

Гипотеза эффективности финансовых рынков или технический анализ: проверка с помощью искусственного интеллекта

The Hypothesis of the Efficiency of Financial Markets or Technical Analysis: Checked with the Help of Artificial Intelligence

DOI 10.12737/24536

Получено: 24 декабря 2016 г. / Одобрено: 13 января 2017 г. / Опубликовано: 17 февраля 2017 г.

**Басовский А.Л.**

Канд. экон. наук,
Мувикс Медиа Ретворкс Лтд,
Израиль, 4951125, ПетахТиква, ул. Эфаль 25
e-mail: basovskiy@yahoo.com

Basovskiy A.L.

Ph.D. in Economics,
Muvix Media Networks Ltd.
25 Efal St. Petach Tikva, 4951125, Israel
e-mail: basovskiy@yahoo.com

Аннотация

Цель работы состояла в проверке гипотезы эффективности финансовых рынков и гипотезы возможности прогнозирования динамики рынков путем технического анализа. Использовались методы искусственного интеллекта и данные дневных цен закрытия всех акций, входящих в S&P 500, с момента их размещения — 3 242 204 цен закрытия. «Обучалась» модель признаков, используемых в техническом анализе, с использованием данных 400 случайно выбранных акций, проверялась модель с использованием оставшихся акций, входящих в S&P 500, за весь период котировки акций. Количество объектов технического анализа в исследуемом множестве составили 20 000. Точности оценок, основанных на методах технического анализа, реализуемых методом искусственного интеллекта оказалась близка к 50%. Полученные результаты ставят под сомнение возможность технического анализа определять, повысится или понизится цена акции, и подтверждают теорию эффективности финансовых рынков, по крайней мере для рынка акций, входящих в S&P 500.

Ключевые слова: эффективность финансовых рынков, технический анализ, модель, искусственный интеллект, S&P 500.

Теория эффективности финансовых рынков предполагает, что вся имеющаяся и доступная участникам рынка информация отражена в цене актива (акции, облигации или иного финансового актива) [1, 2]. Движение цен случайно, инвесторы не могут добиться доходности, превышающей общую доходность рынка. Некоторые экономисты, специализирующиеся в работе с финансовыми рынками, предполагают, что эффективность финансовых рынков — лишь гипотеза. Они предполагают, что с помощью анализа, называемого техническим, возможно прогнозировать направление изменений цен финансовых рынков. Технический анализ предполагает, что цены активов имеют тенденцию изменяться в определенном направлении. Соответственно, определив это направление, можно предсказать, будет ли увеличиваться или уменьшаться цена актива в ближайшем будущем, и, таким образом, получить доходность, выше общей доходности рынка [3].

Одним из наиболее популярных разделов технического анализа является изучение временных рядов, представленных графиками цен. Инвестор, изучая

Abstract

The purpose of work consisted in check of a hypothesis of efficiency of the financial markets and hypotheses of a possibility of forecasting of dynamics of the markets by the technical analysis. Methods of artificial intelligence and data of the day prices of closing of all actions entering S&P 500 used from the moment of their placemen — 3 242 204 prices of closing t. The model of the signs used in the technical analysis, with use of these 400 incidentally chosen actions «studied», the model with use of the remained actions entering S&P 500 for the entire period of stock quotation checked. The quantity of subjects of the technical analysis in the studied set was made to 20 000. To accuracy of the estimates based on the methods of the technical analysis realized by method of artificial intelligence it was close to 50%. The received results call into question a possibility of the technical analysis to define, the action price will raised or will go down, and confirm the theory of efficiency of the financial markets, on extremely measure for the stock market, entering S&P 500.

Keywords: efficiency of the financial markets, technical analysis, model, artificial intelligence, S&P 500.

график, делает предположение о том, какую фигуру формирует график. Выделяют две категории фигур: медвежьей и бычьей, предсказывающие понижение или повышение цены соответственно. Среди каждой из категорий есть множество конкретных фигур: голова и плечи, чашка с ручкой и другие. На рис. 1 приведен пример графика, который формирует бычью фигуру «Чашка с ручкой», сигнализирующую о вероятном повышении цены [4].

