

Предсказание финансовых временных рядов

Что обеспечивает доходность бизнеса, основанного на предсказаниях? Какова методика предсказания временных рядов? Специфика финансовых временных рядов. Как подбирать признаковое пространство? Какой функционал ошибки лучше? Норма прибыли нейросетевой игры на реальных данных.

📖 И будущего, конечно, не знают ни Ангелы Божии, ни демоны; однако, они предсказывают.

Св. Иоанн Дамаскин, Точное изложение православной веры

📖 Я реагирую на события на рынке, как животное реагирует на то, что происходит в джунглях.

Д.Сорос, Алхимия финансов

Введение: Предсказание как вид бизнеса

В этой главе рассмотрено одно из самых популярных практических приложений нейросетей - предсказание рыночных временных рядов. В этой области предсказания наиболее тесно связаны с доходностью, и могут рассматриваться как один из видов бизнеса.

Кому нужно предсказывать рынок?

Предсказание финансовых временных рядов - необходимый элемент любой инвестиционной деятельности. Сама идея *инвестиций* - вложения денег сейчас с целью получения дохода в будущем - основывается на идее прогнозирования будущего. Соответственно, предсказание финансовых временных рядов лежит в основе деятельности всей индустрии инвестиций - всех бирж и небиржевых систем торговли ценными бумагами.

Приведем несколько цифр, иллюстрирующих масштаб этой индустрии предсказаний (Шарп, 1997). Дневной оборот рынка акций только в США превышает \$10 млрд. Депозитарий DTC (Depository Trust Company) в США, где зарегистрировано ценных бумаг на сумму \$11 трлн (из общего объема \$18 трлн), регистрирует в день сделок примерно на \$250 млрд. Еще более активно идет торговля на мировом валютном рынке FOREX. Его дневной оборот превышает \$1000 млрд. Это примерно 1/50 всего совокупного капитала человечества.

Известно, что 99% всех сделок - спекулятивные, т.е. направлены не на обслуживание реального товарооборота, а заключены с целью извлечения прибыли по схеме "купил дешевле - продал дороже". Все они основаны на предсказаниях изменения курса участниками сделки. Причем, что

немаловажно, предсказания участников каждой сделки противоположны друг другу.¹ Так что объем спекулятивных операций характеризует степень различий в предсказаниях участников рынка, т.е. реально - степень непредсказуемости финансовых временных рядов.

Можно ли предсказывать рынок?

Это важнейшее свойство рыночных временных рядов легло в основу теории "*эффективного*" рынка, изложенной в диссертации Луи де Башелье (L.Bachelier) в 1900 г. Согласно этой доктрине, инвестор может надеяться лишь на среднюю доходность рынка, оцениваемую с помощью индексов, таких как Dow Jones или S&P500 для Нью-Йоркской биржи. Всякий же спекулятивный доход носит случайный характер и подобен азартной игре на деньги. В основе непредсказуемости рыночных кривых лежит та же причина, по которой деньги редко валяются на земле в людных местах: слишком много желающих их поднять.

Теория эффективного рынка не разделяется, вполне естественно, самими участниками рынка (которые как раз и заняты поиском "упавших" денег). Большинство из них уверено, что рыночные временные ряды, несмотря на кажущуюся стохастичность, полны скрытых закономерностей, т.е. в принципе хотя бы частично предсказуемы. Такие скрытые эмпирические закономерности пытался выявить в 30-х годах в серии своих статей основатель технического анализа Эллиот (R.Elliott).

В 80-х годах неожиданную поддержку эта точка зрения нашла в незадолго до этого появившейся теории динамического хаоса. Эта теория построена на противопоставлении хаотичности и стохастичности (случайности). Хаотические ряды только выглядят случайными, но, как детерминированный динамический процесс, вполне допускают краткосрочное прогнозирование. Область возможных предсказаний ограничена по времени *горизонтом прогнозирования*, но этого может оказаться достаточно для получения реального дохода от предсказаний (Chorafas, 1994). И тот, кто обладает лучшими математическими методами извлечения закономерностей из зашумленных хаотических рядов, может надеяться на большую норму прибыли - за счет своих менее оснащенных собратьев.

В этой главе мы приведем конкретные факты, подтверждающие частичную предсказуемость финансовых временных рядов, и даже оценим эту предсказуемость численно.

Технический анализ и нейронные сети

В последнее десятилетие наблюдается устойчивый рост популярности технического анализа - набора эмпирических правил, основанных на различного рода индикаторах поведения рынка. Технический анализ сосредотачивается на индивидуальном поведении данного финансового инструмента, вне его связи с остальными ценными бумагами (Pring, 1991).

