

Санкт-Петербургский государственный университет
Математико - механический факультет
Кафедра информационно-аналитических систем

Курсовая работа
студента 341-й группы Евгения Хандыго

Оценка эффективности торговой стратегии

Научный руководитель — Н.Г. Графеева

Санкт-Петербург
2014

St. Petersburg State University
Mathematics and Mechanics
Department of information and analytical systems

Coursework

Eugene Khandygo, student of 341 group

Trading Strategy Effectiveness Evaluating

Scientific director — N.H. Grafeeva

St. Petersburg
2014

1 Введение

Многие не раз слышали о бирже: кино, пресса, Интернет — все они широко освещают эту тему. Тем не менее, большинство людей, как правило, считают, что биржа — это невероятно сложный механизм. С другой стороны, почти везде сейчас можно увидеть рекламу наподобие: «Зарабатывать на бирже легко!». Получается, что биржа может приносить доход, и нам совсем не обязательно знать, как она функционирует?

На самом деле, все не так просто. Обратимся к примеру: представьте, что Вы выращиваете и продаете картошку. Вы собрали очередной урожай в размере трех мешков и поехали на рынок. Пока ехали, мечтали, как продадите картошку по 50 рублей за кг., но, зайдя на базар, увидели массу продавцов, которые уже продают по 30, при этом покупателей практически нет. Что делать — нужно срочно продать товар — и Вы устанавливаете цену в 25 рублей. Начинают подходить придирчивые покупатели — оценивать вид и качество Вашей картошки. И тут, не дойдя до Вас, видят, как конкурент поставил цену в 24 рубля. Вам это не понравилось, и Вы ставите 23, конкурент — 22, 20, 15. Узнав о последней цене, покупатели рассказывают своим друзьям о потрясающе низкой стоимости картофеля на рынке. На следующий день Вы возвращаетесь на рынок и видите сумасшедших покупателей, готовых купить даже перемороженную картошку при практически полном отсутствии продавцов. Оказывается, вечером по новостям передали, что будет метель, мороз и снег, что на неделю парализуется инфраструктура города, и люди запасаются, да еще и узнали о низких ценах. В этот момент Вы становитесь главным на рынке. Вы повышаете цену до тех пор, пока остаются люди, готовые купить Ваш картофель. В итоге продаете последний килограмм по 90 рублей.

А теперь давайте задумаемся над тем, как в такой ситуации поступит умный биржевой делец. Конечно, прежде всего необходимо оценить соотношение спроса и предложения, увидеть тенденцию изменения стоимости товара и оценить новости. Имея необходимую информацию, мы делаем вывод: надо прийти под самое закрытие в первый день торгов и постараться скупить все по 15 рублей, а потом подоспеть к середине следующего дня и продать все по 70 — 80 рублей.

Есть еще один интересный вариант: можно вычислить инсайдера (человека, который заранее знает о новостях на завтра и зарабатывает практически всегда), понять, что он делает (скупает или распродает) и делать это вместе с ним. Или вычислить просто очень умного человека, который чаще других в плюсе.

Кто-то теперь может подумать: «Вот он секрет биржи!»

И снова вынужден Вас разочаровать: тут тоже все совсем не просто. Во-первых, в примере описана сильно упрощенная модель биржи. Во-вторых, большинство людей не представляет, как анализировать ситуацию на бирже.

«Как», — спросите Вы, — «все потеряно? Не видать нам биржевых дивидендов?»

Не стоит делать поспешных выводов: на бирже уже давно встречаются полностью или частично автоматизированные системы, которые торгуют вместо Вас.

В данной работе рассматривается одна из проблем построения таких биржевых роботов.

Конечно, на сегодняшний день существует множество способов построения торговых стратегий — на любой вкус и цвет. Например, если мы имеем временной ряд цены некоторого товара, то можно строить его прогнозы, тем самым получая оценку стоимости товара (с некоторой вероятностью) в будущем. Трейдер, работающий на основе такой модели должен построить прогноз на n шагов вперед, определить, насколько он точен (например, посредством построения доверительного интервала) и принять решение: продавать, покупать или выждать какое-то время. Другим примером может служить построение простейших торговых стратегий на основе статистических критериев. Имея значения временного ряда в скользящем окне, можно проверить гипотезу о наличии тренда на рассматриваемом участке. В случае принятия этой гипотезы, на основании статистики критерия можно сделать вывод о направлении роста тренда. На основе такой модели трейдер проходит по всему временному ряду скользящим окном оптимальной длины и на каждом участке

принимает решение о дальнейших действиях.