Технический анализ субъективен. Оценка, формирует ли конкретный график цен фигуру, является субъективной оценкой эксперта. Практики технического анализа утверждают, что эксперт в области технического анализа может, глядя на график, определить, формирует ли этот график фигуру. Он также может определить, какую именно фигуру формирует график и, таким образом, с высокой вероятностью предсказать, повысится или понизится цена данного актива в ближайшем будущем.

Современные методы искусственного интеллекта позволяют создать математическую модель, которая может распознавать закономерности в данных так же,



Рис. 1. Фигура «Чашка с ручкой» [4, с. 192]

как это делает человек-эксперт, исключив при этом субъективный подход при оценивании. В практике использования методов искусственного интеллекта используют термин «обучить модель» в смысле, что на основании огромного объема данных выявляются параметры модели. Затем эти параметры используются для предсказания результата нового набора входных данных.

Представляет интерес попытка «обучить» математическую модель признаков, используемых в техническом анализе, используя графики большого числа акций за различные периоды времени. Если модель сможет с высокой вероятностью, получив в качестве входных данных график, который не участвовал в процессе «обучения» предсказать, повысится ли цена актива в конце какого-то определенного срока или нет, то это можно будет считать подтверждением точки зрения сторонников технического анализа. Если вероятность предсказания такой модели не будет существенно превышать 50%, можно говорить о том, что гипотеза об эффективности рынков получила еще одно эмпирическое подтверждение. Это же будет означать, что технический анализ не способен обеспечить объективное предсказание.

Описание методов и данных

В настоящей работе использовались временные ряды цен акций 500 компаний, составляющих индекс Стандратс Энд Пуэрс 500 (S&P 500). Была построена математическая модель для классификации графиков цен акций. Затем была проанализирована точность модели. Было оценено, с какой точностью данная модель может, «посмотрев» на график цен акции, предсказать, повысится ли цена этой акции

или понизится. Для вычислений и отображения графиков в данной работе использовалась библиотека Анаконда версии 4.2.9 для языка программирования питон.

Для каждого графика требуется определить, повысится ли цена акции, понизится или останется неизменной. Эта задача, как известно, является задачей классификации. Формальное ее определение следующее. На основании имеющегося конечного множества объектов (выборка), о которых известна их принадлежность к классам, требуется построить математическую модель, которая позволит классифицировать произвольный объект из заданной выборки [5]. Входными данными для рассматриваемой задачи является график цен акций, как правило, несколько сотен чисел. Количество классов строго определено — три: повышение цены, неизменная цена или понижение цены. Существует множество методов для решения задачи классификации. Наиболее простые — это регрессия и наивный байесовский классификатор. Наиболее сложные с точки зрения построения и наиболее дорогостоящие с точки зрения вычислений — это методы, комбинирующие несколько видов нейронных сетей. Для использования в данной работе автор выбрал метод опорных векторов.

Метод опорных векторов предложили В.Н. Вапник и А.Я. Червоненкис [6]. Этот метод первоначально разработан, как и многие другие методы решения задачи классификации, для разделения множества на два класса. Метод опорных векторов сводит задачу классификации к поиску гиперплоскости с максимальным расстоянием между классами. В настоящее время этот метод широко используется для решения задачи классификации с множеством классов. Метод опорных векторов является более

простым и требует меньше вычислений, чем методы нейронных сетей. Метод опорных векторов использовался для решения задач, подобных рассматриваемой в настоящей работе, — входной ряд из нескольких сотен чисел, а количество классов — строго определено и невелико. Примером может послужить задача распознавания рукописных цифр — входным рядом является картинка 20 на 20 пикселей (400 чисел), количество классов, соответственно, 10 (представляющие цифры от нуля до девяти). Метод опорных векторов с использованием радиальной базисной функцией Гаусса показал не менее точные результаты, чем метод нейронных сетей [7].

В данной работе использовался метод опорных векторов с радиальной базисной функцией Гаусса в качестве ядра. Ядром в методе опорных векторов называют функцию, которая переводит входной ряд в другое множество с большей размерностью, что позволяет найти разделяющую гиперплоскость, которая будет не линейна по отношению к исходному множеству.

