы, включая выбор признаков для моделирования, сравнительный анализ моделей прогнозирования изменения цены, а также создание торгового алгоритма с тестированием на исторических данных. Он может применяться различными инвестиционными институтами для эффективного управления капиталом в биржевых торгах. Разработка более сложных и детальных торговых алгоритмов на базе модели машинного обучения позволит увеличить конечную эффективность всей торговой системы.

Ключевые слова: биржевая торговля; алгоритмическая торговля; микроструктурные рыночные данные; биржевая книга заявок; машинное обучение; временные ряды; инвестиции; управление капиталом

Для цитирования: Билев Н.А. Моделирование изменения цены биржевого инструмента на базе микроструктурных рыночных данных. *Финансы: теория и практика.* 2018;22(5):141-153. DOI: 10.26794/2587-5671-2018-22-5-141-153

Modeling Stock Price Changes Based on Microstructural Market Data

N.A. Bilev,

Lomonosov Moscow State University,
Moscow, Russia
<https://orcid.org/0000-0002-0321-785X>

ABSTRACT

In the modern electronic stock exchanges it is possible to analyze event microstructural market data. This data is highly informative and describes physical pricing which allows to find complex patterns in price behaviour. Discovering such patterns manually is hard and time consuming. However, modern machine learning models are capable of solving such

issues automatically adapting to changing price behavior. This research presents a trading system based on a machine learning model and market microstructure data. The data for the research has been collected from the Moscow stock exchange MICEX and provides event changes in the exchange book of applications and the tape of all transactions in a liquid stock exchange instrument. Logistic regression model has been used and compared to neural network models with different configuration. According to the study results, the logistic regression model is not inferior in forecasting quality to more complex models and at the same time it has a high forecast formation rate. It is very important when making trading decisions in the modern trading market. The developed trading system has an average frequency of transactions which allows to avoid expensive infrastructure compared to high-frequency exchange trading. However, it also allows to use the full potential of high-quality microstructural market data. The article describes all the stages of a trading system development, including feature engineering, a comparative analysis of price change forecasting models, and a trading algorithm with testing for historical data. It can be used by various investment institutions for the capital effective management in exchange trading. The development of more complex and detailed trading algorithms based on a machine learning model will increase the ultimate efficiency of the entire trading system.

Keywords: stock market trading; algorithmic trading; market microstructure; limit order book; machine learning; time series; investment; capital management

For citation: Bilev N.A. Modeling stock price changes based on microstructural market data. *Finansy: teoriya i praktika = Finance: Theory and Practice*. 2018;22(5):141-153. (In Russ.). DOI: 10.26794/2587-5671-2018-22-5-141-153

ВВЕДЕНИЕ

Главная функция инвестиционного института — эффективное управление капиталом. На сегодняшний день инвестиционные фонды активно ищут новые способы инвестирования капитала с минимальным риском на фоне напряженной геополитической обстановки в мире. Большинство классических стратегий инвестирования средств на базе фундаментальной оценки стоимости актива подвержены риску в периоды общего спада экономики под воздействием тех или иных глобальных факторов. Автоматизированные алгоритмы торговли направлены на устранение нестабильности биржевых показателей и являются эффективными в условиях высокой неопределенности на рынке. В настоящий момент большинство научных работ в области биржевой автоматизированной торговли посвящены разработке высокочастотных торговых стратегий. Известно, что подобные алгоритмы заметно искажают цены активов, сильно зависят от биржевого регулирования и имеют очень дорогую технологическую инфраструктуру [1, 2]. Исходя из неспособности классических стратегий торговли поддерживать высокую прибыльность инвестиций в моменты неопределенности на рынке и указанных негативных черт высокочастотных стратегий, представляется актуальной разработка алгоритмов торговли, направленных одновременно на стабилизацию рынка и максимизацию доходности инвестиций. Развитие технологий биржевой инфраструктуры сделало возможным анализ событийных микроструктурных рыночных данных (биржевая книга заявок, объемы всех сделок и количество открытых позиций) [3, 4], отражающих в себе много-

го больше информации о состоянии рынка, чем интервальные ценовые индикаторы, используемые во многих научных работах в данной сфере. Микроструктурные рыночные данные описывают физическое формирование цены биржевого инструмента на фондовом рынке. Такие данные являются событийными, что позволяет получать информацию об активе не за определенный интервал времени, а непосредственно по факту возникновения события. При использовании микроструктурных рыночных данных автоматизированная торговая система принимает решения на базе огромного количества информации о формировании цены всеми участниками рынка. Наиболее распространенным инструментом принятия торговых решений являются логические правила, сформированные вручную путем тщательного изучения закономерностей поведения цены на биржевой инструмент за определенный период времени. Такой подход имеет серьезный недостаток, заключающийся в том, что закономерности в рыночных данных постоянно меняются и поиск новых зависимостей требует много времени и других ресурсов. Уровень алгоритмов машинного обучения и вычислительных возможностей в настоящее время позволяет автоматически находить сложные закономерности при условии, что в данных достаточно информации для их выявления. Таким образом, наиболее актуальным подходом для принятия торговых решений является применение алгоритмов машинного обучения на базе микроструктурных событийных рыночных данных, максимально детально описывающих физическое формирование цены на биржевые инструменты.

ПОДХОДЫ К РАЗРАБОТКЕ АВТОМАТИЗИРОВАННЫХ ТОРГОВЫХ СИСТЕМ

Основным ресурсом для создания автоматизированной торговой системы являются данные, на базе которых алгоритм принимает торговые решения. Существуют два основных типа данных:

- микроструктурные рыночные данные (биржевая книга заявок и сведения о всех сделках по торговому инструменту);
- интервальные данные цен активов и ценовые индикаторы.

Использование микроструктурных рыночных данных в основном распространено в высокочастотной алгоритмической торговле [5–8]. Высокочастотные алгоритмы совершают огромное количество мелких по объему сделок и нацелены на использование закономерностей в биржевой книге заявок и ленте всех сделок [9]. При такой высокой частоте заключения сделок скорость совершения транзакции становится очень важным фактором [10]. Поэтому высокочастотная торговля имеет очень дорогую инфраструктуру и самые высокие издержки на комиссионные сборы.

