

М. О. Паньков

старший преподаватель кафедры статистики и эконометрики Санкт-Петербургского государственного экономического университета

ПРОГНОЗИРОВАНИЕ ЦЕН ЗАКРЫТИЯ АКЦИЙ APPLE С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Введение

Прогнозирование цен на акции является одной из ведущих тем в финансовых исследованиях. Целью данной статьи является прогнозирование цен закрытия акций компании Apple с помощью искусственной нейронной сети. Первоначальной задачей является изучение современных исследований по данной тематике. Генеральной задачей исследования является интеграция метода нейронных сетей в область прогнозирования финансовых рынков. Важными задачами работы являются выбор метода стандартизации данных, функции активации, оценки ошибки, а также подходящего алгоритма оптимизации и подбор оптимальной сети, требующий измерения времени обучения сети, а также анализа изменения ошибки в процессе обучения с целью выявления эффекта переобучения.

Успешное прогнозирование цен на акции может принести значительную прибыль в торговле. Технические трейдеры основывают свой анализ на предположении, что модели рыночных цен будут повторяться в будущем, и, таким образом, эти модели могут быть использованы в целях прогнозирования. Так что наиболее распространенным и хорошо разработанным методом, используемым для выполнения прогнозов, является анализ временных рядов.

Краткий обзор исследований

Порой и в научных статьях можно встретить простую манипуляцию терминами, такими как нейронные сети, за которыми ничего, по сути, нет. Р. Х. Ильясовым и Д. А. Куразовой для прогнозирования промышленных индексов Московской биржи использованы методы корреляционно-регрессионного анализа, а также приводятся результаты применения искусственной нейронной сети (Ильясов, Куразова, 2018). В качестве основной модели авторами выбрана модель линейного тренда. Поскольку авторы имеют дело с временными рядами, то, казалось бы, их внимание должно быть сосредоточено на проверке стационарности рядов, отсутствия гетероскедастичности, а также проверке наличия автокорреляции остатков. О мультиколлинеарности речь не идет, поскольку в данной работе использовались только модели с одной объясняющей переменной. Так как ни одна из названных процедур анализа временных рядов не была проведена, возникает вопрос прежде всего о качестве прогнозов и их целесообразности. Авторами выдвигается гипотеза о том, что прогноз индекса промышленного производства можно получить с помощью модели персептрона (Minsky, Papert, 2017). Утверждается целесообразность использования одного скрытого слоя, в то время как точное число скрытых слоев не устанавливается и чаще всего определяется эмпирически с помощью ошибки

прогноза. Точность прогнозов основывается на собственном предположении авторов и никак не подтверждается оценкой качества предсказанных значений. Авторы делают заключение о подтверждении выдвинутой гипотезы о возможности построения прогноза с помощью искусственной нейронной сети, без какой-либо оценки.

Довольно интересный вызов проблемам прогнозирования был сделан Себастьяном Хайнце (Heinze, 2017). В качестве объекта исследования был принят индекс S&P 500, а также его 500 компонентов. Рабочим инструментом исследования выступали язык программирования Python и библиотека для глубокого обучения нейронных сетей Tensor Flow. В качестве единицы исследования были приняты минутные торги за период с апреля по август 2017 г. В результате база данных составила 41 266 наблюдений. Этот набор данных был разбит на тренировочную и тестовую выборки. Разбивка осуществлялась путем отсечения первых 80% наблюдений в качестве тренировочных и 20% последних в качестве тестовых. Многие исследователи предлагают в качестве основного способа формирования тренировочного и тестового набора данных случайный отбор наблюдений, что отвечает включению принципа случайности при построении прогнозов по искусственным нейронным сетям. Такой способ помогает исключить содержание тренда и сезонности при прогнозировании. В работе Хайнце предлагается шкалирование исходных данных методом отклонений от максимальных — минимальных значений. Такой подход нацелен на использование функций активации, область значений которых лежит в пределах $[-1; 1]$ для тангенса или $[0; 1]$ для функции сигмоида. Автор предлагает соблюдать осторожность в отношении того, какая часть данных масштабируется, отмечая, что распространенной ошибкой является масштабирование всего набора данных перед применением методов обучения и разделения тестовой части. Масштабирование сказывается на вычислении статистик, например, мин/макс значений переменной. При выполнении прогнозирования временных рядов в реальной жизни у исследователя нет информации о будущих наблюдениях. Поэтому переход к масштабированию должен проводиться на обучающемся наборе данных, а затем должен применяться к тестовой части совокупности. В противном случае будет использована будущая информация, относящаяся к периоду прогнозирования, которая обычно смещает прогнозные показатели в позитивном направлении. Нейронная сеть рассматриваемого исследования включает четыре скрытых слоя. Первый слой содержит 1024 нейрона, что более чем вдвое превышает размер входов. Последующие скрытые слои всегда в два раза меньше предыдущего и составляют 512, 256 и, наконец, 128 нейронов. Уменьшение количества нейронов для каждого последующего слоя сжимает информацию, которую сеть идентифицирует на предыдущих уровнях. Принцип уменьшения размерности с каждым последующим слоем составляет суть подхода к построению архитектуры искусственной нейронной сети глубокого обучения. Скрытые слои сети преобразуются с помощью функций активации, которые являются важными элементами сетевой архитектуры, поскольку они вносят нелинейность в систему. В качестве функций активации в рассматриваемой работе использована выпрямленная линейная единица (ReLU). Результаты исследования показали после 10 эпох итоговый тест MSE, равный 0,00078. Средняя абсолютная процентная ошибка прогноза на тестовом наборе составила 5,31%, так что результаты исследования С. Хайнце возможно оценить как весьма положительные.

Оценка курса акций или оценка фондового индекса интересны тем, что для их описания можно привлечь множество факторов. По сути, выдвинутые гипотезы о возможном воздействии того или иного фактора могут ограничиваться только фантазией исследователя.