Такой подход психологически обоснован сосредоточенностью брокеров именно на том инструменте, с которым они в данный момент работают. Согласно Александру Элдеру (A.Elder), известному специалисту по техническому анализу (по своей предыдущей специальности - психотерапевту), поведение рыночного сообщества имеет много аналогий с поведением толпы, характеризующимся особыми законами массовой психологии. Влияние толпы упрощает мышление, нивелирует индивидуальные особенности и рождает формы коллективного,

¹ Здесь мы несколько упрощаем ситуацию, забывая что участники сделки могут ориентироваться на разные временные масштабы цикла купли-продажи. Это, однако, не меняет кардинально общий вывод о сложности финансовых предсказаний.

стадного поведения, более примитивного, чем индивидуальное. В частности, стадные инстинкты повышают роль лидера, вожака. Ценовая кривая, по Элдеру, как раз и является таким лидером, фокусируя на себе коллективное сознание рынка. Такая психологическая трактовка поведения рыночной цены обосновывает применение теории динамического хаоса. Частичная предсказуемость рынка обусловлена относительно примитивным коллективным поведением игроков, которые образуют единую хаотическую динамическую систему с относительно небольшим числом внутренних степеней свободы.

Согласно этой доктрине, для предсказания рыночных кривых необходимо освободиться от власти толпы, стать выше и умнее ее. Для этого предлагается выработать *систему игры*, апробированную на прошлом поведении временного ряда и четко следовать этой системе, не поддаваясь влиянию эмоций и циркулирующих вокруг данного рынка слухов. Иными словами, предсказания должны быть основаны на *алгоритме*, т.е. их можно и даже должно порекомендовать компьютеру (LeBeau, 1992). За человеком остается лишь создание этого алгоритма, для чего в его распоряжении имеются многочисленные программные продукты, облегчающие разработку и дальнейшее сопровождение компьютерных стратегий на базе инструментария технического анализа.

Следуя этой логике, почему бы не использовать компьютер и на этапе разработки стратегии, причем не в качестве ассистента, рассчитывающего известные рыночные индикаторы и тестирующего заданные стратегии, а для извлечения *оптимальных* индикаторов и нахождения *оптимальных* стратегий по найденным индикаторам. Такой подход - с привлечением технологии нейронных сетей - завоевывает с начала 90-х годов все больше приверженцев (Beltratti, 1995, Бэстенс, 1997), т.к. обладает рядом неоспоримых достоинств.

Во-первых, нейросетевой анализ, в отличие от технического, не предполагает никаких ограничений на характер входной информации. Это могут быть как индикаторы данного временного ряда, так и сведения о поведении других рыночных инструментов. Недаром нейросети активно используют именно институциональные инвесторы (например, крупные пенсионные фонды), работающие с большими портфелями, для которых особенно важны корреляции между различными рынками.

Во-вторых, в отличие от теханализа, основанного на общих рекомендациях, нейросети способны находить оптимальные для *данного* инструмента индикаторы и строить по ним оптимальную опять же для *данного* ряда стратегию предсказания. Более того, эти стратегии могут быть *адаптивными*, меняясь вместе с рынком, что особенно важно для молодых активно развивающихся рынков, в частности, российского.

Нейросетевое моделирование в чистом виде базируется лишь на данных, не привлекая никаких априорных соображений. В этом его сила и одновременно - его ахиллесова пята. Имеющихся данных может не хватить для обучения, размерность потенциальных входов может оказаться слишком велика. Далее в этой главе мы покажем как для преодоления этих типичных в области финансовых предсказаний трудностей можно воспользоваться опытом, накопленным техническим анализом.

Методика предсказания временных рядов

Для начала обрисуем общую схему нейросетевого предсказания временных рядов (Рисунок 1).

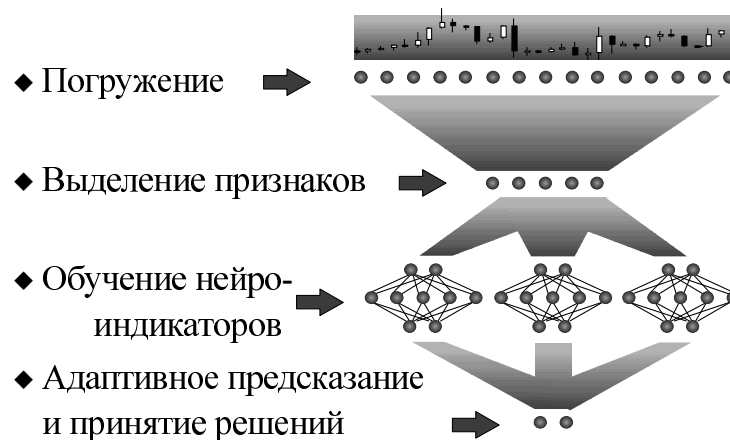


Рисунок 1. Схема технологического цикла предсказаний рыночных временных рядов

Далее в этой главе мы кратко обсудим все этапы этой технологической цепочки. Хотя общие принципы нейро-моделирования применимы к данной задаче в полном объеме, предсказание финансовых временных рядов имеет свою специфику. Именно эти отличительные черты и будут в большей мере затронуты в этой главе.