Интервальные данные представляют собой агрегацию микроструктурных данных за определенный период времени. Таким образом, сокращается количество информации, анализируемой торговым алгоритмом и, соответственно, потенциал эффективности работы всей торговой системы. Однако, в отличие от микроструктурных рыночных данных, интервальные данные более распространены в различном программном обеспечении для анализа биржевой информации.

В настоящее время для принятия торговых решений используются различные инструменты. Данная работа выделяет 2 основных вида моделирования процессов биржевой торговли:

- логические правила;
- модели на базе алгоритмов машинного обучения.

Наиболее популярным видом моделирования биржевой торговли являются логические правила, которые представляют собой набор закономерностей, формирующих правила поведения цены торгового инструмента в определенных условиях. Очень большое количество книг посвящено именно этому подходу. Напротив, научных публикаций по этой теме совсем немного. Это связано, прежде всего, с тем, что торговые стратегии на базе логических правил имеют достаточно простой математический аппарат. Несмотря на это, такие алгоритмы являются наиболее прибыльными

и удобными в использовании. Большой недостаток логических правил заключается в том, что с течением времени они перестают работать за счет того, что закономерности в поведении цены на бирже постоянно меняются. Соответственно, инвестор, принимающий решение на базе логических правил, нуждается в постоянном поиске новых закономерностей в поведении цены. Это влечет за собой издержки на разработку новой торговой системы под новые логические правила.

Все большую популярность во многих сферах, предполагающих анализ данных, набирают методы машинного обучения, представляющие собой эмпирические модели, решающие задачи регрессии, классификации и кластеризации [11]. Решение задачи строится обычно по следующему сценарию:

- 1) определение целевой переменной;
- 2) поиск наиболее информативных переменных-признаков, максимально хорошо описывающих целевую переменную;
- 3) разделение выборки на тренировочную и тестовую;
- 4) обучение модели на тренировочной выборке;
- 5) тестирование модели на тестовой выборке.

В биржевой торговле целевой переменной обычно выступает изменение цены. Иногда авторы используют бинарную целевую переменную, отвечающую на вопрос: вырастет ли цена за следующий определенный интервал времени или нет? В качестве признаков, в зависимости от типа данных, выступают различные агрегаты предыдущей динамики цены и связанных с ней данных [12]. В отличие от логических правил, модель на базе методов машинного обучения способна приспособиться к меняющимся рыночным условиям за счет регулярного обучения модели на поступающих данных для поиска новых закономерностей в поведении цены.

Суммируя вышесказанное, можно сделать вывод, что наиболее перспективным направлением для построения торговой системы является торговая стратегия со средней частотой заключения сделок, которая принимает торговые решения с помощью модели машинного обучения, построенной на базе микроструктурных рыночных данных. Такой подход имеет следующие преимущества:

- микроструктурные данные имеют очень низкую латентность и, соответственно, дают очень высокое качество симуляции рыночных событий;
- по сравнению с высокочастотным видом торговых стратегий, среднечастотный вид не нуждается в существенных расходах на инфраструктуру;

- низкие комиссионные сборы в сравнении с высокочастотным типом торговли;
- отсутствует негативное влияние на биржевую инфраструктуру;
- длительная работоспособность стратегии за счет постоянного переобучения модели.

ПОСТАНОВКА ЗАДАЧИ МОДЕЛИРОВАНИЯ ИЗМЕНЕНИЯ ЦЕНЫ АКТИВА

В данной работе используются микроструктурные рыночные данные по биржевому инструменту, фьючерсу на индекс RTS с экспирацией 15 марта 2018 г., за период с 26.02.2018 по 07.03.2018 г. Данные были собраны с биржи ММВБ и состоят из событий изменения биржевой книги заявок и рыночных сделок за указанный период. Биржевая книга заявок фиксирует очередь цен на покупку $p_b(t)$ и на продажу $p_a(t)$ с соответствующими объемами выставленных заявок на покупку $v_b(t)$ и на продажу $v_a(t)$. Для моделирования изменения цены используется секундная дискретизация данных, т.е. на каждую секунду имеется актуальное состояние биржевой книги заявок. В качестве второго основного источника данных имеются объемы сделок на покупку $d_b(t)$ и на продажу $d_s(t)$. Кроме этого, используются интервальные биржевые данные с периодом 1 минута. Эти данные отражают изменения цены, и на каждый интервал времени имеется 4 значения: цена открытия интервала $po(t)$, цена закрытия интервала $pc(t)$, максимальная цена интервала $ph(t)$, минимальная цена интервала $pl(t)$.

Моделируемая переменная представляет собой изменения цены за будущий интервал. В данном исследовании под ценой подразумевается средняя цена между лучшей ценой покупки и лучшей ценой продажи:

$$p_t = \frac{p_a^1(t) + p_b^1(t)}{2}. \quad (1)$$

Динамика изменения цены с секундной величиной интервала имеет очень много шума [13]. Поэтому для подсчета моделируемой переменной, а именно изменения цены за k интервалов, мы используем сглаживание за соответствующее количество интервалов:

$$m_b(t) = \frac{1}{k} \sum_{i=1}^k p_{t-i}, \quad (2)$$

$$m_a(t) = \frac{1}{k} \sum_{i=1}^k p_{t+i}, \quad (3)$$

$$l_t = \begin{cases} 1, m_a(t) > m_b(t) + \alpha \\ -1, m_a(t) < m_b(t) - \alpha \\ 0, \text{ в остальных случаях} \end{cases}, \quad (4)$$

где α — минимальное изменения цены, при котором мы считаем, что цена выросла или упала. Моделируемая переменная l_t принимает 3 значения: 1 — рост, -1 — падение, 0 — цена значительно не изменилась. Таким образом, в данной работе решается задача классификации на 3 класса.

