



НАУЧНЫЙ ЖУРНАЛ НАУКА И МИРОВОЗЗРЕНИЕ

АНАЛИЗ ДАННЫХ ЦЕНОВЫХ КОТИРОВОК АКЦИЙ КОМПАНИЙ МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ

Алексеев Антон Викторович

студент, Московский политехнический университет РФ, г. Москва

Царькова Наталья Ивановна

научный руководитель, канд. пед. наук, профессор кафедры «Прикладная информатика»,
Московский политехнический университет, РФ, г. Москва

Аннотация. Экономика и финансы — самые распространённые предметные области, в которых нейросетевые методы используются чаще всего. Алгоритмы нейросетей представляются в виде математического ядра, их можно применять для принятия решения в интеллектуальных и экспертных системах, для моделирования оболочек, для баз знаний и во многих других областях. Применение нейросетевых технологий имеет большое значение для прогнозирования динамики экономических показателей и финансовых рядов.

Ключевые слова: Большие данные, нейронные сети, фондовый рынок, python, предсказания, временной ряд.

Прогнозирование фондового рынка — классическая проблема на стыке финансов и информатики. В отношении этой проблемы знаменитая гипотеза эффективного рынка дает пессимистический взгляд и подразумевает, что финансовый рынок эффективен.

Некоторые исследования пытаются измерить различные уровни эффективности для зрелых и развивающиеся рынков, в то время как другие исследования пытаются построить эффективные прогнозы модели фондовых рынков.

Временной ряд — это последовательность наблюдений, проводимых с относительно регулярными интервалами. Примеры временных рядов: выпуск продукта, количество проданных страховых продуктов и т. д. Временные ряды широко используются в статистике, обработке сигналов, распознавании речи, эконометрике, финансах, астрономии, инженерии и многих других областях знаний, где значение измеряется во времени. Анализ временных рядов объединяет множество моделей для извлечения значимой информации из данных для прогнозирования будущей стоимости на основе ретроспективной информации. Временные ряды всегда отсортированы по времени. Модели временных рядов отражают факт того, что наблюдения, близкие по времени, будут более тесно связаны, чем наблюдения, расположенные дальше друг от друга.

Актуальность исследования заключается в совершенствовании методов анализа данных ценовых котировок акций компаний, расширение статистики баз данных. Объектом исследования послужили теоретические и практические разработки в применении различных моделей временных рядов к ценовым котировкам акций компаний. Предмет исследования — теоретические и методологические аспекты применения статистических методов в анализе временных рядов. В результате исследования должны быть

проанализированы данные исследуемых временных рядов котировок акций, а также модели обучения на основе шумовых маркеров. Полученные модели должны применяться для построения прогнозов на будущие периоды. Прогнозирование временных рядов традиционно выполняется эконометрикой с использованием различных методов.

В машинном обучении искусственные нейронные сети классы моделей, вдохновленные биологическими нейронными сетями, которые содержат взаимосвязанные узлы или блоки адаптивной обработки. Какие делает ИНС важными из-за их адаптивного характера, эта особенность делает ИНС - хорошо подходящим инструментом для аппроксимации сильно нелинейных и многомерных функций. Поскольку окончательная модель нейронной сети предсказывает временные ряды значения, она используется слоями LSTM в своей архитектуре для противодействия времени, связанным с этим проблемами, такие как «Проблема исчезающего градиента». В разнице между LSTM и другими традиционными рекуррентными нейронами сетями — это способность обрабатывать и прогнозировать временные ряды большого размера без потерь важной информации.

Таким образом, общая схема для анализа данных ценовых котировок ценных бумаг будет выглядеть следующим образом:

1. Определение нужного временного отрезка, на котором будет происходить выборка. Создание базы данных.
2. Указание входных и выходных значений для нейронной сети.
3. Нормализация данных.
4. Предварительная обработка данных. Обучение на шумовых маркерах.
5. Основное обучение нейронной сети.
6. Создание прогноза и тестирование итоговой модели.