Вивек Паланиаппан в своей работе (Palaniappan, 2018) уделяет особое внимание методам машинного обучения для получения прогнозов. Из-за сложности динамики фондового рынка данные о ценах на акции часто заполнены шумом, который может отвлечь машинное обучение от изучения тренда и структуры. По этой причине автор предлагает убрать часть шума, сохранив тенденции и структуру данных. Для решения этой задачи он применяет Вейвлет-преобразование. Вейвлет-преобразование тесно связано с преобразованием Фурье, которое автор предлагает в качестве альтернативы для анализа случайных колебаний. В работе было принято решение использовать автокодировщики из-за интерпретируемости кодирования по сравнению с вероятностями ограниченных машин Больцмана. Сложенные автокодировщики хорошо сжимают данные и воспроизводят их снова. Автора интересует исключительно часть сжатия, так как она представляет информацию, необходимую для воспроизведения данных, каким-то образом закодированную в сжатой форме. Это говорит о том, что сжатые данные могут в некотором роде быть характеристиками данных, из которых извлекаются функции.

Входные данные сжимаются в любое желаемое количество нейронов и сеть вынуждена перестраивать исходные данные с помощью автокодировщика. Это вынуждает модель извлекать ключевые элементы данных, которые можно интерпретировать как функции. Важным моментом, который отмечает В. Паланиаппан, является то, что эта модель на самом деле попадает под неконтролируемое обучение, так как нет входных и выходных пар, но вход и выход одинаковы.

Этот подход отличается от рассмотренной статьи С. Хайнце большей сложностью, но полученные результаты прогнозов весьма схожи.

Как уже отмечалось, методы временных рядов являются основополагающими в анализе данных котировок финансовых рынков. В работе Селен Ю Ху (Ху, 2014) предпринята попытка объединить традиционную технику анализа временных рядов с информацией на веб-сайте тренда Google и финансовом веб-сайте Yahoo для прогнозирования еженедельных изменений в цене акций. В качестве рабочего инструмента автором выбран язык программирования R. Суть работы состоит в подборе лучшей модели для прогноза, причем первоначальный анализ обнаружил значительную корреляцию между новостями и еженедельными изменениями цены акций. К такого рода проявлениям нужно относиться осторожно, поскольку вполне вероятно, что может доминировать ряд значимых наблюдений, которые не отражают общую тенденцию. В целом, еженедельные изменения цены акций в разные интервалы значений новостей ведут себя так, как ожидает автор. Однако присутствуют и некоторые исключения, на которые стоит обратить внимание при составлении прогнозов. В данной работе есть ряд ограничений и возможности улучшения в методах сбора и анализа данных. Так, в работе использована экспоненциальная функция, чтобы имитировать, как быстро исчезает влияние новостей. Данный процесс может быть имитирован с помощью других функций, обеспечивающих лучший результат. Работа интересна тем, что автор попытался включить в исследование и фундаментальный, и технический анализ. Технический анализ раскрыт в статье фрагментарно, так как производились сравнения двух моделей, таких как ARMA/ARIMA и регрессионной модели с фиктивными переменными.

Основной теорией, лежащей в основе прогнозирования цены акций, является гипотеза эффективного рынка (ЕМН), которая утверждает, что цена акций отражает всю доступную информацию, и каждый индивид имеет определенный уровень доступа к ней. Суть ЕМН заключается в предположении, что рынок мгновенно реагирует на новости, и никто не может превзойти рынок в долгосрочной перспективе. Анализ временных рядов охватывает большое количество методов прогнозирования.

Исследователями разработаны многочисленные модификации базовой модели ARIMA, и достигнут значительный успех в этих методах. Модификации такого подхода включают кластеризацию временных рядов из моделей ARMA с урезанными данными. Также интересен метод нечеткой нейронной сети (Marcek, 2004). Имеются примеры использования для прогнозов методов опорных векторов (Pai, Lin, 2005). Почти все эти исследования показывают, что дополнительные факторы должны быть приняты во внимание для расширения базовой или так называемой «неизменной модели». Наиболее распространенным и важным из таких факторов является онлайн новостная информация, связанная с акциями. Многие исследователи пытаются использовать текстовую информацию в публичных СМИ для оценки новостей. Для выполнения этой задачи разрабатываются различные механизмы, которые не затрагиваются в данной работе, так как требуют отдельного и детального исследования. Исследования всех этих процессов основаны на сложных алгоритмах, которые способны выполнять извлечение и оценку текста из онлайн-источников.

В работе Ина Йибин основной задачей был поиск наилучшего способа прогнозирования котировки акций на основе исторических данных (Yibin, 2019). Выбор лучшей модели проводился на основе средней ошибки аппроксимации (MAPE) и стандартного отклонения остатков (RMSE). В качестве основных моделей были использованы методы скользящих средних, линейная регрессия, библиотека градиентного бустинга (XGBoost), метод долгой краткосрочной памяти (LSTM). По результатам исследований автором сделан вывод, что лучший результат прогноза показала линейная регрессия. Этот вывод вряд ли можно считать вполне обоснованным и достоверным, так как при анализе регрессии не выполнена оценка гетероскедастичности и автокорреляции, которые являются частой компонентой, влияющей на построение моделей котировок акций.

В работе Ю. Сонг, Ю. Жу, Р. Хан также рассматривается задача лучшего прогнозирования котировки цены акции. Задача решается в рамках прогнозирования финансовых временных рядов, поскольку фондовый рынок по своей природе является динамичным, нелинейным, сложным, непараметрическим и хаотичным (Song, Zhou, Han, 2018). Кроме того, многие макроэкономические условия, такие как политические события, политика компании, общие экономические условия, индекс цен на сырьевые товары, процентные ставки, ожидания инвесторов, решения институциональных инвесторов и психологические факторы, также рассматриваются в качестве влияющих факторов. В работе представлены пять моделей искусственного интеллекта (ИИ) для целей прогнозирования. Среди моделей ИИ нейронные сети с обратным распространением (BPNN), нейронные сети с радиальной базой (RBFNN), нейронная сеть общей регрессии (GRNN), регрессия с опорным вектором (SVMR), регрессия с опорным вектором с наименьшими квадратами (LS-SVMR) являются наиболее широко используемыми и хорошо разработанными методами. Авторами изучаются еженедельные скорректированные цены закрытия трех акций: China Bank, Vanke A и Kweichou Moutai. Каждый набор данных о ценах акций содержит 427 наблюдений, начиная с 3 января 2006 г. по 11 марта 2018 г. Весь набор данных авторы разбивают на тренировочный набор (80%) и тестовый набор (20%). По результатам исследования лучшие оценки прогнозов для всех наборов данных показала сеть обратного распространения ошибки. Кроме того, авторами проанализирована сеть SVRM с различными функциями активации. Были рассмотрены линейная, полиномиальная, сигмоидная функции, а также сеть радиально-базисных функций (RBF). Линейное ядро является лучшим в этой задаче прогнозирования и последовательно превосходит три других ядра. Стоит также учесть, что ядро RBF является ядром по умолчанию во многих библиотеках из-за его гибкости в отношении