Для решения поставленной задачи на базе сырых данных строятся признаки для классификации. Для предсказания следующего изменения цены логично использовать предыдущие изменения, предполагая, что динамика цены в будущем зависит от динамики цены k периодов назад. Помимо этого, будем использовать усредненные прошлые изменения:

$$p_{mean}(t) = \frac{1}{k} \sum_{i=1}^k p_{t-i}. \quad (5)$$

Объемы на различных ценовых уровнях в биржевой книге заявок физически формируют цену на биржевой инструмент. Появление больших объемов может сдерживать цену или формировать определенный паттерн ее поведения в будущем. В качестве признаков будем использовать суммы объемов от лучшей цены до d -й цены для заявок на покупку и на продажу:

$$bs_a(t, d) = \sum_{i=1}^d v_a^i, \quad (6)$$

$$bs_b(t, d) = \sum_{i=1}^d v_b^i. \quad (7)$$

Рыночные сделки отражают настроение участников торгов и создают изменения в биржевой книге заявок. В динамике объемов рыночных сделок формируются закономерности в поведении цены в будущем. Для того чтобы классификатор мог распознать эти закономерности, создадим набор признаков на базе объемов рыночных сделок. На каждый интервал времени будем подсчитывать как количество сделок на покупку и продажу, так и суммарный объем сделок в двух направлениях по отдельности. Для устранения шума в данных будем использовать усредненные значения за k периодов в прошлом:

$$dc_{mean}(t) = \frac{1}{k} \sum_{i=1}^k dc_{t-i}, \quad (8)$$

$$ds_{mean}(t) = \sum_{i=1}^k ds_{t-i}, \quad (9)$$

где dc_t — количество сделок за период t ; ds_t — сумма объемов сделок за период t .

Самым распространенным инструментом технического анализа является график японских свечей и различные индикаторы, построенные на его базе. Многие участники рынка принимают решения о покупке или продаже биржевых активов, руководствуясь именно этими инструментами. Такое поведение большого количества игроков отражается в поведении цены. В связи с этим построим популярные индикаторы технического анализа и будем использовать их как дополнительные признаки в модели классификации. В качестве индикаторов были выбраны расстояние до простого скользящего среднего $sma(t)$, средний истинный диапазон $atr(t)$ и индекс относительной силы $rsi(t)$ [14]:

$$sma(t) = p_t - \frac{1}{k} \sum_{i=0}^k pc_{t-i}, \quad (10)$$

$$tr(t) = \max(|ph_t - pl_t|, |ph_t - pc_{t-1}|, |pl_t - pc_{t-1}|), \quad (11)$$

$$atr(t) = \frac{1}{k} \sum_{i=1}^k tr_{t-i}, \quad (12)$$

$$u_t = |pc_t - pc_{t-1} > 0|, d_t = |pc_t - pc_{t-1} < 0|, \quad (13)$$

$$rsi(t) = 100 - \frac{100}{1 + \frac{\sum_{i=1}^k u_{t-i}}{\sum_{i=1}^k d_{t-i}}}. \quad (14)$$

Значения локальных экстремумов в динамике цены биржевого инструмента могут влиять на ее изменение в будущем. Зачастую в окрестности данных точек формируется относительно большое количество объемов в биржевой книге заявок, сдерживая цену в направлении экстремума. Помимо этого, участники биржевых торгов склонны выставлять заявки на фиксацию убытка за локальными экстремумами. Соответственно, цена имеет специфичную динамику вблизи данных ценовых уровней. Для того чтобы модель классификации учитывала данную специфику, будем использовать в качестве признаков расстоя-

ние от текущей цены биржевого инструмента до ближайшего локального максимума и минимума:

$$dp_{max}(t) = ph_{t-1} - p_t; ph_{t-2} < ph_{t-1} > ph_t, \quad (15)$$

$$dp_{min}(t) = p_t - ph_{t-1}; ph_{t-2} > ph_{t-1} < ph_t. \quad (16)$$

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ МОДЕЛИРОВАНИЯ

Для решения задачи классификации изменения цены биржевого инструмента на три класса, а именно: «рост», «падения», «отсутствие значимого изменения» — воспользуемся тремя моделями классификации. Временной горизонт предсказания равен 60 секундам. Значимым будем считать изменение цены более чем на 70 ценовых пунктов. Так как для торговой системы время обработки данных критично, модель должна иметь простую архитектуру и давать прогноз быстро. В связи с этим для исследования была выбрана модель логистической регрессии [15], которая сравнивается с ее обобщениями — моделями искусственной нейронной сети с различной конфигурацией.

Архитектуры моделей:

Нейронная сеть с одним скрытым слоем:

- 1) входной слой, размерность (69,);
- 2) скрытый слой, размерность (64,), функция активации “elu” [16];
- 3) пропуск связей с вероятностью 50% (Dropout) [17];
- 4) выходной слой, размерность (3), функция активации “softmax” [11].

Нейронная сеть с двумя скрытыми слоями:

- 1) входной слой, размерность (69,);
- 2) скрытый слой, размерность (64,), функция активации “elu”;
- 3) пропуск связей с вероятностью 50%;
- 4) скрытый слой, размерность (32,), функция активации “elu”;
- 5) пропуск связей с вероятностью 50%;
- 6) выходной слой, размерность (3,), функция активации “softmax”.