Для прогнозирования ценовых котировок акций компаний используют множество способов, но наиболее эффективными являются модели экспоненциального сглаживания и нейронные сети. Наиболее популярные модели — это модели LSTM, причем подбор параметров для этих моделей должен происходить по алгоритмам основанных на обучении с учителем

В дальнейшей работе также возможно улучшение сети LSTM путем обучения модели в процессе непосредственной работы с полученными данными. На графике обучения модели LSTM видно, что с каждой новой эпохой обучения, точность обучения модели увеличивается. При этом точность, предсказанная моделью на каждой отдельной эпохе обучения, составляет приблизительно 58%. Таким образом, учитывая, что длительность формирования модели составляло около 60 лет, а качество прогнозирования для каждой из эпох обучения было оценено, как близкое к 64%, можно сделать вывод, что для прогнозирования временных рядов, имеющих большой размах, LSTM является хорошей моделью.

Для реализации поставленной цели были выбраны методы прогнозирования с учителя на основе рекуррентной нейронной сети с архитектурой LSTM, написанной на языке python в средстве разработки jupyter notebook, основной метрикой точности прогноза будет служить RMSE.

Структура набора данных, содержит помимо даты также цену открытия, нижние и верхние границы ценового диапазона, а также цену закрытия, помимо этого можно посмотреть объем, проторгованный за тот же промежуток времени:

	Date	Open	High	Low	Close	Volume	OpenInt
0	1984-09-07	0.42388	0.42902	0.41874	0.42388	23220030	0
1	1984-09-10	0.42388	0.42516	0.41366	0.42134	18022532	0
2	1984-09-11	0.42516	0.43668	0.42516	0.42902	42498199	0
3	1984-09-12	0.42902	0.43157	0.41618	0.41618	37125801	0
4	1984-09-13	0.43927	0.44052	0.43927	0.43927	57822062	0

Рисунок 1. Формат даты временного ряда

Данный набор данных описывает ежедневное колебание ценовых котировок акций компаний AAPL, NFLX, MA, BA. Для реализации концепции глубокого обучения разделим полученный набор данных на обучающую выборку и тестовую, для тестирования возьмем последние полгода. Во время обучения в качестве цели использовалось изменение курса акций. Обучение проходило по методу «перекрестной проверки» — это обучение и тестирование, которые проводились одновременно. Каждый раз для обучающей выборки вычислялась оценка изменения курса акций, в результате чего оценивалось значение каждого из факторов. После каждого тестирования проводилось изменение обучающей и тестовой выборок, и формировалась новая оценка изменения, которая служила для расчета новой цели. Данная техника обучения называется «метод последовательных приближений». Обработка входных и выходных данных ценовых котировок акций компаний Для правильной работы нейронной сети необходимо понять, какие данные подавать на вход, для этого, определим какой столбец имеет наибольшее влияние на цену акции компании.

Для того, чтобы определить, какой столбец использовать необходимо построить матрицу корреляции этих параметров между собой:

	Open	Close	Low	High
Open	1.000000	0.994434	0.997605	0.996495
Close	0.994434	1.000000	0.996758	0.997724
Low	0.997605	0.996758	1.000000	0.995781
High	0.996495	0.997724	0.995781	1.000000

Рисунок 2. Матрица корреляции значений ряда

Как видно из рисунка, все значения имеют близкое к единице значение, это связано с тем, что в течении дня цена может колебаться, как возле одного значения, так и не значительно уйти от него. Но все же в ряду присутствует волатильность, что позволяет нам определить, значение наиболее тесно связанное с ценой акции. Таким образом, можно заметить, что наибольшим влиянием на значение ряда является цена закрытия цены. Для правильной работы нейронной сети необходимо понять, какие данные подавать на вход, для этого, определим какой столбец имеет наибольшее влияние на цену акции компании. Для того, чтобы определить, какой столбец использовать необходимо построить матрицу корреляции этих параметров между собой:



	Open	Close	Low	High
Open	1.000000	0.994434	0.997605	0.996495
Close	0.994434	1.000000	0.996758	0.997724
Low	0.997605	0.996758	1.000000	0.995781
High	0.996495	0.997724	0.995781	1.000000

Рисунок 3. Матрица корреляции значений ряда