различных ресурсов данных, хотя в данной задаче полученные результаты не вполне качественны. Стоило бы попробовать другие функции активации, чтобы сделать сравнение при выполнении проектов прогнозирования. Хотя сеть обратного распространения ошибки показала лучшие результаты, чем SVMR с линейным ядром, у них оказались схожие оценки ошибок прогнозирования, возможно, благодаря тому, что обе нейронные сети имеют средневзвешенное значение, которое фиксирует некоторую линейную зависимость в сети.

В работе А. Эллиота и С. Х. Хсу представлены подходы для прогнозирования котировок акций индекса S&P 500 (Elliot, Hsu, 2017). Приведено сравнение четырех основных моделей прогнозирования: линейная регрессия, обобщенная линейная модель, рекуррентная нейронная сеть (в качестве которой представлена long short term model — LSTM) и алгоритм онлайн-проектирования. Выделены сильные и слабые стороны каждой модели. Авторы подчеркивают важность стационарности исходных данных. Для линейных моделей регрессии это условие обязательно; для обобщенной модели условие стационарности также играет роль, но в меньшей мере оказывает отрицательное влияние на прогнозы. В эмпирическом тестировании модель RNN работает лучше по сравнению с тремя другими моделями, потому что рекуррентная нейронная сеть мгновенно обновляет ввод через LSTM. Авторы выдвигают гипотезу о том, что любая последовательность значений, которая соответствует котировкам финансового рынка, представляет собой так называемый мартингейл, где лучшим прогнозом для следующего значения является тот, который учитывает все состояния до сих пор, т. е. это то, где мы сейчас находимся. Используя алгоритм ОТВ, авторы преобразовывали оригинальные модели в динамические, которые могут обновлять информацию каждый раз, когда появляются новые данные. Тем самым появляется возможность продолжать корректировать веса экспертов в соответствии с самым последним исходом. При тестировании авторы обнаружили, что линейная модель никогда не показывает лучший результат прогнозов и работает хуже при увеличении «памяти» модели (включая количество точек задержки), т. е. включение краткосрочных периодических трендов не улучшило производительность модели. В качестве метрики ошибки использовался показатель RMSE.

Для построения GLM (Обобщенная линейная модель) модели авторы использовали оценку максимального правдоподобия в качестве весов прогноза, в качестве функции использовали лог — линейное нормальное распределение. Результаты тестирования показали ошибку ниже, чем для линейной модели, тем не менее ошибка оказалась существенной.

В рассмотренном исследовании была предложена рекуррентная нейронная сеть с кратковременной памятью (LSTM-RNN), которая позволяет сделать еще один шаг к созданию архитектуры памяти в отдельных узлах. Как известно, рекуррентные нейронные сети (RNN) — это нейронные сети, которые имеют некоторый способ запоминания предыдущих значений, введенных в них, и вывода на основе как ввода, так и сохраненного в памяти значения. Что делает RNN особенно удачной при прогнозировании данных временных рядов, так это то, что сеть позволяет выражать долгосрочные зависимости в выходных данных, но при этом не требуется сложной архитектуры, позволяющей вводить переменные данные. Кроме того, если во временном ряду существуют долгосрочные зависимости, то нейронная сеть с учетом достаточного количества узлов может теоретически их моделировать. Это менее реалистично на практике, поскольку нейронная сеть построена так, чтобы следовать градиентному спуску функции потерь. Если более простая модель — это локальные минимумы, то вполне вероятно, что модель «застрянет» в этих минимумах. Кроме того, если RNN построена слишком сложным способом, она может

найти очень глубокие минимумы в своих тренировочных данных, но только потому, что она подвержена переобучению. Наконец, RNN предполагает стационарность, так как после окончания обучения веса сети фиксируются.

Размер набора данных в статье составил 4000 наблюдений. При тестировании и сравнении LSTM-RNN с моделью Мартингейла авторами было обнаружено, что в целом LSTM-RNN либо показывал результаты, аналогичные модели Мартингейла, либо хуже. По сути, решение для точного прогнозирования показателей финансового рынка не было найдено. Возможно, это связано с недостаточным обзором всех возможностей моделирования, а также пренебрежением статистическими тестами, которые помогли бы определить, каким образом стоит преобразовать данные.

В работе М. Горбани и Е. К. Ченг предлагают использовать анализ главных компонент (PCA) в качестве основного метода для прогнозирования будущих значений котировок акций (Ghorbani, Chong, 2018). В рассмотренной статье представлено описание общего метода прогнозирования цены акций с использованием ковариаций в случаях уменьшения размерности с помощью анализа главных компонент. Прогнозирование цен на акции основано на исторических данных о ценах с использованием ковариационной информации. Как известно, ковариационная матрица набора данных хорошо аппроксимируется классической оценкой максимального правдоподобия или эмпирической ковариацией, если число наблюдений достаточно велико по сравнению с числом предикторов. Проблема возникает, когда размерность матрицы превышает количество наблюдений. В этом случае полученные оценки не являются надежными. Численные оценки в рассматриваемой работе основаны на достаточном наборе исходных данных для получения надежной эмпирической ковариационной матрицы. Продемонстрированный авторами подход имеет сходство с методами подпространственной фильтрации. Такие методы основаны на ортогональном разложении пространства данных с шумом на подпространство сигналов и подпространство шума. Такое разложение возможно в предположении низкоранговой модели для данных и при наличии оценки ковариационной матрицы шума. Эта задача может быть выполнена на основе модифицированного разложения по сингулярным значениям (SVD) матриц данных. Ортогональное разложение на подпространство сигнала и шума может выполняться разложением единичных значений или, аналогично, разложением по собственным значениям ковариационной матрицы с зашумленным сигналом. Авторы проиллюстрировали подход на ежедневных значениях цены акций для General Electric и пяти других компаний. Дневные исторические данные о ценах были преобразованы в Ханкель-матрицы 15 различных размерностей, чтобы исследовать влияние продолжительности наблюдения на мощность оценок. Результаты исследования показали лучшую производительность предложенного метода по сравнению с многомерным условным средним.