Графики цен акций являются графиками временных рядов. Точка ряда может представлять разные периоды — от минуты до года. В данной работе рассматриваются графики, точкой в которых является торговый день. Под торговым днем здесь и далее понимался день, в который проходили торги по данной акции на бирже, как правило, это рабочие дни, за исключением дней праздников и форс-мажорных обстоятельств, влияющих на биржу в целом или на котировки акции данной компании. Такие графики наиболее просты для понимания, широко используются на практике и наиболее часто встречаются в научной и учебной литературе, курсах по техническому анализу. Кроме того, одним из основополагающих принципов технического анализа является так называемая фрактальность — подобие самим себе.

Утверждается, что характеристики фигур сохраняются вне зависимости от периодичности графика [3]. Например, пусть определили бычью фигуру на дневном графике цен (где каждая точка по оси X — это один торговый день), которая занимает 100 точек (торговых дней), и предположили, что цена повысится через 20–50 дней. Если увидим похожую фигуру на недельном графике, где каждая точка — это неделя работы биржи, то можем сделать вывод, что цена повысится через 20–50 недель.

В настоящей работе использовались графики по 300 чисел. Это означает, что график покрывает период немногим более календарного года. Каждое число — это цена закрытия акции, т.е. цена, которая установилась на конец торгового дня и считается ценой акции до начала следующего торгового дня.

В качестве критерия повышения или понижения цены было определено следующее правило. Сравнивалась цена закрытия через 130 торговых дней (что примерно соответствует шести календарным месяцам) после последней точки графика (далее — «будущая» цена) с величиной последней точки графика. Если цена увеличилась на 10% и более, будем считать, что график сигнализирует о повышении цены акции. Пример такого графика приведен на рис. 2, где представлен временной ряд цен на акции компании «Корпорация Терадата», символ на нью-йоркской фондовой бирже — TDC.

Если будущая цена остается в пределах плюс-минус 10% от величины последней точки графика, будем считать, что цена осталась неизменной. И соответственно, если уменьшилась на 10% и более, примем, что цена снизилась.

Графики были построены по принципу скользящего окна с интервалом в двадцать торговых дней с даты начала торгов по данной акции по 31 марта



Рис. 2. График повышения

2016 г. Например, торги по акции компании «Майкрософт» начались 13 марта 1986 г. Первый график — это первые 300 торговых дней, т.е. период с 13 марта 1986 г. по 20 мая 1987 г. Второй график — это 300 торговых дней, начиная с 10 апреля 1986 г. (20 рабочих дней после 13 марта) и т.д. Интервал в двадцать дней был выбран по следующей причине. С одной стороны, слишком малый интервал, например в один или два дня, привел бы к тому, что алгоритм рассматривал огромное количество фактически идентичных графиков. С другой стороны, слишком большое расстояние между начальными точками графиков привело бы к потере ценных входных данных. Фигуры, которые, как было указано, определяются субъективно на глаз, как правило, формируются на нескольких десятках или даже сотнях точек. Поэтому 20 точек — несущественная потеря. При этом количество входных данных сокращается на порядок, что заметно увеличит скорость вычислений.

При этом если в период времени между начальной точкой графика и днем, используемым в качестве критерия типа графика, т.е. если в один из 430 торговых дней компания проводила дробление или консолидацию акций, эти периоды не участвовали в анализе. Иными словами, в данной работе мы анализировали только те случаи, когда цена акции изменяется в результате торгов на бирже.

Когда эксперт по техническому анализу смотрит на график, он определяет форму фигуры по расположению точек относительно друг друга. Например, Чарльз Кирпатрик [3] подробно описал ряд фигур исключительно в терминах соотношений различных частей фигуры относительно друг друга. В связи

с этим абсолютные значения цены могут затруднить вычисления, так как алгоритм будет сравнивать несопоставимые величины. Например, одна акция торгуется в диапазоне 10–20 долларов за акцию, а другая 300–600 долларов за акцию. При этом они формируют похожие фигуры на графике. В связи с изложенным в качестве входного ряда для каждой акции использовались не абсолютные данные о цене, выраженные в долларах, а доходность каждого торгового дня относительно первого торгового дня. Доходность рассчитывалась по формуле

$$r_i = (p_i - p_1) / p_1, \tag{1}$$

где r_i — доходность i -го дня; p_i — цена акции в i -й день; p_1 — цена акции в первый день данного ряда (первая точка на графике).