Метод погружения. Теорема Такенса

Начнем с этапа погружения. Как мы сейчас убедимся, несмотря на то, что предсказания, казалось бы, являются *экстраполяцией* данных, нейросети, на самом деле, решают задачу *интерполяции*, что существенно повышает надежность решения. Предсказание временного ряда сводится к типовой задаче нейроанализа - аппроксимации функции многих переменных по заданному набору примеров - с помощью процедуры *погружения* ряда в многомерное пространство (Weigend, 1994). Например, d -мерное *лаговое* пространство ряда X_t состоит из d значений ряда в последовательные моменты времени: $\mathbf{X}_{t-d} = (X_{t-1}, \dots, X_{t-d})$.

Для динамических систем доказана следующая *теорема Такенса*. Если временной ряд порождается динамической системой, т.е. значения X_t есть произвольная функция состояния такой системы, существует такая *глубина погружения* d (примерно равная эффективному числу степеней свободы данной динамической системы), которая обеспечивает *однозначное* предсказание следующего значения временного ряда (Sauer, 1991). Таким образом, выбрав достаточно большое d можно гарантировать однозначную зависимость будущего значения ряда от его d предыдущих значений: $X_t = f(\mathbf{X}_{t-d})$, т.е. предсказание временного ряда сводится к задаче интерполяции функции многих переменных. Нейросеть далее можно использовать для восстановления этой неизвестной функции по набору примеров, заданных историей данного временного ряда.

Напротив, для случайного ряда знание прошлого ничего не дает для предсказания будущего. Поэтому, согласно теории эффективного рынка, разброс предсказываемых значений ряда на следующем шаге при погружении в лаговое пространство не изменится.

Отличие хаотической динамики от стохастической (случайной), проявляющееся в процессе погружения, иллюстрирует Рисунок 2.

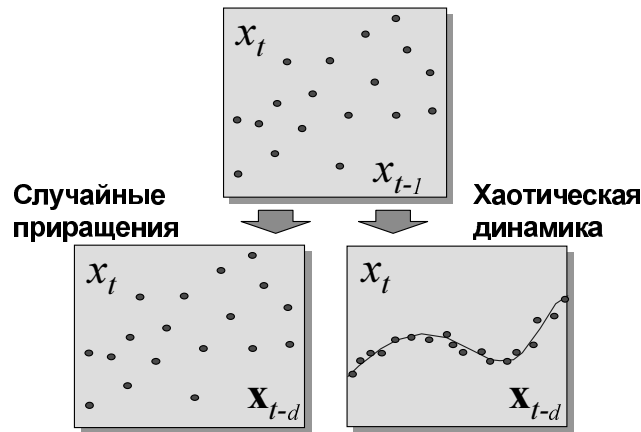


Рисунок 2. Проявляющееся по мере погружения ряда различие между случайным процессом и хаотической динамикой

Эмпирические свидетельства предсказуемости финансовых рядов

Метод погружения позволяет количественно измерить предсказуемость реальных финансовых инструментов, т.е. проверить или опровергнуть гипотезу эффективности рынка. Согласно последней, разброс точек по всем координатам лагового пространства одинаков (если они - одинаково распределенные независимые случайные величины). Напротив, хаотическая динамика, обеспечивающая определенную предсказуемость, должна приводить к тому, что наблюдения будут группироваться вблизи некоторой гиперповерхности $X_t = f(\mathbf{X}_{t-d})$, т.е. экспериментальная выборка формирует некоторое множество размерности меньшей, чем размерность всего лагового пространства.

Для измерения размерности можно воспользоваться следующим интуитивно понятным свойством: если множество имеет размерность D , то при разбиении его на все более мелкие покрытия кубиками со стороной ε , число таких кубиков растет как ε^{-D} . На этом факте основывается определение размерности множеств уже знакомым нам методом box-counting. Размерность множества точек определяется по скорости возрастания числа ячеек (boxes), содержащих все точки множества.² Для ускорения алгоритма размеры ε берут кратными 2, т.е. масштаб разрешения измеряется в битах.