Нейронная сеть с тремя скрытыми слоями:

- 1) входной слой, размерность (69,);
- 2) скрытый слой, размерность (64,), функция активации “elu”;
- 3) пропуск связей с вероятностью 50%;
- 4) скрытый слой, размерность (32,), функция активации “elu”;
- 5) пропуск связей с вероятностью 50%;
- 6) скрытый слой, размерность (16,), функция активации “elu”;

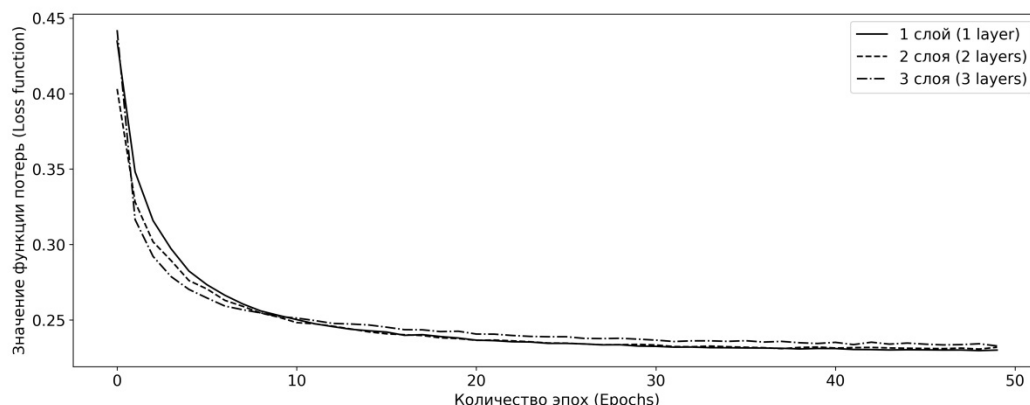


Рис. 1 / Fig. 1. Динамика функций потерь моделей нейронных сетей с различным количеством слоев / The dynamics of loss functions of neural network models with different numbers of layers

Источник / Source: составлено автором / compiled by the author.

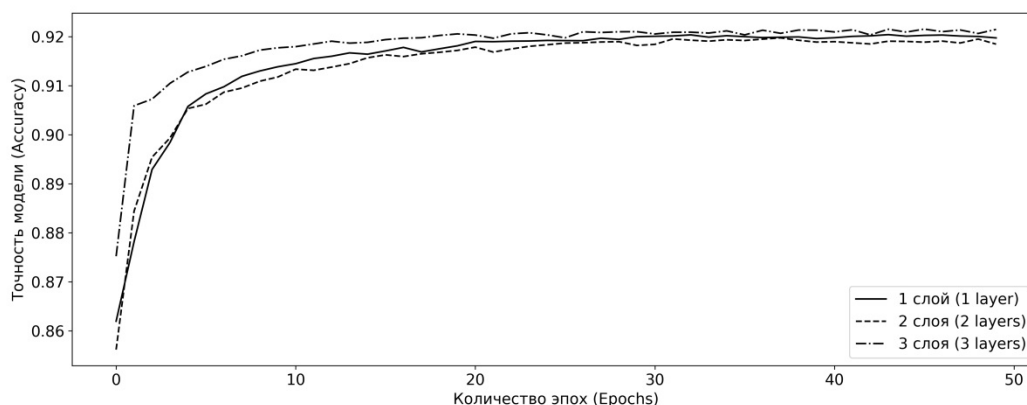


Рис. 2 / Fig. 2. Динамика точности моделей нейронных сетей с различным количеством слоев / The dynamics of accuracy of neural network models with different numbers of layers

Источник / Source: составлено автором / compiled by the author.

- 7) пропуск связей с вероятностью 50%;
- 8) выходной слой, размерность (3,), функция активации “softmax”.

Логистическая регрессия:

- 1) входной слой, размерность (69,);
- 2) выходной слой, размерность (3,), функция активации “softmax”.

Каждая сеть обучалась с использованием оптимизатора “adam” [18], минимизируя логарифмическую функцию потерь (“logistic loss”) [19] с размером подвыборки, равным 6000 наблюдений и количеством эпох, равным 100. Логистическая регрессия обучалась с помощью стохастического градиентного спуска. Весь набор наблюдений был разделен на обучающую и тестовую выборку по времени для оценки качества модели с помощью вневыборочного тестирования. Разделение по времени представлено на рис. 2. Размер обучающей выборки составляет 139 836 наблюдений (70% от

общего количества наблюдений), размер тестовой выборки равен 59 931 наблюдений (30% от общего количества наблюдений). Динамика значений функций потерь и точности моделей показана на рис. 3 и 4 соответственно. Результаты тестирования представлены в табл. 1. Динамика функций потерь и метрик точности моделей относительно количества эпох отражена на рис. 1, 2.

Увеличение глубины модели нейронной сети не позволяет увеличить ее качество, а только сокращает количество итераций для достаточного обучения модели. Кроме того, глубина сети увеличивает время ответа модели при предсказании. Таким образом, предпочтение отдается нейронной сети с одним скрытым слоем как наиболее эффективной из рассмотренных моделей для решения поставленной задачи. Результаты логистической регрессии не отличаются значительно для решения задачи предсказания изменений цены от результатов

Результаты тестирования моделей / Modeling results

	Значение функции потерь / Loss functions values	Точность / Accuracy
Нейронная сеть – 1 слой / Neural net – 1 Layer	0,2262	0,9242
Нейронная сеть – 1 слой / Neural net – 1 Layer	0,2277	0,9226
Нейронная сеть – 1 слой / Neural net – 1 Layer	0,2278	0,9237
Логистическая регрессия / Logistics Regression	0,2309	0,9235

Источник / Source: составлено автором / compiled by the author.

лучшей модели нейронной сети. При этом в отличие от модели нейронной сети модель логистической регрессии легко интерпретируема. В табл. 2 представлены 30 наиболее высоких коэффициентов при объясняющих переменных в логистической регрессии для классов «рост» и «падение».

Таблица с коэффициентами (см. табл. 2) позволяет интерпретировать поведение модели при текущем состоянии рынка. К примеру, для класса «рост» коэффициент при переменной «изменение цены за 10 секунд» положительный, а коэффициент при переменной «изменение цены за 60 секунд, усредненное за 10 прошлых периодов» отрицательный. Это означает что при текущем состоянии рынка после длительного падения цены и наличии краткосрочного роста модель склонна будет рекомендовать продолжение роста более чем на 70 ценовых пунктов как минимум в течение следующих 60 секунд.

