

Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.585.5

Киндрук
Павел Алексеевич

Применение машинного обучения
для технического анализа рынка Форекс

АВТОРЕФЕРАТ

на соискание академической степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель
Сиротко С. И.
кандидат физико-математических
наук, доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

В мировой финансовой системе очевидные макроэкономические закономерности порой уживаются с хаотическими изменениями курсов валют и процентных ставок. Колебания цен валют, которые бывают довольно большими, сулят потенциально высокий доход от операций на валютном рынке. Поэтому многие финансовые институты, в частности, банки и инвестиционные компании проявляют активный интерес к подобного рода деятельности.

В техническом анализе для построения разнообразных эмпирических зависимостей используются сложные математические модели, например, теория игр, машинное обучение, математическая статистика и теория вероятностей. Знание некоторых особенностей кратковременного поведения цен позволяет на порядок повысить эффективность торговли на бирже.

За последние два десятилетия рынки стали более динамичными, и торговля с использованием алгоритмов машинного обучения может через какое-то время вытеснить традиционную биржевую торговлю. Как управляющие хедж-фондами, так и трейдеры в настоящее время сосредоточены на разработке программ, которые помогают их ежедневной торговле в стремлении увеличить прибыль.

Тем не менее, люди по-прежнему сохраняют преимущество, видя общую картину. Компьютеры по-прежнему не очень хорошо умеют определять моменты смены тренда на рынке и делать прогнозы с учетом реакции других участников рынка, таких как политики и банкиры, или предвидеть, как рынки будут двигаться. В дополнение к этому, данные, необходимые для прогнозирования рынков, становятся все более и более сложными.

На основании вышеизложенного можно выделить актуальную проблему обработки данных ценовых колебаний валют и построения моделей, способных прогнозировать ценовые движения на рынке.

Диссертационная работа посвящена разработке методов формирования набора обучающих данных на основе информации о ценовых колебаниях и исследованию применимости методов машинного обучения для решения задач прогнозирования направления движения цены на рынке Форекс.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является изучение применимости методов машинного обучения для прогнозирования ценовых движений на рынке Форекс и построение моделей, способных прогнозировать направление движения цены в следующий временной период.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) Изучить основные подходы технического анализа и структуру используемых для анализа рынка данных.
- 2) Изучить методы прогнозирования экономических временных рядов.
- 3) Сформулировать задачу машинного обучения для прогнозирования движения цены на рынке Форекс.
- 4) Построить и обучить модели, прогнозирующие ценовые движения.
- 5) Провести исследования применимости данных моделей на практике.

Объектом исследования являются ценовые движения, генерируемые участниками рынка.

Предметом исследования является подходы и методы машинного обучения для задач прогнозирования экономических временных рядов.

Основой диссертационной работы являются основные положения технического анализа. Их суть состоит в следующем:

- 1) Движения рынка учитывают все, т.е. любое политическое или экономическое событие уже было или будет учтено участниками рынка и соответственно отражено в цене актива.
- 2) Направление движения цен не хаотично, т.е. существуют предпосылки которые определяют куда цена будет двигаться.
- 3) История циклична, т.е. если рынок при определенных показателях двигался в каком-то направлении, то в следующий раз при тех же показателях он будет двигаться также.

Возможность прогнозирования движения цен значительно повышает эффективность работы на рынке.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя, заключается в формулировке целей и задач исследования.

Опубликованность результатов диссертации

Публикации по теме диссертации отсутствуют.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, и приложений. В первой главе представлен анализ предметной области, рассмотрены основные методы технического анализа, выявлены существующие подходы к применению машинного обучения в техническом анализе. Во второй главе изучены подходы к прогнозированию временных рядов методами машинного обучения, рассмотрены методы формирования обучающей выборки из данных временного ряда, выявлены методы построения и оценки точности моделей при прогнозировании временных рядов. В третьей главе приведены результаты исследований, сформулированы задачи классификации для прогнозирования ценового движения, выявлены некоторые недостатки в практическом применении построенных моделей.

Общий объем работы составляет 94 страниц, из которых основного текста – 47 страниц, 23 рисунка на 11 страницах, 29 таблиц на 12 страницах, список использованных источников из 27 наименований на 2 страницах и 2 приложения на 22 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В первой главе проведен анализ методов, математических и статистических моделей, используемых в техническом анализе рынка. Рассмотрены графические методы технического анализа. Изучены методы машинного обучения, применяемые в техническом анализе рынка. Рассмотрены варианты формулирования задач прогнозирования ценовых движений, как задач машинного обучения.

На сегодняшний день наиболее популярными методами технического анализа являются цифровые фильтры и математические аппроксимации (индикаторы). Многие трейдеры используют их в совокупности с графическим подходом. Такой метод позволяет опытному человеку видеть направления движения рынка и использовать эти знания.

В задачах прогнозирования движения цен самая частая проблема, которая возникает при построении алгоритмов машинного обучения, – это диапазон входных данных. Чаще всего он не ограничен или его трудно определить. В таких случаях для нормализации данных диапазон задают либо вручную либо на основе имеющихся данных. Значения лежащие вне заданного диапазона приводятся к его границе.

В качестве входных данных для алгоритмов применяются исторические данные цен (авторегрессионные модели) и значения технических индикаторов. Также осуществляются попытки использовать информацию из неструктурированных источников, таких как финансовые новости.