В качестве примера типичного рыночного временного ряда возьмем такой известный финансовый инструмент, как индекс котировок акций 500 крупнейших компаний США, S&P500, отражающий среднюю динамику цен на Нью-Йоркской бирже. Рисунок 3 показывает динамику индекса на протяжении 679 месяцев. Размерность (информационная) приращений этого ряда, подсчитанная методом box-counting, показана на следующем рисунке (Рисунок 4).

² Вообще говоря можно определить целое семейство т.н. размерностей Реньи: $D_q = \log \sum_i p_i^q / (1-q) \log \varepsilon$,

определяемых через относительные числа заполнения ячеек p_i . Причем D_0 называют размерностью Хаусдорфа, D_1 - информационной, а D_2 - корреляционной размерностями. Чем выше степень q , тем меньше эффективное число ячеек, и соответственно - тем меньше размерность. Для сопоставимости с другими применениями box-counting в данных лекциях мы будем опираться на информационную размерность D_1 , численно равную энтропии, деленной на число бит данной степени разрешения ε .

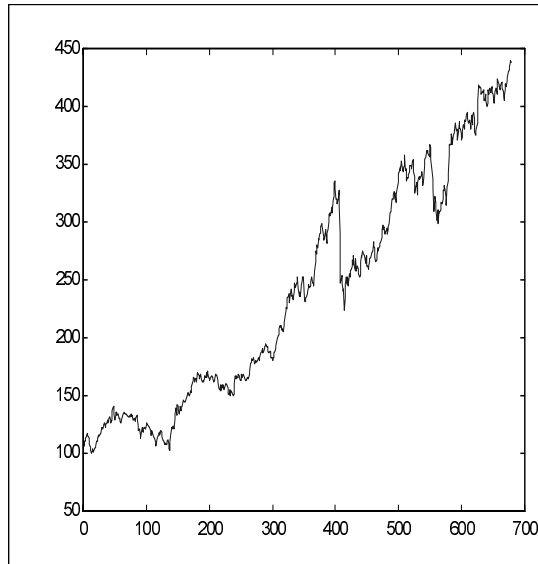


Рисунок 3. Временной ряд 679 значений индекса S&P500, используемый на протяжении данной главы в качестве примера

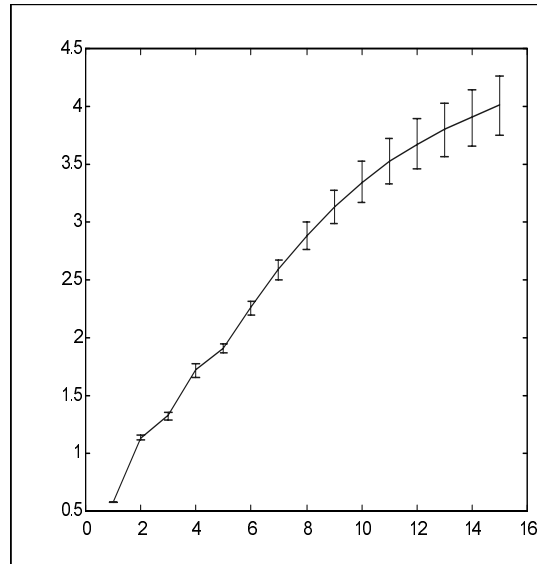


Рисунок 4. Информационная размерность приращений ряда S&P500

Как следует из последнего рисунка, в 15-мерном пространстве погружения экспериментальные точки формируют множество размерности примерно 4. Это значительно меньше, чем 15, что мы получили бы исходя из гипотезы эффективного рынка, считающей ряд приращений *независимыми* случайными величинами.

Таким образом, эмпирические данные убедительно свидетельствуют о наличии некоторой предсказуемой составляющей в финансовых временных рядах, хотя здесь и нельзя говорить о полностью детерминированной хаотической динамике³. Значит попытки применения нейросетевого анализа для предсказания рынков имеют под собой веские основания.

Заметим, однако, что теоретическая предсказуемость вовсе не гарантирует достижимость практически значимого уровня предсказаний. Количественную оценку предсказуемости конкретных рядов дает измерение *кросс-энтропии*, также возможное с помощью методики box-counting. Для примера приведем измерения предсказуемости приращений индекса S&P500 в зависимости от глубины погружения.⁴ Кросс-энтропия $I(X_t, \mathbf{X}_{t-d}) = H(X_t) - H(X_t | \mathbf{X}_{t-d})$, график которой приведен ниже (Рисунок 5), измеряет дополнительную информацию о следующем значении ряда, обеспеченную знанием d прошлых значений этого ряда.