Как видно из рис. 3, график абсолютных цен и график доходности визуально идентичны. Данные о ценах акций были взяты с сайта finance.yahoo.com. Были использованы данные о 504 акциях (4 из 500 компаний, входящих в индекс Стандратс Энд Пуэрс 500, имеют по два класса акций, которые торгуются под разными символами, например, акции GOOG и GOOGL компании Алфавит, управляющей компанией Гугл). Всего были обработаны 3 242 204 цен закрытия. На основании этих данных были смоделированы 130 033 графиков.

Целью алгоритма, решающего задачу классификации, является определение, к какому множеству относится новый объект, т.е. тот, который был не известен на стадии обучения алгоритма. В статистике выделяют два типа ошибок подобного рода алгоритмов: смещение и дисперсию. Под смещением, на-



Рис. 3. График абсолютных цен акции и график доходности относительно первого торгового дня

помним, понимают недостаточную точность связей между переменными. Дисперсия — это противоположное явление, а именно слишком большая чувствительность к малейшим колебаниям входного набора данных. В любом случае новый объект может быть классифицирован неправильно.

Для выявления и решения проблемы дисперсии применяют следующую технику, которая использовалась в настоящей работе. Вся совокупность объектов входного ряда разбивается на три множества: обучающий набор данных, набор данных для перекрестной проверки и проверочный набор данных. Обучающий набор данных используют непосредственно для «обучения» алгоритма (модели). Перекрестный набор данных используют для оптимизации параметров алгоритма. При этом алгоритм многократно проходит цикла «обучения» с различными параметрами, и выбираются параметры, наиболее точно классифицирующие перекрестный набор данных. Наконец, проверочный набор данных используется для окончательной проверки алгоритма и оценки его точности.

В данной работе под ошибкой алгоритма понималась доля графиков, классифицированных неправильно. Например, график, за которым фактически последовало повышение цены, отнесен к классу «неизменная цена». Такое определение ошибки наиболее просто и понятно по сравнению с другими показателями ошибки, используемыми в статистике. Предварительный анализ данных позволяет также использовать этот показатель, так как в данных отсутствует так называемое искажение класса, т.е. ситуация, когда один из классов представлен непропорционально малым количеством объектов, чем другие классы.



Рис. 4. Распределение классов среди объектов обучающего множества

Как видно из диаграммы на рис. 4, все классы были представлены сопоставимыми числами, выраженными в десятках процентов. Поэтому можем использовать в качестве ошибки показатель, определенный выше.

Для определения типа ошибки алгоритма — дисперсии или смещения — используют кривые обучения. Для этого модель повторно «обучают» на возрастающем числе объектов обучающего множества. По вертикальной оси откладывают ошибку алгоритма, а по горизонтальной — количество единиц обучающего набора данных. Нанеся на график две линии — одну, отражающую ошибку обучающего набора, а вторую — ошибку перекрестного набора данных, можно определить тип ошибки алгоритма. При высоких значениях ошибки обоих графиков имеет место проблема смещения. При снижающейся ошибке обучающего набора и при стабильно высокой ошибке перекрестного набора имеет место проблема дисперсии. На рис. 5 приведены кривые обучения рассматриваемой задачи.