Существуют два подхода к прогнозированию движения цен на рынке Форекс: прогнозирование будущих значений и прогнозирование трендов. Первый подход достаточно сложно реализовать, поскольку шумные и нестационарные. Второй подход более популярен, однако при его реализации часто упускают специфику рынка. Просто знать направление движения недостаточно, нужно быть уверенным, что изменение цены достаточно велико, чтобы покрыть комиссию и разницу в курсах покупки и продажи.

Во второй главе посвящена изучению подходов к прогнозированию временных рядов методами машинного обучения. Рассмотрены вопросы формулирования задач прогнозирования временных рядов, как задач

машинного обучения, подготовки данных, удаления трендов из временных рядов и оценки обученных моделей.

Применение методов машинного обучения к прогнозированию экономических временных рядов имеет ряд особенностей. Во-первых нельзя случайным образом отсортировать выборку, т.к. данные связаны между собой временной переменной и изменение их порядка приведет к созданию нового временного ряда. Во-вторых данные для тестирования следует брать из промежутка, который следует непосредственно за обучающим, т.к. характеристики экономических временных рядов меняются с течением времени.

Задачу прогнозирования движения цен на рынке Форекс следует рассматривать как задачу классификации направления движения цен.

При подготовке данных для задачи классификации из задачи рекомендуется удалить все глобальные и локальные тренды. Их поиск является достаточно трудной задачей, однако простая схема вычитания предыдущего значения цены из текущей позволяет убрать большинство линейных трендов.

При подготовке данных для некоторых методов машинного обучения необходима нормализация или стандартизация данных. Стандартизация данных чаще всего невозможна, поскольку экономические временные ряды редко соответствуют нормальному распределению, а если и соответствуют то лишь на отдельно взятых промежутках. Нормализация экономических временных рядов также представляет трудности, т.к. диапазон значений входных данных ограничен. Обычно диапазон значений задается вручную или формируется на основе входных данных.

Для получения представления о точке отсчета для сравнения получаемых моделей строится базовый прогноз. Он представляет из себя простой алгоритм, всегда дает одинаковые результаты на одних и тех же входных данных. Для задач классификации такой алгоритм чаще всего выдает мажоритарный класс.

В третьей главе проведены исследования применимости методов машинного обучения к задаче прогнозирования движения цен на рынке Форекс, рассмотрены методы формирования обучающей выборки из исходных ценовых колебаний, изучены результаты применения различных методов классификации на практике.

С точки зрения машинного обучения сформулированы задачи классификации результатом которых косвенно или опосредовано является прогнозирование направления движения цены.

Прогнозирование только направления движения цены не позволяет построить успешную торговую стратегию, потому что данный подход не дает представлений о том насколько значимо ценовое движение. При недостаточном ценовом движении комиссия брокера, а также спред (разница цен покупки и продажи) чаще всего полностью поглощает полученную прибыль и переходит в убыток.

При прогнозировании торгового сигнала закладывается промежуток, который покрывает спред и комиссию. Однако прогнозирование сигналов продажи на основе цен предложения не учитывает возможное расширение спреда. Т.е. даже при удачном прогнозе при закрытии сделки на продажу при увеличившемся спреде возможен убыток.

Подход к прогнозированию торгового сигнала в одну сторону, так чтобы спред учитывался только при открытии сделки, предоставляет наилучшие шансы при успешном прогнозе. Однако все еще существует шанс, что несколько неудачных прогнозов перечеркнут сотни удачных.

Результаты моделей, основанных на авторегрессии, дают довольно низкие результаты. В это же время модели, использующие технические индикаторы в качестве входных параметров, показывают более высокие результаты.

Общий уровень точности полученных моделей оставляет желать лучшего. Однако он может быть интересен только в задаче прогнозирования направления движения цены, которая мало применима в реальности. В реальности наибольший интерес представляет точность прогнозирования торгового сигнала о покупке или продаже. Получить приемлемую точность прогнозирования такого (хотя бы 0.5) весьма затруднительно. На отдельно взятых промежутках это возможно, однако в общем случае рассмотренные модели не в состоянии достичь такого результата.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

Предложен метод формирования классов для задачи прогнозирования движения цен на рынке Форекс на основе исходных данных о ценовых движениях. Данный метод позволяет использовать полученные сигналы в реальной торговле, не прибегая к иным методам оценки рынка.

Рассмотрены 2 подхода к формированию набора данных для обучения моделей в задаче прогнозирования движения цен: с использованием предыдущих значений цен и с использованием значений технических индикаторов. При использовании значений технических индикаторов построенные модели имели большую точность прогнозирования.

Исследована применимость метода опорных векторов, глубоких и сверточных нейронных сетей для задачи прогнозирования движения цен. Глубокие нейронные сети на наборе данных из значений технических индикаторов показывали наилучшую точность классификации. Однако результаты использования торговых сигналов полученных моделей не позволяют достичь прибыльности стратегии на любом временном промежутке. Это возможно лишь на небольшом количестве отдельно взятых временных промежутков.

Рекомендации по практическому использованию результатов

Предложенный метод формирования классов для задачи прогнозирования движения цен может применяться при построении набора данных для различных методов машинного обучения.

Использование значений технических индикаторов в наборе входных данных для задачи прогнозирования движения цен повышает точность классификации, по сравнению с использованием предыдущих значений ценовых движений.