ФОРМИРОВАНИЕ И ТЕСТИРОВАНИЕ АЛГОРИТМА ТОРГОВЛИ БИРЖЕВЫМ ИНСТРУМЕНТОМ НА БАЗЕ МОДЕЛИ ИЗМЕНЕНИЯ ЦЕНЫ

Основной задачей построенной модели является предсказание будущих изменений цены для принятия торговых решений. Таким образом, модель является основой для системы заключения сделок на бирже [20]. Заключение сделок на бирже с целью получения прибыли является рискованной деятельностью, поэтому нам важно, чтобы алгоритм при совершении сделки делал как можно меньше ошибок. Сделок может быть немного, но большинство из них должны быть прибыльными. Таким образом, при совершении сделок наибольшее значение играет метрика «точности»

(precision) модели — доля правильных ответов из всех тех наблюдений, которые она предсказала правильными. Напротив, меньшую значимость имеет метрика «полноты» модели (recall) — доля правильных ответов модели из всех правильных наблюдений. Данные метрики качества модели рассчитываются из матрицы ошибок классификатора (табл. 3) [21].

$$precision = \frac{TP}{TP + FP}, \quad (17)$$

$$recall = \frac{TP}{TP + FN}. \quad (18)$$

Модель логистической регрессии в качестве ответа отдает оценку вероятности в виде величины от 0 до 1, исходя из сигмоидальной функции от линейной комбинации признаковых показателей на оптимизированные веса.

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (19)$$

Соответственно, при принятии решения имеется возможность совершать сделки только в тех случаях, когда ответ модели имеет определенную точность. Таким образом, торговая система контролирует риск получения убытка при совершении сделок на бирже. Рассмотрим описанные метрики в зависимости от выбранного порога уверенности алгоритма. Будем считать, что

$$l(tr) = \begin{cases} 1, & \sigma(x) > tr \\ 0, & \sigma(x) \leq tr \end{cases}, \quad (20)$$

где l — сигнал модели для совершения сделки на покупку или продажу. Данная величина принима-

Наиболее высокие коэффициенты при объясняющих переменных в модели логистической регрессии / Highest ratios for explanatory variables in the logistic regression model

Признак / Feature	Коэффициент «Рост» / Coefficient "Increase"	Коэффициент «Падение» / Coefficient "Decrease"
Расстояние от скользящего среднего с периодом 5 минут до цены	1,610 498	-1,635 850
Изменение цены за 10 секунд	0,939 757	-0,913 961
Изменение цены за 60 секунд, усредненное за 10 прошлых периодов	-0,519 591	0,438 025
Количество сделок на продажу в секунду, усредненное за 60 прошлых периодов	0,511 877	-0,247 752
Суммарный объем сделок на продажу в секунду, усредненный за 60 прошлых периодов	-0,495 679	0,369 316
Изменение цены за 60 секунд, усредненное за 5 прошлых периодов	-0,376 149	0,360 490
Количество сделок на продажу в секунду, усредненное за 30 прошлых периодов	-0,352 787	0,191 294
Изменение цены за 10 секунд, усредненное за 10 прошлых периодов	-0,329 322	0,236 513
Изменение цены за 30 секунд	0,291 211	-0,243 507
Сумма объемов на продажу в биржевой книге заявок с глубиной равной 10, усредненная за 60 прошлых периодов	0,287 177	-0,114 863
Расстояние от цены до ближайшего локального максимума	0,275 646	-0,402 231
Количество сделок на покупку в секунду, усредненное за 10 прошлых периодов	0,262 265	-0,197 380
Расстояние до скользящей средней с периодом 10 минут	-0,252 081	0,180 693
Сумма объемов на продажу в биржевой книге заявок с глубиной, равной 20, усредненная за 60 прошлых периодов	-0,251 218	0,263 777
Суммарный объем сделок на покупку в секунду, усредненный за 30 прошлых периодов	0,250 966	-0,174 902

Источник / Source: составлено автором / compiled by the author.

ет значение 1 в качестве положительного сигнала и 0 — в качестве отрицательного сигнала; tr — величина порога уверенности алгоритма. Метрики качества моделей классификации для роста и падения цены представлены на рис. 3. Изменение размера выборки относительно порога уверенности модели показано на рис. 4.

К примеру, для уровня порога уверенности модели 0,9 матрицы ошибок для моделей роста и падения представлены в табл. 4 и 5 соответственно.

Таким образом, из общего количества положительных ответов равного 299 модель роста оказа-

лась правой 266 раз. Модель падения из общего количества положительных ответов 298 оказалась права 281 раз. Метрики качества моделей при том же уровне уверенности представлены в табл. 6.

Рассмотрим базовую торговую стратегию на базе построенной модели. Стратегия торгует заданным количеством инструмента и способна открывать как длинные позиции (покупать актив), так и короткие позиции (продавать актив). Торговля ведется одним торговым инструментом. Алгоритм покупает актив (открывает длинную позицию), если отсутствуют открытые позиции и уверенность модели на по-

Таблица 3 / Table 3

Схема матрицы ошибок / Error matrix diagram

		Ответы модели / Predictions	
		1	0
Истинные значения / True values	1	Правильные положительные ответы (TP) / Correct positive answers	Неправильные отрицательные ответы (FN) / Incorrect negative answers
	0	Неправильные положительные ответы (FP) / Incorrect positive answers	Правильные отрицательные ответы (TN) / Correct negative answers

Источник / Source: составлено автором / compiled by the author.