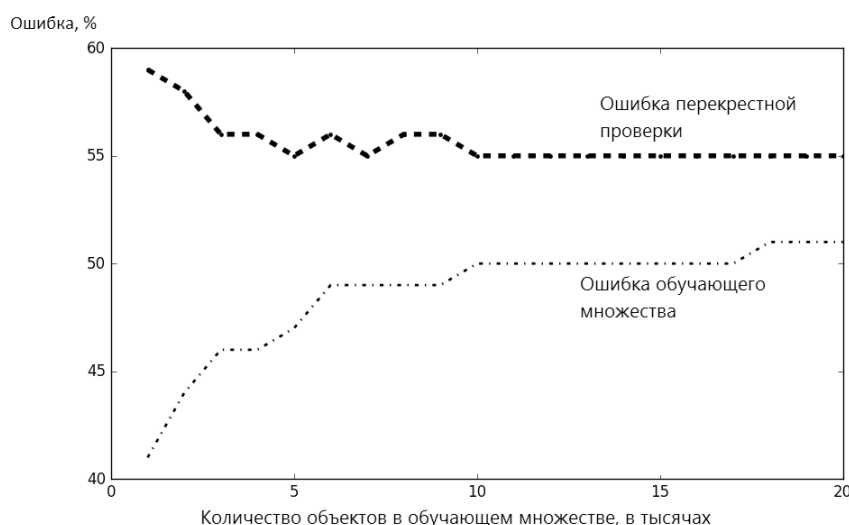


Рис. 5. Кривые обучения

Как видно из рис. 5, «обучали» алгоритм 20 тыс. раз, каждый раз увеличивая количество объектов обучающего множества на тысячу — от одной до двадцати тысяч. При этом множество перекрестной проверки составляло пять тысяч объектов, не входящих в обучающее множество. Таким образом, наблюдалась ошибка смещения в чистом виде: ошибка перекрестной проверки и ошибка обучающего множества стабильно высоки (ошибка около 50% — очень высокая ошибка) независимо от дополнительного количества входных данных начиная с 10 тыс. объектов обучающего множества. При этом ошибка перекрестной проверки не изменяется начиная с 10 тыс. объектов обучающего множества. Это позволяет в целях упрощения вычислений использовать 10 тыс. объектов обучающего множества для оптимизации параметров модели.

Как было указано выше, использовался метод опорных векторов с радиальной базисной функцией Гаусса. В этой модели есть два параметра, которые требуют оптимизации. Первый — так называемое *C*-значение, которое влияет на размер расстояния между гиперплоскостью, разделяющей классы и объектами класса. *C* с практической точки зрения параметр *C* снижают, чтобы решить проблему дисперсии. По умолчанию параметр принимает значение 1. Второй параметр — это значение γ (гамма) радиальной базисной функции Гаусса:

$$K(x, x') = e^{\gamma \cdot \|x - x'\|^2} \quad (2)$$

Для оптимизации параметров *C* и γ был использован следующий алгоритм. Для каждого из параметров был взят ряд, представленный девятью значениями геометрической прогрессии. Для *C* были использованы значения от 10^{-4} до 10^4 . Для γ — от 10^{-6} до 10^{-2} . Для каждой комбинации параметров алгоритм был «обучен» на значениях обучающего множества (10 тыс. объектов), затем была оценена ошибка предсказаний, используя в качестве входного ряда множество перекрестной проверки (5 тыс. объектов). Всего была выполнена 81 итерация. Оптимальными параметрами явились значения $C = 0,1$

и $\gamma = 1$. При данных параметрах ошибка перекрестной проверки была наименьшей и составила 53%.

Для «обучения» алгоритма все множество графиков было поделено на обучающее множество и проверочное множество. В обучающее множество вошли графики 400 случайным образом отобранных акций (из 504 акций компаний, входящих в индекс Сандерс Энд Пуэрс) — всего 105 049 графиков. В проверочное множество — графики других 100 акций, всего 24 983 графика. Графики остальных четырех акций не использовались. Для предварительного анализа (построение кривых обучения) и оптимизации использовались исключительно данные из первого множества. Проверочное множество использовалось только для окончательной оценки ошибки алгоритма.

Описание результатов

Ошибка алгоритма с оптимизированными параметрами составила 53%. Это означает, что алгоритм смог точно классифицировать только 47% графиков.

В данной работе была предпринята дополнительно попытка «обучить» алгоритм с параметрами *C* и γ по умолчанию ($C = 1$, а γ рассчитывается как отношение единицы к количеству объектов входного ряда). Кроме того, был использован упрощенный вариант метода опорных векторов — линейный метод опорных векторов, не использующий функцию ядра.

Результаты «обучения» каждого из этих алгоритмов приведены в табл. 1. Для всех методов использовались одни и те же входные данные и тот же способ оценки ошибки, описанные выше. Полученные оценки показывают, что наилучшие результаты обеспечивает применение оптимизированного метода опорных векторов с радиальной функцией Гаусса в качестве ядра ($C = 0,1; \gamma = 1$). Использование даже этого метода на огромном массиве данных приводит к точности классификации только около 50%. Это означает, что верные и неверные объективные оценки, основанные на использовании методов технического анализа, равновероятны.