³ График размерностей с ростом глубины погружения не выходит на насыщение, которое зафиксировало бы существование замкнутой детерминистической системы

⁴ Предсказуемость подсчитывалась методом box-counting с использованием искусственных примеров, как это будет описано чуть ниже.

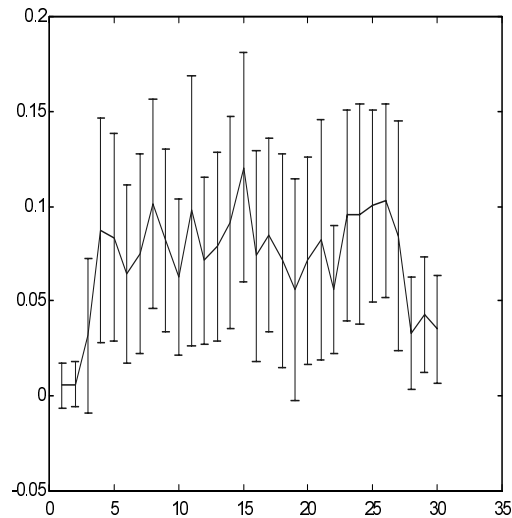


Рисунок 5. Предсказуемость знака приращений ряда индекса S&P500 в зависимости от глубины погружения (ширины "окна"). Увеличение глубины погружения свыше 25 сопровождается снижением предсказуемости.

Далее в этой главе мы оценим какой доход в принципе способен обеспечить такой уровень предсказуемости.

Формирование входного пространства признаков

Как иллюстрирует Рисунок 5, увеличение ширины окна погружения ряда приводит в конце концов к понижению предсказуемости - когда повышение размерности входов уже не компенсируется увеличением их информативности. В этом случае, когда размерность лагового пространства d слишком велика для данного количества примеров, приходится применять специальные методики формирования *пространства признаков* с меньшей размерностью. Способы выбора признаков и/или увеличения числа доступных примеров, специфичные для финансовых временных рядов будут описаны ниже.

Выбор функционала ошибки

Для обучения нейросети недостаточно сформировать обучающие наборы входов-выходов. Необходимо также определить ошибку предсказаний сети. Среднеквадратичная ошибка, используемая по умолчанию в большинстве нейросетевых приложений, не имеет большого "финансового смысла" для рыночных рядов. Поэтому в отдельном разделе мы рассмотрим специфичные для финансовых временных рядов функции ошибки и покажем их связь с возможной нормой прибыли.

Например, для выбора рыночной позиции надежное определение знака изменения курса более важно, чем понижение среднеквадратичного отклонения. Хотя эти показатели и связаны между собой, сети оптимизированные по одному из них будут давать худшие предсказания другого. Выбор адекватной функции ошибки, как мы покажем далее в этой главе, должен опираться на

некую идеальную стратегию и диктоваться, например, максимизацией прибыли (или минимизацией возможных убытков).

Обучение нейросетей

Основная специфика предсказания временных рядов лежит в области предобработки данных. Процедура обучения отдельных нейросетей стандартна. Как всегда, имеющиеся примеры разбиваются на три выборки: *обучающая*, *валидационная* и *тестовая*. Первая используется для обучения, вторая - для выбора оптимальной архитектуры сети и/или для выбора момента остановки обучения. Наконец, третья, которая вообще не использовалась в обучении, служит для контроля качества прогноза обученной нейросети.

Однако, для сильно зашумленных финансовых рядов существенный выигрыш в надежности предсказаний способно дать использование *комитетов* сетей. Обсуждением этой методики мы и закончим данную главу.

В литературе имеются свидетельства улучшения качества предсказаний за счет использования нейросетей с обратными связями. Такие сети могут обладать локальной памятью, сохраняющей информацию о более далеком прошлом, чем то, что в явном виде присутствует во входах. Рассмотрение таких архитектур, однако, увело бы нас слишком далеко от основной темы, тем более, что существуют альтернативные способы эффективного расширения "горизонта" сети, за счет специальных способов погружения ряда, рассмотренных ниже.

Формирование пространства признаков

Ключевым для повышения качества предсказаний является эффективное кодирование входной информации. Это особенно важно для труднопредсказуемых финансовых временных рядов. Все рекомендации, описанные в главе о предобработке данных, применимы и здесь. Имеются, однако, и специфичные именно для финансовых временных рядов способы предобработки данных, на которых мы подробно остановимся в данном разделе.