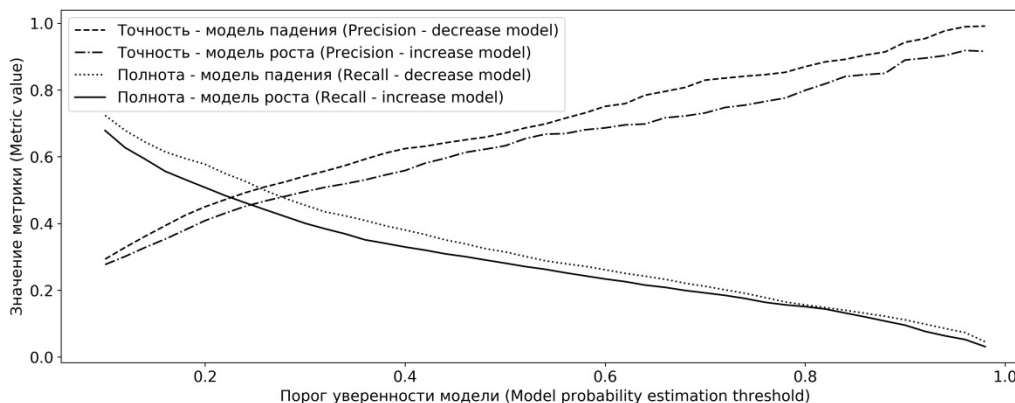


Рис. 3 / Fig. 3. Метрики качества моделей классификации в зависимости от порога уверенности модели / Quality metrics of classification models depending on the model confidence threshold

Источник / Source: составлено автором / compiled by the author.

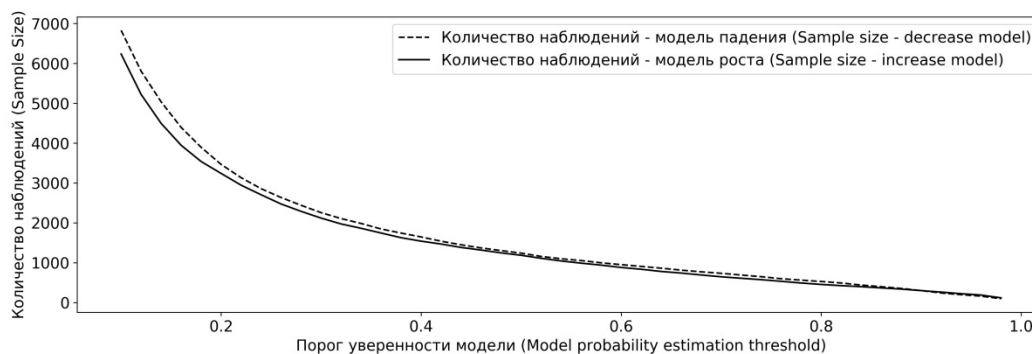


Рис. 4 / Fig 4. Изменение размера выборки относительно порога уверенности модели / Changing the sample size relative to the model confidence threshold

Источник / Source: составлено автором / compiled by the author.

купку выше заданного значения, т.е. когда сигнал модели на покупку положительный.

$$\begin{cases} l_b(tr_b) = 1 \\ p = 0 \end{cases} \quad (21)$$

Аналогично алгоритм продает актив (открывает короткую позицию), если отсутствуют открытые позиции и сигнал модели на продажу положительный.

$$\begin{cases} l_s(tr_s) = 1 \\ pos = 0 \end{cases}, \quad (22)$$

где l_b – сигнал модели на покупку; l_s – сигнал модели на продажу; p – позиция алгоритма. Позиция алгоритма p принимает три значения: $pos = 1$, если текущая позиция длинная; $pos = -1$, если текущая позиция короткая; $pos = 0$, отсутствует позиция.

Таблица 4 / Table 4

Матрица ошибок для модели роста / Error matrix for "Increase" model

		Ответы модели / Predictions	
		1	0
Истинные значения / True values	1	Правильные положительные ответы (TP) = 266 / Correct positive answers	Неправильные отрицательные ответы (FN) = 2523 / Incorrect negative answers
	0	Неправильные положительные ответы (FP) = 33 / Incorrect positive answers	Правильные отрицательные ответы (TN) = 57109 / Correct negative answers

Источник / Source: составлено автором / compiled by the author.

Таблица 5 / Table 5

Матрица ошибок для модели падения / Error matrix for "Decrease" model

		Ответы модели / Predictions	
		1	0
Истинные значения / True values	1	Правильные положительные ответы (TP) = 281 / Correct positive answers	Неправильные отрицательные ответы (FN) = 244 / Incorrect negative answers
	0	Неправильные положительные ответы (FP) = 17 / Incorrect positive answers	Правильные отрицательные ответы (TN) = 57389 / Correct negative answers

Источник / Source: составлено автором / compiled by the author.

Таблица 6 / Table 6

Метрики качества моделей / Model quality metrics

	Модель роста / "Increase" model	Модель падения / "Decrease" model
Точность / Precision	0,8896	0,9430
Полнота / Recall	0,0954	0,1113

Источник / Source: составлено автором / compiled by the author.

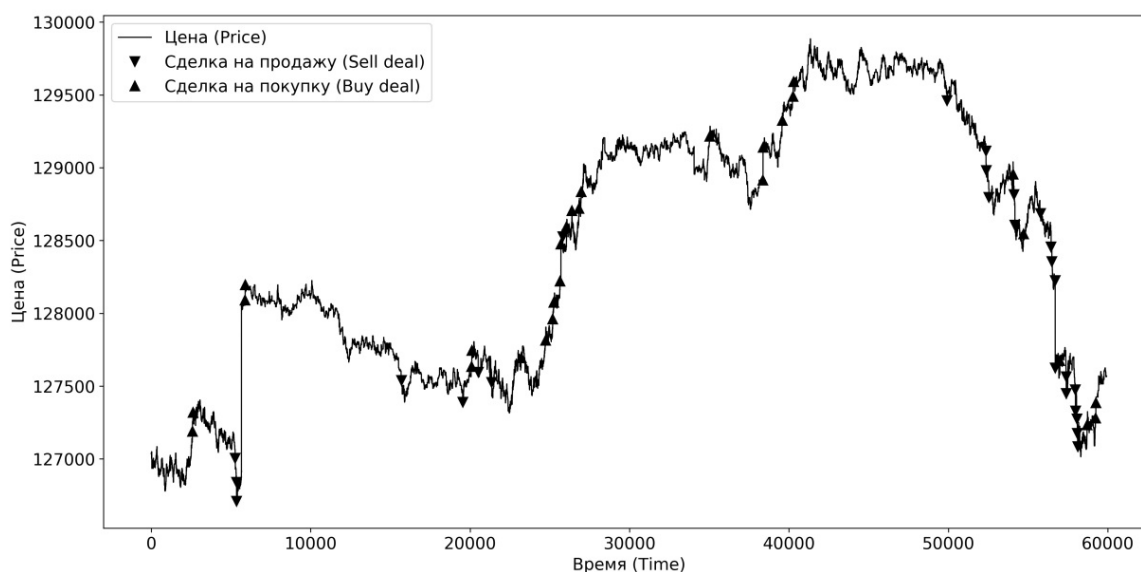


Рис. 5 / Fig. 5. Сделки стратегии при тестировании на исторических данных / Strategy deals when testing on historical data

Источник / Source: составлено автором / compiled by the author.