База данных

В нашем исследовании предпринята попытка прогнозирования цены закрытия акций компании Apple с учетом таких факторов, как цена открытия торгов акций, максимальная, минимальная цена за торговую сессию, а также объем торгов. Цены закрытия акций компании Apple представлены на рис. 1 в период с 2012 по 2017 г. Помимо информации о текущем активе в модель были включены данные о 28 крупнейших компаниях, акции которых представлены на Нью-Йоркской фондовой бирже, а также бирже NASDAQ. Все активы представлены через аббревиатуры, в соответствии с тикерами ранее названной биржи: AAPL, AXP, HD, CVX, MCD, MSFT, KO, XOM, GS, CAT, JNJ, MRK, CSCO, UNH, TRV, PG, WMT, JPM, PFE, VZ, V, UTX, MMM, WBA, IBM, DIS, DWDP, BA, INTC.

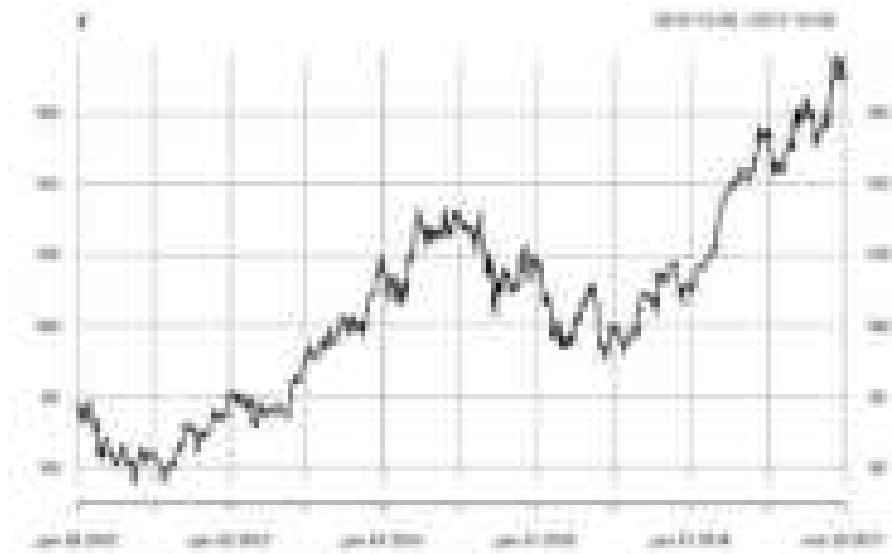


Рис. 1. Цены закрытия акций Apple в долларах США за период с 6 декабря 2012 г. по 5 декабря 2017 г.

Исходная база для анализа представляет собой матрицу размерностью 1259×174 , где в качестве строк выступают торговые сессии за пять лет, а в качестве столбцов — переменные с информацией о ценах закрытия, максимальной, минимальной, объема торгов, а также скорректированных ценах закрытия на выплаченные дивиденды. Для исключения мультиколлинеарности из вышеназванной базы данных пришлось убрать скорректированную цену закрытия по всем акциям, так как эта переменная фактически дублирует переменную цены закрытия.

Для качественной оценки прогнозов исходные данные были разбиты на тренировочную и тестовую выборки в пропорции 80 на 20%, соответственно. Оценки точности обученной нейронной сети по тренировочной выборке не имеют смысла, в то время как оценки точности по тестовой выборке являются прогнозными оценками (Ciaburro, Venkateswaran, 2017). Подготовленные данные для обучения и тестирования нейронных сетей представляют собой:

- тренировочную матрицу факторов размерностью 1007×144 ,
- тренировочный вектор результатов, в данном исследовании — это цены закрытия акций компании Apple размерностью 1007×1 ,
- тестовую матрицу факторов размерностью 252×144 ,
- тестовый вектор результатов 252×1 .

Подбор параметров сети

Исходный массив данных был представлен переменными разной размерности. Для оптимизации обучения искусственных нейронных сетей была применена функция стандартизации *z-scores*:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j},$$

где x_{ij} — значение i -го наблюдения j -й переменной, \bar{x}_j — средняя величина переменной, σ_j — стандартное отклонение переменной.

В исследовании в качестве функции активации была использована Rectified Linear Unit (ReLU). Данная функция активации очищает ряд значений от отрицательных, а для положительных принимает линейные значения:

$$z = b + \sum xw_i$$

$$y = \begin{cases} z, & \text{если } z > 0 \\ 0, & \text{если } z \leq 0 \end{cases}$$

Для оценки функции потерь (loss function) была использована среднеквадратичная ошибка (MSE). Данная метрика считается подходящей для измерения эффективности инструментов прогнозирования, которая представляет собой среднеквадратическую разницу между фактическим и прогнозируемым значением цены (Chollet, 2015). Учитывая исторические данные, одним из эффективных способов оценки будущих цен является простое использование многомерного условного среднего в качестве точечной оценки, поскольку оно сводит к минимуму среднеквадратичную ошибку оценки.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2,$$

где Y_i — целевое значение, Y'_i — прогнозное значение сети.

В качестве оценивания точности прогнозов применялась средняя абсолютная ошибка.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y'_i|.$$

В качестве оптимизатора искусственных нейронных сетей был использован алгоритм RMSprop. RMSprop — это неопубликованный алгоритм оптимизации, разработанный для нейронных сетей, впервые предложенный Джеффом Хинтоном в своих лекциях (Hinton, 2014). Оценка оптимизатора может быть получена как:

$$M(g^2)_t = \beta M(g^2)_{t-1} + (1-\beta) \left(\frac{\partial E}{\partial w} \right)^2.$$

Тогда обновление весов происходит следующим образом:

$$w_t = w_{t-1} \frac{\gamma}{\sqrt{M(g^2)_t}} \left(\frac{\partial E}{\partial w} \right),$$

где $M(g^2)_t$ — скользящее среднее квадратов градиентов, $\frac{\partial E}{\partial w}$ — градиент функции потерь относительно весов, γ — коэффициент скорости обучения, β — параметр скользящей средней (по умолчанию — 0,9).