Существующие методики построения трейдеров способны демонстрировать неплохие результаты, однако, большинство из них имеют слабости:

- **Модель данных.** Если модель данных представляет собой временной ряд, то мы фактически вынуждены рассматривать зависимость цены от времени. Сложно поверить в то, что такая модель будет соответствовать действительности. Кроме того, даже на приведенном выше примере видно, что информации только о ценах недостаточно.
- **Оценка качества.** Для организации «правильного» процесса обучения, необходимо уметь оценивать, насколько хорошо отдельно взятый трейдер соответствует требованиям. В настоящее время подавляющее большинство обучающих алгоритмов для биржевых трейдеров в качестве значения фитнес-функции используют доход на обучающей выборке, который на самом деле может являться случайной величиной. [1]

2 Постановка задачи

Целью данной работы являлось создание нового способа построения фитнес-функций для оценки качества биржевых трейдеров. Новый метод должен обладать следующими свойствами:

- **Прозрачность:** простота понимания работы алгоритма и интерпретации результатов.
- **Эффективность:** новый метод должен давать качественно лучшие результаты, нежели оценка дохода на тренировочном периоде.

3 Контекст задачи

Проблема, которая рассматривается в этой работе, на самом деле является лишь частью одной большой задачи построения эффективного биржевого робота. Модель алгоритма была предложена моим научным руководителем, и именно в рамках этой модели будет решаться поставленная задача.

Алгоритм предполагает, что помимо временного ряда цены мы так же обладаем некоторой дополнительной информацией за интересующий нас период времени. Это может быть подробная информация о сделках во время торгов или, например, информация о поведении курса доллара за последний год.

Приведем основные шаги алгоритма:

1. На основе имеющихся данных необходимо выбрать индикатор, на основе которого трейдер сможет принимать решение о том, какое действие необходимо совершить.
2. Построение модели трейдера (например, нейронной сети) на основе выбранного индикатора.
3. Построение фитнес-функции.
4. Поиск параметров модели на основе фитнес-функции с применением генетического алгоритма.

Стоит сказать, что каждый из представленных шагов является темой отдельного исследования. В первом пункте пока еще не понятно, каким образом оценивать качество индикатора. Первое, что приходит на ум в отношении второго пункта, это различные классификаторы/предсказатели, однако, тут еще предстоит выяснить, какой из них наиболее легко и «прозрачно» подойдет для дальнейшей модификации генетическим алгоритмом. Возможно, и вовсе стоит обратиться к другим

моделям. Четвертый же пункт, в силу применения генетических алгоритмов, не может гарантировать оптимальности решения (большинство решений достигают лишь локального экстремума), поэтому необходимо провести исследование на предмет его модификаций.

4 Решение

Пусть наш трейдер должен удовлетворять m требованиям. Каждое требование накладывает ограничение на некоторый показатель трейдера. Например, выдвигается требование: средний доход трейдера за период (час, день, неделя или др.) должен быть равен a у.е., при этом мы хотим, чтобы трейдер был как можно более стабильным, из чего следует условие: допустимое отклонение от средней величины не должно быть слишком велико. Например, мы определяем допустимое отклонение дохода от среднего на b у.е., где $b > 0$. Пойдем дальше: рассмотрим доход как случайную величину ξ . В этом случае доход должен удовлетворять некоторому закону распределения. Какому именно — зависит от самих требований и их природы. Для примера потребуем, чтобы ξ имела нормальное распределение. Таким образом, выдвинутое требование к доходу трейдера можно записать следующим образом:

$$\xi \sim \mathcal{N}(a, b/3)^1$$

Действуя аналогичным образом, преобразуем m требований в m независимых случайных величин. Рассмотрим теперь $\bar{\xi} = (\xi_1, \dots, \xi_m)$ — многомерную случайную величину, компоненты которой независимы.

Рассмотрим ковариационную матрицу $\Sigma_{\bar{\xi}} = (\sigma_{ij})$ — аналог дисперсии для одномерной случайной величины, где

$$\sigma_{ij} = \begin{cases} \mathbb{D}\xi_i, & i = j \\ \text{cov}(\xi_i, \xi_j), & i \neq j \end{cases}, \quad i, j \in [1 : m]$$

В дополнение к ковариационной матрице можно рассмотреть матрицу корреляций, элементы которой характеризуют меру линейной зависимости между соответствующими случайными величинами. Используя корреляционную матрицу, существующие m требований можно другими l требованиями, которые будут характеризовать зависимость показателей трейдера. Например, мы допускаем увеличение риска при доходе выше нормы, значит можем задать соответствующий элемент корреляционной матрицы равным небольшому положительному числу.