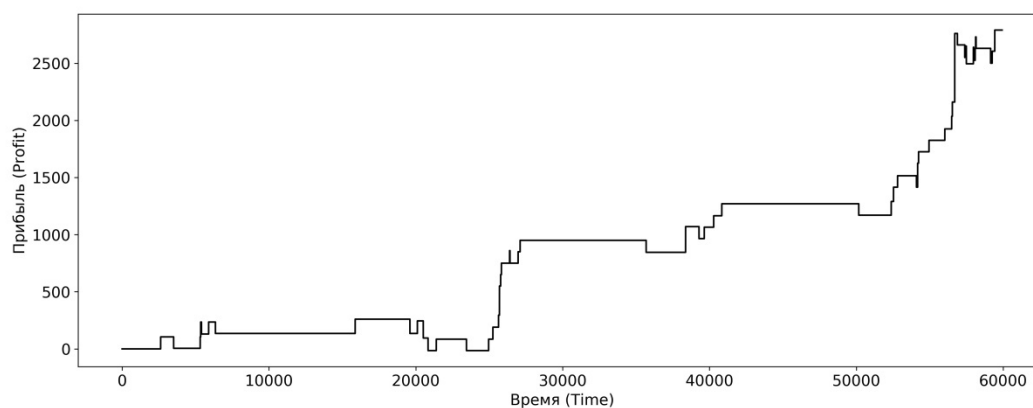


Рис. 6 / Fig. 6. Динамика прибыли стратегии при тестировании стратегии на исторических данных / Dynamics of profit strategy when testing strategies on historical data

Источник / Source: составлено автором / compiled by the author.

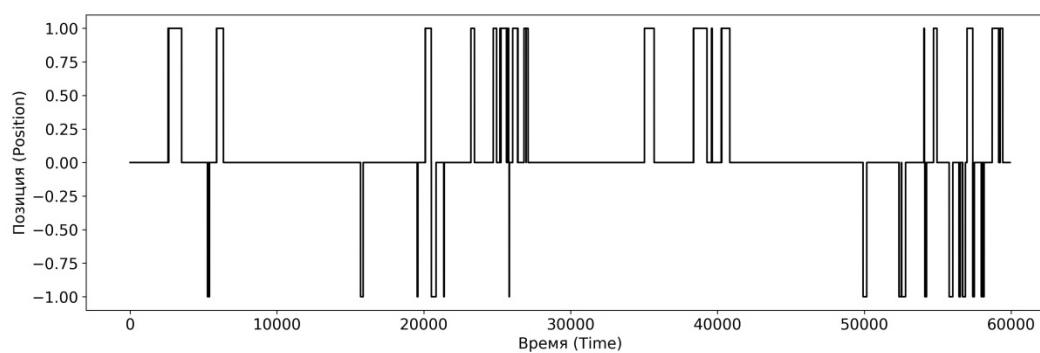


Рис. 7 / Fig. 7. Динамика изменения позиции стратегии при тестировании на исторических данных / Dynamics of change in strategy position when testing on historical data

Источник / Source: составлено автором / compiled by the author.

Закрывание позиций осуществляется при достижении алгоритма заданного уровня прибыли или убытка. Условие закрытия длинной позиции:

$$\begin{cases} p_t - p_b > tp \\ p_b - p_t > sl \\ pos = 1. \end{cases} \quad (23)$$

Условие закрытия короткой позиции:

$$\begin{cases} p_s - p_t > tp \\ p_t - p_s > sl, \\ pos = -1 \end{cases} \quad (24)$$

где p_b — цена покупки актива во время открытия длинной позиции; p_s — цена продажи актива во время открытия короткой позиции; sl — заданный уровень убытка в сделке; tp — заданный уровень прибыли в сделке.

Рассмотрим результаты тестирования вышеописанной торговой стратегии на исторических данных. Тестирование проводилось на сформированной на этапе построения модели, тестовой выборке. Зафиксированы следующие параметры стратегии:

- порог уверенности модели на покупку $tr_b = 0,9$;
- порог уверенности модели на продажу $tr_s = 0,9$;
- уровень убытка в сделке $sl = 100$;
- уровень прибыли в сделке $tp = 100$.

Динамика показателей стратегии при тестировании на исторических данных показана на рис. 5–7. Отражены сигналы модели на продажу и на покупку, время и цены открытия длинных и коротких позиций, динамика прибыли стратегии, динамика позиции стратегии.

Статистика торговой стратегии при тестировании на исторических данных приведена в табл. 7.

Таким образом, модель, построенная на базе микроструктурных рыночных данных с использованием методов машинного обучения, позволила

Показатели качества торговой стратегии / Quality indicators of a trading strategy

Показатель / Indicator	Значение / Value
Прибыль / Profit	2790
Количество прибыльных сделок / The number of profitable trades	46
Количество убыточных сделок / The number of losing trades	8
Количество сделок на покупку / The number of transactions for the purchase	28
Количество сделок на продажу / Number of sales transactions	26

Источник / Source: составлено автором / compiled by the author.

создать алгоритм принятия решений, реализующих прибыльную торговую стратегию.