Результаты исследования

В нашей работе рассматриваются модели однослойной нейронной сети и нейронные сети глубокого обучения (Keras Neural Network Sequential Models или NNSM). Обучение и архитектура искусственных нейронных сетей проводились с помощью библиотеки глубокого обучения Keras для языка программирования R. В табл. 1 приведены краткие характеристики предложенных искусственных нейронных сетей.

Таблица 1

Характеристики искусственных нейронных сетей

Однослойная нейронная сеть		
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 64)	9280
dense_2 (Dense)	(None, 1)	65
Total params: 9,345		
Trainable params: 9,345		
Non-trainable params: 0		
Сеть глубокого обучения с двумя скрытыми слоями		
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 64)	9280
dense_2 (Dense)	(None, 64)	4160
dense_3 (Dense)	(None, 1)	65
Total params: 13,505		
Trainable params: 13,505		
Non-trainable params: 0		
Сеть глубокого обучения с тремя скрытыми слоями		
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 64)	9280
dense_2 (Dense)	(None, 64)	4160
dense_3 (Dense)	(None, 64)	4160
dense_4 (Dense)	(None, 1)	65
Total params: 17,665		
Trainable params: 17,665		
Non-trainable params: 0		

На рис. 2 представлен условный граф NNSM с входной информацией (верхний слой графа), одним скрытым слоем и выходным результатом (нижний слой графа) для задач регрессии. Заметим, что визуализации архитектур нейронных сетей в данном исследовании условны по причине большого объема как входных данных, так и высокого числа нейронов в скрытых слоях.

Данная архитектура сети является наиболее простой для обучения и менее энергоемкой (табл. 2).

В данном исследовании предполагалось не только выбрать самую подходящую архитектуру для прогнозирования цены закрытия, но и изучить процесс изменения ошибки и точности прогнозов в процессе обучения сети. Результаты показали, что ошибка прогноза стабилизируется с 30 эпохи обучения и дестабилизируется к последним эпохам (рис. 3). В данном случае возникает предположение о появлении эффекта переобучения нейронной сети. В результате обучения оценка функции потерь MSE составила 0,16 для оценочных выборок внутри сети и 0,002 для тренировочных выборок. Качество результатов весьма высокое для однослойной сети, тем не менее в работе были рассмотрены также сети глубокого обучения.

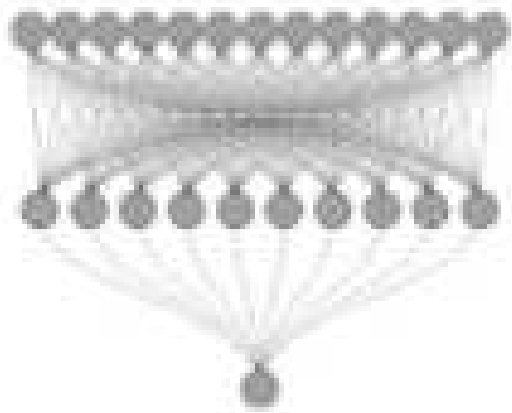


Рис. 2. Условный граф NNSM с одним скрытым слоем

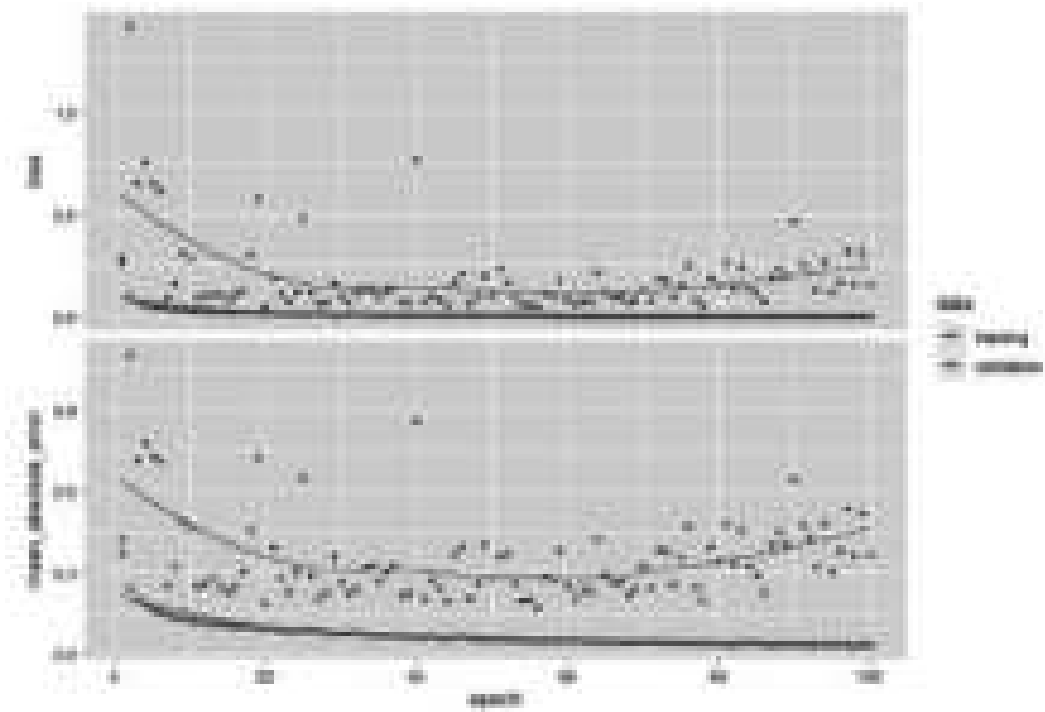


Рис. 3. MSE (loss) и MAE по тренировочным и оценочным выборкам в течение 100 эпох для NN с одним скрытым слоем

Искусственная нейронная сеть с двумя скрытыми слоями, каждый из которых состоит из 64 нейронов, условно представлена в виде графа на рис. 4. В ходе глубокого обучения были установлены стабилизация функции потерь MSE, а также оценки точности прогноза с 30-й эпохи. В результате обучения, включавшего 100 эпох, MSE достигла оценки 0,13 для оценочных выборок и 0,002 для тренировочной выборки, соответственно. Оценка качества прогноза, представленная MAE, составила 0,25 для валидации и 0,03 для тренировочной выборки (рис. 5). Для данной модели характерна стабилизация оценки с 30-й и до заключительной эпохи, что не дает повода для предположения о переобучении сети. Как следствие, эта архитектура сети может гарантировать постоянные оценки точности прогнозов и функции потерь.