Способы погружения временного ряда

Начнем с того, что в качестве входов и выходов нейросети *не* следует выбирать сами значения котировок, которые мы обозначим C_t . Действительно значимыми для предсказаний являются *изменения* котировок. Поскольку эти изменения, как правило, гораздо меньше по амплитуде, чем сами котировки, между последовательными значениями курсов имеется большая корреляция - наиболее вероятное значение курса в следующий момент равно его предыдущему значению: $\langle C_{t+1} \rangle = C_t + \langle \Delta C_t \rangle = C_t$. Между тем, как это уже неоднократно подчеркивалось, для повышения качества обучения следует стремиться к статистической независимости входов, то есть к отсутствию подобных корреляций.

Поэтому в качестве входных переменных логично выбирать наиболее статистически независимые величины, например, изменения котировок ΔC_t или логарифм относительного приращения $\log(C_t / C_{t-1}) \approx \Delta C_t / C_{t-1}$. Последний выбор хорош для длительных временных рядов, когда уже заметно влияние инфляции. В этом случае простые разности в разных частях ряда будут иметь различную амплитуду, т.к. фактически измеряются в различных единицах. Напротив, отношения последовательных котировок не зависят от единиц измерения, и будут

одного масштаба несмотря на инфляционное изменение единиц измерения. В итоге, большая стационарность ряда позволит использовать для обучения большую историю и обеспечит лучшее обучение.

Отрицательной чертой погружения в лаговое пространство является ограниченный "кругозор" сети. Технический анализ же, напротив, не фиксирует *окно* в прошлом, и пользуется подчас весьма далекими значениями ряда. Например, утверждается, что максимальные и минимальные значения ряда даже в относительно далеком прошлом оказывают достаточно сильное воздействие на психологию игроков, и, следовательно, должны быть значимы для предсказания. Недостаточно широкое окно погружения в лаговое пространство не способно предоставить такую информацию, что, естественно, снижает эффективность предсказания. С другой стороны, расширение окна до таких значений, когда захватываются далекие экстремальные значения ряда, повышает размерность сети, что в свою очередь приводит к понижению точности нейросетевого предсказания - уже из-за разрастания размера сети.⁵

Выходом из этой, казалось бы, тупиковой ситуации являются альтернативные способы кодирования прошлого поведения ряда. Интуитивно понятно, что чем дальше в прошлое уходит история ряда, тем меньше деталей его поведения влияет на результат предсказаний. Это обосновано психологией субъективного восприятия прошлого участниками торгов, которые, собственно, и формируют будущее. Следовательно, надо найти такое представление динамики ряда, которое имело бы избирательную точность: чем дальше в прошлое - тем меньше деталей, при сохранении общего вида кривой. Весьма перспективным инструментом здесь может оказаться т.н. вейвлетное разложение (wavelet decomposition). Оно эквивалентно по информативности лаговому погружению, но легче допускает такое сжатие информации, которое описывает прошлое с избирательной точностью.

Понижение размерности входов: признаки

Подобного рода сжатие информации является примером извлечения из непомерно большого числа входных переменных наиболее значимых для предсказания *признаков*. Способы систематического извлечения признаков уже были описаны в прошлых главах. Их можно (и нужно) с успехом применять и к предсказанию временных рядов.

Важно только, чтобы способ представления входной информации по возможности облегчал процесс извлечения признаков. Вейвлетное представление является примером удачного, с точки зрения извлечения признаков, кодирования (Kaiser, 1995). Например, на следующем рисунке (Рисунок 6) изображен отрезок из 50 значений ряда вместе с его реконструкцией по 10 специальным образом отобранным вейвлет-коэффициентам. Обратите внимание, что несмотря на то, что для этого потребовалось в пять раз меньше данных, непосредственное прошлое ряда восстановлено точно, а более далекое - лишь в общих чертах, хотя максимумы и минимумы отражены верно. Следовательно, можно с приемлемой точностью описывать 50-мерное окно всего лишь 10-мерным входным вектором.

⁵ Напомним, что согласно нашим прошлым оценкам минимальная ошибка нейропредсказаний возрастает с ростом числа входов: $\varepsilon \geq \sqrt{d/P}$, где P - число примеров.

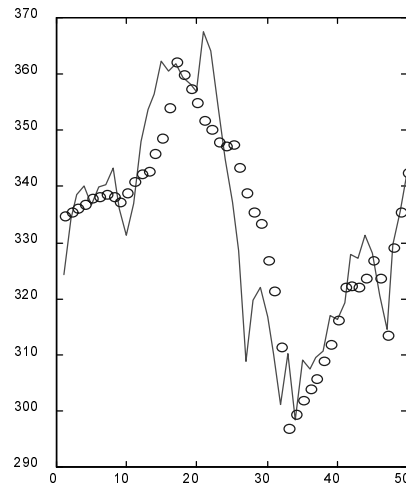


Рисунок 6. Пример 50-мерного окна (сплошная линия) и его реконструкции по 10 вейвлет-коэффициентам (о).