ЗАКЛЮЧЕНИЕ

В ходе исследования удалось разработать торговую систему, основанную на модели машинного обучения и микроструктурных рыночных данных. Для реализации данной системы был проведен сравнительный анализ ряда моделей машинного обучения, а именно: нейронных сетей и логистической регрессии. В ходе экспериментов моделирования наилучшие показатели качества продемонстрировала модель логистической регрессии. В качестве признаков для моделей использовались различные агрегаты биржевой книги заявок и рыночных сделок, а также классические индикаторы технического анализа. На базе построенной модели машинного обучения

разработан торговый алгоритм, показывающий стабильный положительный результат при тестировании на исторических данных. Описанные методы и алгоритмы могут применяться различными инвестиционными институтами для эффективного управления капиталом в биржевых торгах.

Видится несколько направлений дальнейших исследований. Во-первых, использование большего количества данных для обучения модели позволит повысить качество классификации. Во-вторых, возможно применение оптимизации параметров моделей, а также использование принципиально других архитектур нейронных сетей. Помимо этого, разработка более сложных и детальных торговых алгоритмов на базе модели машинного обучения позволит увеличить конечную эффективность всей торговой системы.

REFERENCES

1. Cartea A., Penalva J. Where is the value in high frequency trading? *Quarterly Journal of Finance*. 2012;2(3). DOI: 10.1142/S 2010139212500140
2. Brunzell T. High-frequency trading — to regulate or not to regulate — that is the question? Does scientific data offer an answer? *Journal of Business and Financial Affairs*. 2013;2(1):1–4. DOI: 10.4172/2167–0234.1000e121
3. O'Hara M. High frequency market microstructure. *Journal of Financial Economics*. 2015;116(2):257–270. DOI: 10.1016/j.jfineco.2015.01.003
4. Hoffmann P. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics*. 2014;113(1):156–169. DOI: 10.1016/j.jfineco.2014.04.002
5. Kercheval A.N., Zhang Y. Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance*. 2015;15(8):1315–1329. DOI: 10.1080/14697688.2015.1032546

6. Anderson E., Merolla P., Pribula A. Adaptive strategies for high frequency trading. *Journal of Stellar MS&E*. 2008. URL: <https://web.stanford.edu/class/msande444/2008/OrderBook.pdf> (accessed 28.08.2018).
7. Kearns M., Nevmyvaka Yu. Machine learning for market microstructure and high frequency trading. In: High frequency trading: New realities for traders, markets and regulators. London: Risk Books; 2013:91–124. URL: <https://www.cis.upenn.edu/~mkearns/papers/KearnsNevmyvakaHFTRiskBooks.pdf> (accessed 28.08.2018).
8. Huang H., Kercheval A. N. A generalized birth-death stochastic model for high-frequency order book dynamics. *Quantitative Finance*. 2012;12(4):547–557. DOI: 10.1080/14697688.2012.664926
9. Goldstein M.A., Kumar P.K., Graves F.C. Computerized and high-frequency trading. *The Financial Review*. 2014;49(2):177–202. DOI: 10.1111/fire.12031
10. Angel J., L. Harris L., Spatt C.S. Equity trading in the 21st century. *Quarterly Journal of Finance*. 2011;1(1):1–53. DOI: 10.1142/S 2010139211000067
11. Bishop C.M. Pattern recognition and machine learning. New York: Springer-Verlag; 2006. 738 p.
12. Patel J., Shah S., Thakkar P., Kotecha K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*. 2015;42(1):259–268. DOI: 10.1016/j.eswa.2014.07.040
13. Tsantekidis A., Passalis N., Tefas A., Kannianen J., Gabbouj M., Iosifidis A. Forecasting stock prices from the limit order book using convolutional neural networks. In: IEEE 19th Conf. on business informatics (CBI) (Thessaloniki, Greece, 24–27 July 2017). New York: IEEE; 2017. DOI: 10.1109/CBI.2017.23
14. Achelis S.B. Technical analysis from A to Z: Covers every trading tool... from the Absolute Breadth Index to the Zig Zag. New York: McGraw-Hill Co.; 2000. 380 p.
15. Cox D.R. The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society: Series B*. 1958;20(2):215–242. URL: <https://www.nuffield.ox.ac.uk/users/cox/cox48.pdf> (accessed 28.08.2018).
16. Clevert D.-A., Unterthiner T., Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: 4th Int. conf. on learning representations (ICLR 2016) (San Juan, Puerto Rico, 2–4 May, 2016). URL: <https://arxiv.org/pdf/1511.07289.pdf>
17. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 2014;15(1):1929–1958. URL: <http://jmlr.org/papers/volume15/srivastava14a.old/srivastava14a.pdf> (accessed 28.08.2018).
18. Kingma D.P., Ba J. Adam: A method for stochastic optimization. In: 3rd Int. conf. on learning representations (ICLR 2015) (San Diego, 7–9 May, 2015). URL: <https://arxiv.org/pdf/1412.6980.pdf> (accessed 28.08.2018).
19. Rosasco L., De Vito E. D., Caponnetto A., Piana M., Verri A. Are loss functions all the same? *Neural Computation*. 2004;16(5):1063–1076. DOI: 10.1162/089976604773135104
20. Li X., Deng X., Zhu S., Wang F., Xie H. An intelligent market making strategy in algorithmic trading. *Frontiers of Computer Science*. 2014;8(4):596–608. DOI: 10.1007/s11704-014-3312-6
21. Powers D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*. 2011;2(1):37–63. URL: https://bioinfopublication.org/files/articles/2_1_1_JMLT.pdf (accessed 28.08.2018).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Николай Александрович Билев — аспирант кафедры математических методов анализа экономики, экономический факультет, Московский государственный университет им. М. В. Ломоносова, Москва, Россия
bilevn@gmail.com

ABOUT THE AUTHOR

Nikolai A. Bilev — post-graduate student of the Department of Mathematical Methods of Economic Analysis, Faculty of Economics, Lomonosov Moscow State University, Moscow, Russia
bilevn@gmail.com

Статья поступила 10.06.2018; принята к публикации 08.10.2018.

Автор прочитал и одобрил окончательный вариант рукописи.

The article was received on 10.06.2018; accepted for publication on 08.10.2018.

The author read and approved the final version of the manuscript.