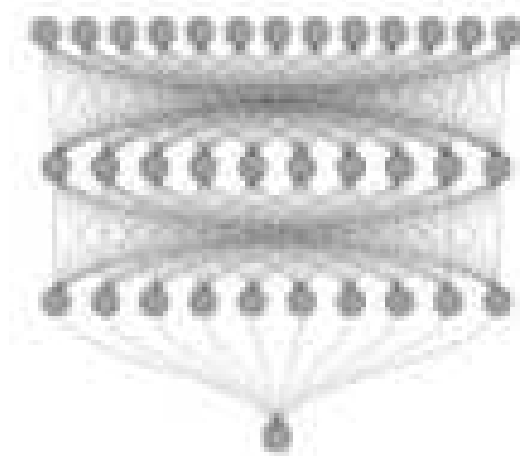


Рис. 4. Условный граф NNSM с двумя скрытыми слоями

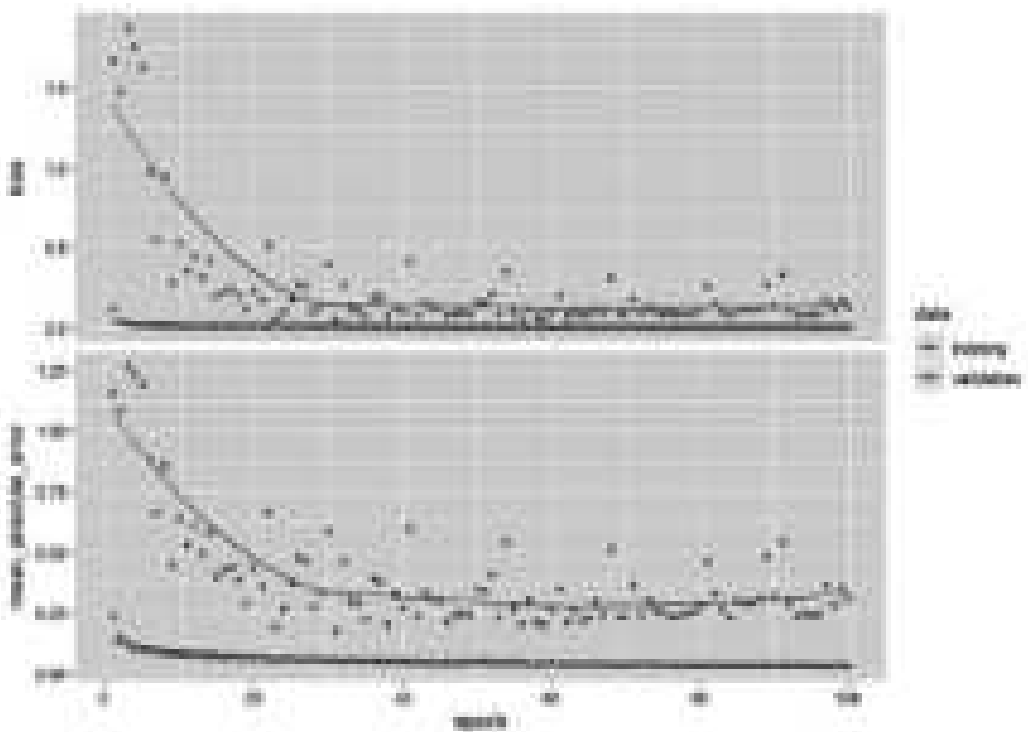


Рис. 5. MSE и MAE по тренировочным и оценочным выборкам в течение 100 эпох для DLN с двумя скрытыми слоями

В ходе исследования наблюдался рост ошибки прогноза с увеличением числа скрытых слоев сети. Тем не менее, как уже отмечалось, для выбора наилучшей архитектуры сети была предпринята попытка разбора сети глубокого обучения с тремя скрытыми слоями (рис. 6).

Последняя рассмотренная модель показала наихудшие результаты оценок MSE и MAE для 100 эпох обучения (рис. 7). Это может указывать на отсутствие необходимости в увеличении скрытых слоев модели. В данном исследовании сравнивались результаты нейронных сетей с различным числом скрытых слоев, при прочих равных условиях.