Еще один возможный подход — использование в качестве возможных кандидатов в пространство признаков различного рода индикаторов технического анализа, которые автоматически подсчитываются в соответствующих программных пакетах (таких как MetaStock или Windows On Wall Street). Многочисленность этих эмпирических признаков (Colby, 1988) затрудняет пользование ими, тогда как каждый из них может оказаться полезным в применении к данному ряду. Описанные ранее методы позволят отобрать наиболее значимую комбинацию технических индикаторов, которую и следует затем использовать в качестве входов нейросети.

Метод искусственных примеров (hints)

Одним из самых "больных мест" в финансовых предсказаниях является дефицит примеров для обучения нейросети. Финансовые рынки, вообще говоря, не стационарны (особенно российские). Появляются новые финансовые инструменты, для которых еще не накоплена история, изменяется характер торговли на прежних рынках. В этих условиях длина доступных для обучения нейросети временных рядов весьма ограничена.

Однако, можно повысить число примеров, используя для этого те или иные априорные соображения об *инвариантах* динамики временного ряда. Это еще одно физико-математическое понятие, способное значительно улучшить качество финансовых предсказаний. Речь идет о генерации искусственных примеров, получаемых из уже имеющихся применением к ним различного рода преобразований.

Поясним основную мысль на примере. Психологически оправдано следующее предположение: игроки обращают внимание, в основном, на форму кривой цен, а не на конкретные значения по осям. Поэтому если немного растянуть по оси котировок весь временной ряд, то полученный в результате такого преобразования ряд также можно использовать для обучения наряду с исходным. Мы, таким образом, удвоили число примеров за счет использования априорной информации, вытекающей из психологических особенностей восприятия временных рядов участниками рынка.⁶ Более того, мы не просто увеличили число примеров, но и ограничили класс функций, среди которых ищется решение, что также повышает качество предсказаний (если, конечно, использованный инвариант соответствует действительности).

⁶ Естественно, таким образом можно не только удвоить число примеров, но и учетверить их и т.д.

Приведенные ниже результаты вычисления предсказуемости индекса S&P500 методом box-counting (см. Рисунок 7, Рисунок 8) иллюстрируют роль искусственных примеров. Пространство признаков в данном случае формировалось методом ортогонализации, описанным в главе о способах предобработки данных. В качестве входных переменных использовались 30 главных компонент в 100-мерном лаговом пространстве. Из этих главных компонент были выбраны 7 признаков - наиболее значимые ортогональные линейные комбинации. Как видно из этих рисунков, лишь применение искусственных примеров оказалось способным в данном случае обеспечить заметную предсказуемость.

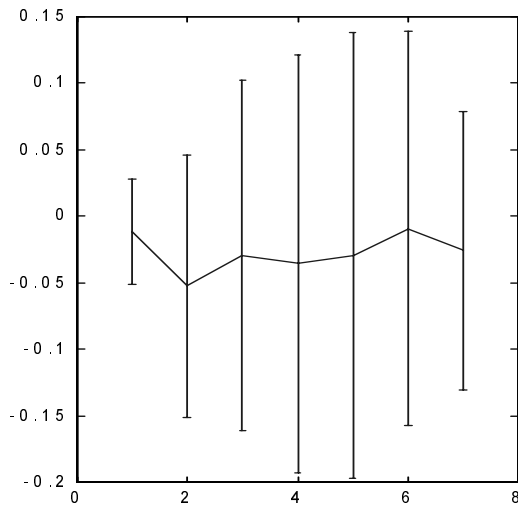


Рисунок 7. Предсказуемость знака изменения котировок индекса S&P500

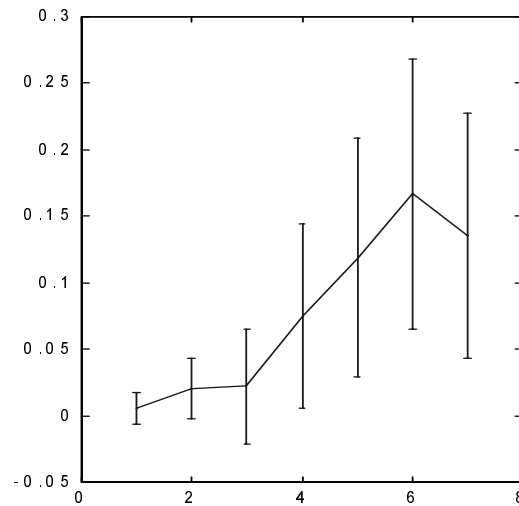


Рисунок 8. Предсказуемость знака изменения котировок индекса S&P500 после учетверения числа примеров методом растяжения по оси цен

Обратите внимание, что использование ортогонального пространства признаков привело к некоторому повышению предсказуемости по сравнению с обычным способом погружения: с 0.12 бит (Рисунок 5) до 0.17 бит (Рисунок 8). Чуть позже, когда пойдет речь о влиянии предсказуемости на прибыль, мы покажем, что за счет этого норма прибыли может увеличиться почти в полтора раза.