Теперь, когда все $m + l$ требований формализованы, можно рассмотреть «полноценную» многомерную случайную величину $\bar{\xi}$. Будем называть ее *моделью идеальных требований*.

Перейдем к построению фитнес-функции.

Мы рассматриваем некоторую особь из поколения генетического алгоритма и хотим понять, насколько она соответствует модели идеальных требований. Протестируем этого трейдера на тренировочных данных. При этом будем фиксировать значения интересующих нас показателей в каждый момент времени. В результате получим многомерный временной ряд показателей конкретно взятого трейдера. На его основе мы можем оценить некоторые параметры соответствующих случайных величин и степень зависимости между ними: математическое ожидание, дисперсию, корреляцию и др. Используя эти данные, уже можно сделать вывод о том, что рассматриваемая особь не удовлетворяет выдвинутым требованиям. В противном случае мы можем продолжить сравнение.

Здесь я вижу два подхода:

- **Метрики.** На пространстве распределений можно задать метрику. То есть мы можем оценить соответствие особи требованиям на основе расстояния между ними.

¹В соответствии с «правилом трех сигма».

- **Статистические критерии согласия.** Можно проверить статистическую гипотезу о том, что распределение показателей рассматриваемого трейдера совпадает с моделью идеальных требований.
 - **Параметрические критерии согласия** — критерии, в которых вычисление тестовой статистики основано на параметрах распределений.
 - **Непараметрические критерии согласия** — критерии, основанные на вычислении эмпирической функции распределения выборки.

Я склоняюсь к мнению, что для построения фитнес-функции проверка статистическими критериями предпочтительнее. И дело вот в чем: результатом работы критерия является p -value — это максимальный уровень доверия, с которым выдвинутая гипотеза не отвергается, и эта величина нормирована, то есть лежит на отрезке от нуля до единицы. Таким образом, легко задать критерий того, что трейдер достаточно «хорош»: достаточно ограничить p -value снизу каким-нибудь «разумным» значением. Метрика, в свою очередь, таким свойством не обладает, поэтому придется либо подбирать оптимальный параметр, либо довольствоваться утверждением наподобие: «Этот трейдер лучше остальных».

Корректность модели требований

Возникает справедливый вопрос: возможно ли, что нашим требованиям не может удовлетворять никакой трейдер? Что ж, если рассматривать числовые характеристики требований, то в их отношении достаточно не забывать о здравом смысле. Например, миллионные доходы при мизерном стартовом капитале и минимальных рисках скорее всего недостижимы. К тому же, можно обратиться к статистике торгов и попытаться вычислить, кто сколько заработал, а на основе этой информации составлять требования к стратегии. Если никакие подходы не помогают определить необходимые константы, я полагаю, стоит обратиться к экономистам.

А как же быть с распределениями интересующих нас показателей? На первый взгляд, этот вопрос может показаться сложно разрешимым: действительно, вряд ли кто-то задумывается над тем, как должны быть распределены его доходы при торговле на бирже. Однако, при ближайшем рассмотрении все становится проще.

Во-первых, необходимо понимать природу требований. Например, доход трейдера на всяком участке зависит от множества малозначимых не взаимосвязанных факторов, и, следовательно, должен быть распределен нормально (в соответствии с природой нормального распределения²). Можно привести и другой пример: понятно, что идеальная кривая роста дохода трейдера должна быть прямой. Можно рассматривать квадратичное отклонение реальной кривой доходности от идеальной, нормировать его и наложить на результат некоторое требование. Такая случайная величина должна быть распределена как хи-квадрат, исходя из определения.

Во-вторых, необходимо учитывать специфику ограничиваемого показателя. Например, доход может быть как положительным, так и отрицательным. Соответственно, некорректно требовать, чтобы он был распределен в соответствии с, например, логнормальным законом (поскольку логнормально распределенная случайная величина всегда положительна).

В третьих, можно использовать различные специфики самих распределений: симметричность, скорость роста, положительность и др.

²Одним из условий выполнения центральной предельной теоремы является условие Линдберга[2], которое и позволяет интерпретировать нормальное распределение как сумму большого числа малозначимых независимых факторов.

Сложность вычислений

Следующая проблема, которую следует обсудить — это вопрос о сложности вычислений: пусть построенная фитнес-функция работает корректно, но не будут ли затраты на ее вычисление слишком велики?

Для ответа на этот вопрос, разобьем процесс работы фитнес-функции на этапы и попробуем оценить каждый из них.