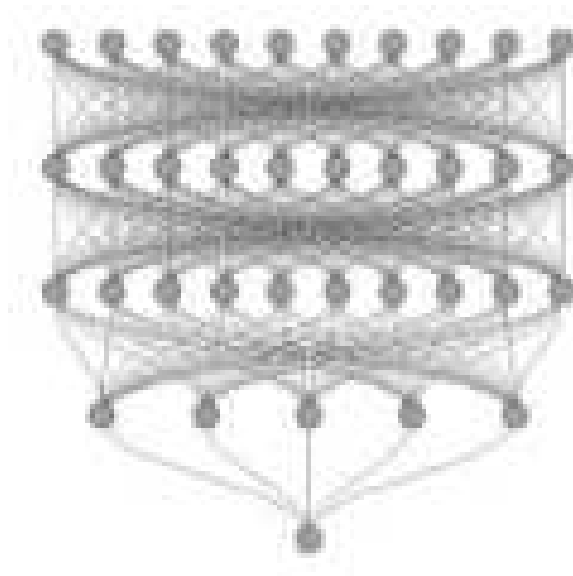


Рис. 6. Условный граф NNSM с тремя скрытыми слоями

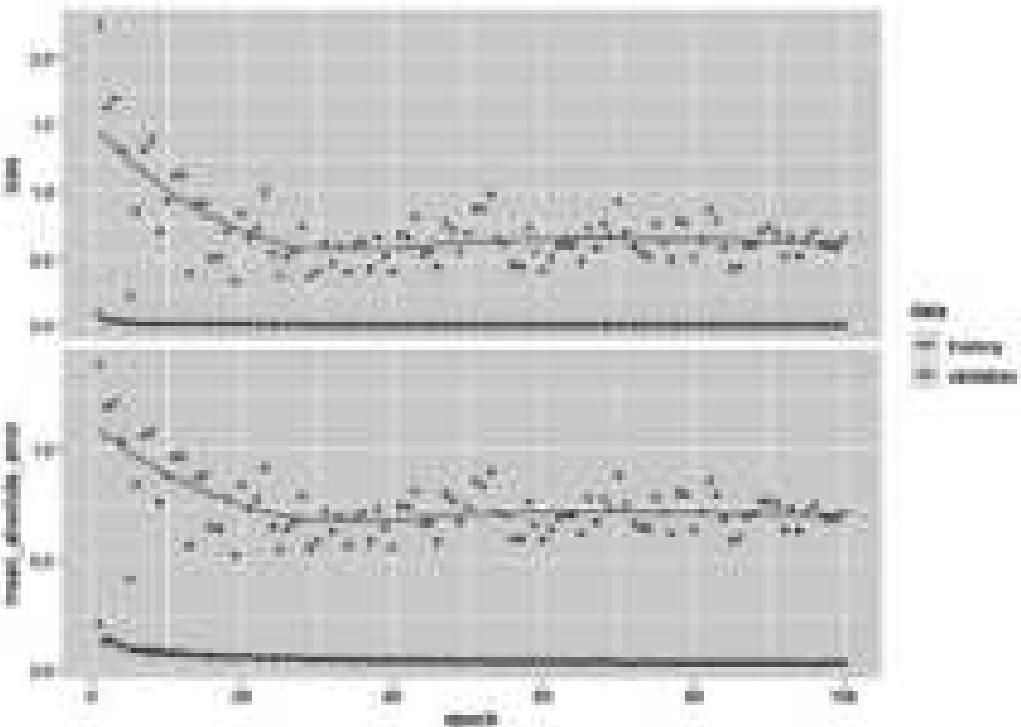


Рис. 7. MSE и MAE по тренировочным и оценочным выборкам в течение 100 эпох для ИНС с тремя скрытыми слоями

Для обучения нейронной сети с меньшим количеством скрытых слоев требуется меньше затрат времени. Расчет времени, требуемый для обучения сети, представлен в табл. 2. Более глубокая нейронная сеть требует больших затрат времени, что и было подтверждено в ходе исследования. Данный индикатор свидетельствует о целесообразности выбора более простой модели, при прочих равных условиях.

Таблица 2

Результаты времени, затраченного системой на обучение нейронных сетей

Нейронная сеть	Время, сек		
	Пользователь, сек	Система, сек	Затраты времени в программе, сек
ИНС с 1 скрытым слоем	30,89	9,44	136,42
ИНС с 2 скрытыми слоями	38,36	15,02	162,76
ИНС с 3 скрытыми слоями	43,05	20,48	186,29

В результате, вышеупомянутые нейронные сети были переобучены на период 30 эпох вследствие стабилизации ошибки и отсекаемого возможного эффекта переобучения, например для однослойной нейронной сети. Результаты, полученные при повторном исследовании, показали лучшие оценки. Оценки точности прогнозов по тестовым выборкам показали наилучшие результаты по однослойной нейронной сети, как и по тренировочным выборкам (MSE 0,003 и MAE 0,04) (табл. 3).

Таблица 3

Результаты оценок точности прогноза нейронной сети по тестовой выборке

Вид нейронной сети	Оценки точности	
	MSE	MAE
NN с 1 скрытым слоем	0,003	0,04
DLN с 2 скрытыми слоями	0,013	0,09
DLN с 3 скрытыми слоями	0,03	0,14

На рис. 8 представлены результаты прогнозных значений относительно фактических по данным тестовой выборки. Как подчеркивалось ранее, тестовые значения выборки были перемешаны случайно при разделении совокупности на тестовую и тренировочную выборки. В качестве основной модели была принята NNSM с одним скрытым слоем.

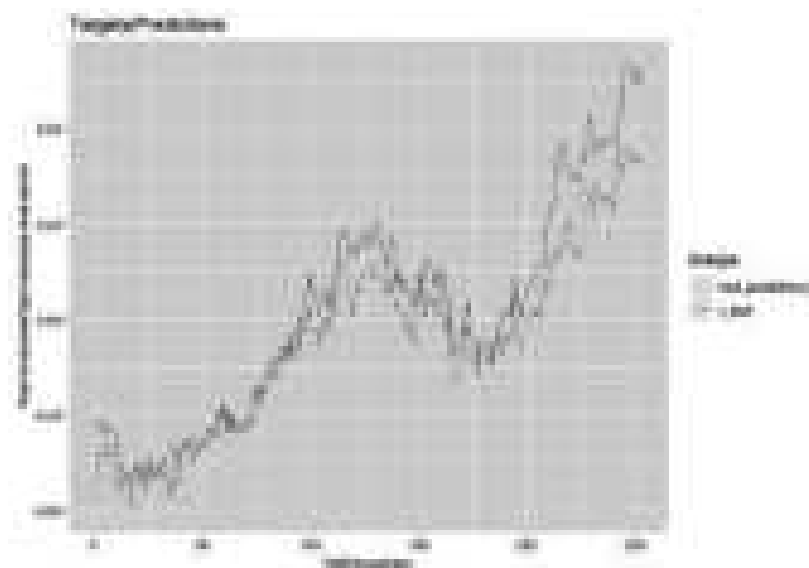


Рис. 8. Фактические значения и результаты прогнозов по тестовой выборке

Полученные с помощью нейронной сети NNSM прогнозы довольно близки к фактическим значениям. Оценка функции потерь MSE, равная 0,3% для NNSM с одним скрытым слоем минимальна по отношению к прочим рассмотренным архитектурам так же, как и MAE, равная 4% (табл. 3). Кроме того, данная архитектура требует меньших затрат времени для обучения (табл. 2). Преимущества библиотеки Keras для построения сетей глубокого обучения заключаются в простом синтаксисе кода в R, а также в высокой популярности.