Близость точности оценок, основанных на реализации методов технического анализа, реализуемых

Таблица 1

Точность классификации графиков цен акций различными алгоритмами

Описание метода	Ошибка классификации, %	Точность классификации, %
Оптимизированный метод опорных векторов с радиальной функцией Гаусса в качестве ядра ($C = 0,1 \gamma = 1$)	53	47
Метод опорных векторов с радиальной функцией Гаусса в качестве ядра с параметрами по умолчанию ($C = 1 \gamma = 10^{-5}$)	56	44
Оптимизированный метод опорных векторов без функции ядра ($C = 0,001$)	60	40
Метод опорных векторов без функции ядра с параметром по умолчанию ($C=1$)	60	40

методом искусственного интеллекта к 50%, свидетельствует о том, что, «посмотрев» на график цен, этот интеллект, не обладающий субъективностью, может предсказать, повысится или понизится цена акции с той же вероятностью, как это сделает человек, который проверяет основные положения теории вероятностей, подбрасывая монеты. Другими словами, можно предположить, эксперт в области технического анализа, изучающий графики акций и увидевший за свою карьеру все графики акций, входящих сегодня в индекс S&P 500, глядя на новый график цена, будет иметь тот же шанс на ошибку, что и человек, впервые увидевший график цен на акции.

Отметим, что возможности методов искусственного интеллекта ограничены, хоты и постоянно развиваются. Кроме того, в данной работе был использован лишь один из известных типов алгоритмов классификации. Интерпретация того, что является графиком цены и что считать повышением или понижением, дана достаточно узко. Данная работа не может считаться окончательным опровержением принципов технического анализа. Однако она ставит под сомнение возможность технического анализа определять, повысится или понизится цена акции в течение ближайших месяцев. Таким образом, данная работа служит еще одним подтверждением теории эффективных рынков.

Литература

1. Efficient Capital Markets: A Review of Theory and Empirical Work. Eugene F. Fama. *The Journal of Finance*. Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual

Meeting of the American Finance Association New York, N.Y. December, 28–30, 1969 (May, 1970), pp. 383–417.

2. Ivo Welch. *Corporate Finance*. 2nd edition, 2011. — 768 p.
3. Charles D. *Technical Analysis*. Kirkpatrick II. Pearson Education Inc., 2011. — 320 p.
4. William O'Neil. *The How to Make Money in Stocks Complete Investing System: Your Ultimate Guide to Winning in Good Times and Bad Paperback*. Au McGraw-Hill Education, 2010. — 500 p.
5. Marco Taboga. *Lectures on Probability Theory and Mathematical Statistics*. — 2nd Edition. CreateSpace Independent Publishing Platform, 2012. — 656 p.
6. Corinna Cortes, Vladimir Vapnik. Support-vector networks. *Journal of Machine Learning*. September 1995, Vol. 20, Issue 3, pp. 273–297.
7. The MNIST database of handwritten digits. YannLeCun, Corinna Cortes, Christopher J.C. Burges. URL: <http://yann.lecun.com/exdb/mnist/>

References

1. Efficient Capital Markets: A Review of Theory and Empirical Work. Eugene F. Fama. *The Journal of Finance*. Vol. 25, No. 2, *Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York*, N.Y. December, 28–30, 1969 (May, 1970), pp. 383–417.
2. Ivo Welch. *Corporate Finance*. 2nd edition, 2011. — 768 p.
3. Charles D. *Technical Analysis*. Kirkpatrick II. Pearson Education Inc., 2011. — 320 p.
4. William O'Neil *The How to Make Money in Stocks Complete Investing System: Your Ultimate Guide to Winning in Good Times and Bad Paperback*. Au McGraw-Hill Education, 2010. — 500 p.
5. Marco Taboga. *Lectures on Probability Theory and Mathematical Statistics*. — 2nd Edition. CreateSpace Independent Publishing Platform, 2012. — 656 p.
6. Corinna Cortes, Vladimir Vapnik. Support-vector networks. *Journal of Machine Learning*. September 1995, Vol. 20, Issue 3, pp. 273–297.
7. The MNIST database of handwritten digits. YannLeCun, Corinna Cortes, Christopher J.C. Burges. URL: <http://yann.lecun.com/exdb/mnist/>