1. Тестирование трейдера.
2. Сбор статистики, вычисление ее характеристик.
3. Сравнение полученной выборки с моделью идеальных требований.

Заметим, что тестирование трейдера и вычисление характеристик выборки можно объединить в одну итерацию. Например, для вычисления выборочных средних, необходимо накапливать сумму, а в конце просто разделить ее на число наблюдений. Таким образом, первые два пункта не увеличивают сложность.

Сравнение производится на основе некоторого статистического критерия, который также можно разбить две фазы:

1. Вычисление тестовой статистики.
2. Сравнение полученной величины с табличным значением.

Таким образом, сложность может повыситься только на этапе вычисления статистики критерия.

Для примера приведем две тестовых статистики непараметрических критериев ([3]):

- **Критерий согласия Ватсона.** Проверяется гипотеза вида $H_0: F_n(x) = F(x, \theta)$. Используется следующая статистика:

$$U_n^2 = \sum_{i=1}^n \left(F(x_i, \theta) - \frac{i - 0.5}{n} \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^n F(x_i, \theta) - 0.5 \right)^2 + \frac{1}{12n} ,$$

где n — объем выборки, $x_1, \dots, x_n$ — упорядоченные по возрастанию элементы выборки.

- **Критерий согласия Купера.** Проверяется гипотеза вида $H_0: F_n(x) = F(x, \theta)$. Используется следующая статистика:

$$V_n = D_n^+ + D_n^- .$$

Здесь:

$$D_n^+ = \max \left(\frac{i}{n} - F(x_i, \theta) \right) ,$$

$$D_n^- = \max \left(F(x_i, \theta) - \frac{i - 1}{n} \right) ,$$

где $i \in [1 : n]$, n — объем выборки, $x_1, \dots, x_n$ — упорядоченные по возрастанию элементы выборки.

На приведенных примерах видно, что вычисление соответствующих статистик имеет сложность $O(n \log n)$ (в силу необходимости сортировки выборки). Конечно, можно использовать и другие критерии согласия. Однако, следует заметить, что все непараметрические критерии согласия будут иметь ту же сложность (в силу того, что вычисление эмпирической функции распределения при отсортированной по возрастанию элементов выборке будет иметь сложность $O(n)$).

5 Дополнения

Полученное представление трейдера позволяет модифицировать алгоритм построения биржевых роботов в целом.

- **Комплексные требования.** Можно строить трейдеров, которые удовлетворяют достаточно сложным требованиям.
- **Развитие трейдера в процессе его работы.** Описанный процесс позволяет подстраиваться под изменения ситуации на бирже в режиме реального времени. Представим, что в наше хранилище данных непрерывно поступает новая информация. После того, как в тренировочные (устаревшие) данные добавлен новый блок (достаточно большой сегмент данных, чтобы оправдать перерасчет), мы запускаем трейдера только на новых данных. Перерасчет статистики трейдера не требуется начинать заново, поскольку нас интересует статистика внутри скользящего окна. Единственное, что требуется пересчитать — это статистику критерия (или расстояние). Таким образом, мы можем вовремя понять, что наш трейдер перестает удовлетворять требованиям в полной мере. Тогда нам надо лишь продолжить прерванный процесс генерации. В силу того, что трейдеры из последнего существующего поколения находятся ближе к результату, этот процесс займет существенно меньшее количество времени.
- **Кластеризация.** Расстояния на пространстве распределений позволяют проводить кластеризацию трейдеров. В контексте генетических алгоритмов это означает, что мы можем разбить всех трейдеров на некоторые семейства. Понятно, что, скрещивать между собой особей из одного семейства — не самая лучшая идея. А вот скрещивание разных семейств должно дать неплохой результат. Здесь можно провести аналогию с точками на плоскости: если у нас есть скопления точек, близких к некоторым точкам на окружности, то скрещивание точек из таких семейств (сумма координат пополам) даст точку на хорде (или близкую к ней), что ближе к центру круга. Таким образом, в какой-то момент мы получим семейство точек, которое лежит в окрестности центра. Конечно, в случае распределений все может оказаться не так хорошо. Но это косвенно дает надежду на разрешение проблемы локальности решений.

Список литературы

- [1] Lars Kestner Quantitative Trading Strategies — 1st edition — McGraw-Hill, 2003.
- [2] А. Н. Ширяев Вероятность. — Том 1. — М., 2004. — с. 421 – 430.
- [3] Лемешко Б. Ю., Горбунова А. А. О применении и мощности непараметрических критериев согласия Купера, Ватсона и Жанга — Измерительная техника. 2013. № 5. — с. 3 – 9.