Выводы

Обзор литературы и проведенное исследование показали, что текущее направление развития остается актуальным. Можно сказать, что методы машинного обучения, как и методы анализа временных рядов, сегодня являются неотъемлемой частью прогнозирования финансовых рынков. Пока существует фондовый рынок, проблема прогнозирования курса акций будет сохранять актуальность и порождать новые методы. Перебор архитектур сети, а также выбор иных характеристик, например функций активации или стандартизации исходных данных, могут дать иные результаты.

Источники

- Boston Dynamics. URL: <https://www.bostondynamics.com/> (дата обращения: 08.04.2019).
- Chollet F.* Keras. 2015. URL: <https://github.com/fchollet/keras> (дата обращения: 08.04.2019).
- Ciaburro G., Venkateswaran B.* Neural Network With R. Smart Models Using CNN, RNN, deep Learning, and Artificial Intelligence Principles. Packt Publishing Ltd., 2017.
- Elliot A., Hsu C. H.* Time Series Prediction: Predicting Stock Price. 2017. URL: <https://arXiv preprint arXiv:1710.05751> (дата обращения: 08.04.2019).
- Ghorbani M., Chong E. K. P.* Stock Price Prediction using Principle Components. 2018. URL: <https://arXiv preprint arXiv:1803.05075> (дата обращения: 08.04.2019).
- Heinze S.* A Simple Deep Learning Model for Stock Price Prediction Using Tensor Flow. URL: <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877> (дата обращения: 08.04.2019).
- Hinton G.* Neural Networks for Machine Learning. 2014 URL: <https://www.coursera.org/learn/neural-networks/supplement/K6F7N/lecture-slides-and-resources> (дата обращения: 08.04.2019).
- И'Ясов Р. Kh., Куразова Д. А.* Forecasting the Situation in the Financial Market and Evaluating its Impact on the Industry Development // Дайджест-финансы. 2018. № 1 (245). URL: <https://cyberleninka.ru/article/n/forecasting-the-situation-in-the-financial-market-and-evaluating-its-impact-on-the-industry-development> (дата обращения: 08.04.2019).
- Le Cun Y., Bengio Y., Hinton G.* Deep Learning // Nature. 2015. Vol. 521. N 7553. P. 436.
- Marcek D.* Stock Price forecasting: Statistical, Classical and Fuzzy Neural Network Approach // MDAI / Ed. by V. Torra, Y. Narukawa. Vol. 3131. Springer, 2004.
- Minsky M., Papert S. A.* Perceptrons: An Introduction to Computational Geometry. MIT press, 2017.
- Pai P. F., Lin C. S.* A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting // Omega. 2005. Vol. 33. N 6. P. 497–505.
- Palaniappan V.* AIAlpha: Using Machine Learning to Predict Stocks. URL: <https://medium.com/engineer-quant/alphaai-using-machine-learning-to-predict-stocks-79c620f87e53>.
- Xu S. Y.* Stock Price Forecasting Using Information from Yahoo Finance and Google Trend. 2014. URL: www.econ.berkeley.edu/sites/default/files/Selene%20Yue%20Xu.pdf
- Yibin Ng.* Machine Learning Techniques applied to Stock Price Prediction. URL: <https://towardsdatascience.com/machine-learning-techniques-applied-to-stock-price-prediction-6c1994da8001> (дата обращения: 08.04.2019).
- Yue-Gang Song, Yu-Long Zhou, Ren-Jie Han.* Neural Networks for Stock Price Prediction // Journal of Difference Equations and Applications. 2018. P. 1–14.

References

- Boston Dynamics*. Available at: <https://www.bostondynamics.com/> (assessed: 08.04.2019).
- Chollet F. *Keras*. 2015. Available at: <https://github.com/fchollet/keras> (assessed: 08.04.2019).
- Ciaburro G., Venkateswaran B. *Neural Network With R. Smart Models Using CNN, RNN, deep Learning, and Artificial Intelligence Principles*. Packt Publishing Ltd., 2017.
- Elliot A., Hsu C. H. *Time Series Prediction: Predicting Stock Price*. 2017. Available at: <https://arXiv preprint arXiv:1710.05751>.
- Ghorbani M., Chong E. K. P. *Stock Price Prediction using Principle Components*. 2018. Available at: <https://arXiv preprint arXiv:1803.05075> (assessed: 08.04.2019).
- Heinze S. *A simple deep learning model for stock price prediction using Tensor Flow*. Available at: <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877> (assessed: 08.04.2019).
- Hinton G. *Neural Networks for Machine Learning*, 2014. Available at: <https://www.coursera.org/learn/neural-networks/supplement/K6F7N/lecture-slides-and-resources> (assessed: 08.04.2019).
- Il'Yasov R. Kh., Kurazova D. A. Forecasting the situation in the financial market and evaluating its impact on the industry development. *Digest Finance*, 2018, N 1 (245). Available at: <https://cyberleninka.ru/article/n/forecasting-the-situation-in-the-financial-market-and-evaluating-its-impact-on-the-industry-development> (assessed: 08.04.2019).
- Le Cun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, N 7553, p. 436.
- Marceck D. Stock price forecasting: Statistical, classical and fuzzy neural network approach. *MDAI*, 2004, vol. 3131, pp. 41–48.
- Minsky M., Papert S. A. *Perceptrons: An introduction to computational geometry*. MIT press, 2017.
- Pai P. F., Lin C. S. A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, 2005, vol. 33, N 6, pp. 497–505.
- Palaniappan V. *ALAlpha: Using machine learning to predict stocks*. Available at: <https://medium.com/engineer-quant/alphaai-using-machine-learning-to-predict-stocks-79c620f87e53> (assessed: 08.04.2019).
- Song Y. G., Zhou Y. L., Han R. J. *Neural networks for stock price prediction*. Available at: <https://arXiv preprint arXiv:1805.11317>.
- Xu S. Y. *Stock price forecasting using information from Yahoo finance and Google trend*. UC Brekley, 2014.
- Yibin Ng. *Machine Learning Techniques applied to Stock Price Prediction*. Available at: <https://towardsdatascience.com/machine-learning-techniques-applied-to-stock-price-prediction-6c1994da8001> (assessed: 08.04.2019).