Другой, менее тривиальный, пример удачного использования такого рода подсказок (hints) для нейросети в каком направлении искать решение - использование скрытой симметрии в валютной торговле. Смысл этой симметрии в том, что валютные котировки могут рассматриваться с двух "точек зрения", например как ряд DM/\$ или как ряд \$/DM. Возрастание одного из них соответствует уменьшению другого. Это свойство можно использовать для удвоения числа примеров: каждому примеру вида $(X_{t-d+1}, \dots, X_{t-1}, X_t) \rightarrow X_{t+1}$ можно добавить его симметричный аналог $(-X_{t-d+1}, \dots, -X_{t-1}, -X_t) \rightarrow -X_{t+1}$. Эксперименты по нейросетевому предсказанию показали, что для основных валютных рынков учет симметрии поднимает норму прибыли примерно в два раза, конкретно - с 5% годовых до 10% годовых, с учетом реальных транзакционных издержек (Abu-Mostafa, 1995).

Измерение качества предсказаний

Хотя предсказание финансовых рядов и сводится к задаче аппроксимации многомерной функции, оно имеет свои особенности как при формировании входов, так и при выборе выходов нейросети. Первый аспект, касающийся входов, мы уже обсудили. Теперь коснемся особенностей выбора выходных переменных. Но прежде ответим на главный вопрос: как измерить качество финансовых предсказаний. Это поможет определить наилучшую стратегию обучения нейросети.

Связь предсказуемости с нормой прибыли

Особенностью предсказания финансовых временных рядов является стремление к получению максимальной прибыли, а не минимизации среднеквадратичного отклонения, как это принято в случае аппроксимации функций.

В простейшем случае ежедневной торговли прибыль зависит от верно угаданого знака изменения котировки. Поэтому нейросеть нужно ориентировать именно на точность угадывания знака, а не самого значения. Найдем как связана норма прибыли с точностью определения знака в простейшей постановке ежедневного вхождения в рынок (Рисунок 9).

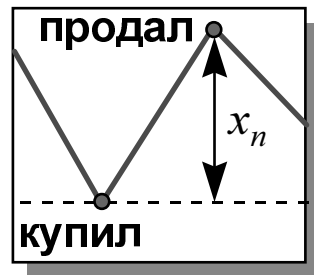


Рисунок 9. Ежедневное вхождение в рынок

Обозначим на момент t : полный капитал игрока K_t , относительное изменение котировки $x_t = \Delta C_t / C_t$, а в качестве выхода сети возьмем степень ее уверенности в знаке этого изменения $y_t \in [-1, 1]$. Такая сеть с выходной нелинейностью вида $y = \tanh(a)$ обучается предсказывать знак изменения и выдает прогноз знака с амплитудой пропорциональной его вероятности. Тогда возрастание капитала на шаге t запишется в виде:

$$K_t = K_{t-1} [1 + |x_t| \delta \operatorname{sgn}(x_t y_t)]$$

где δ - доля капитала, "в игре". Выигрыш за все время игры:

$$K_t = K_0 \exp \left(\sum_{k=1}^t \ln [1 + x_k \delta \operatorname{sgn}(y_k)] \right)$$

нам и предстоит максимизировать, выбрав оптимальный размер ставок δ . Пусть в среднем игрок угадывает долю $p = \frac{1}{2} + \varepsilon$ знаков и, соответственно, ошибается с вероятностью $q = \frac{1}{2} - \varepsilon$. Тогда логарифм нормы прибыли,

$$\langle \ln(K_t / K_0) \rangle = t \langle p \ln(1 + |x|\delta) + q \ln(1 - |x|\delta) \rangle,$$

а следовательно и сама прибыль, будет максимальным при значении $\delta = (p - q) \langle |x| \rangle / \langle x^2 \rangle$ и составит в среднем:

$$\langle \ln(K_t / K_0) \rangle \approx t (p - q)^2 \frac{\langle |x| \rangle^2}{2 \langle x^2 \rangle} = 2at\varepsilon^2.$$

Здесь мы ввели коэффициент $a = \langle |x| \rangle^2 / \langle x^2 \rangle \leq 1$. Например, для Гауссова распределения $a \approx 0.8$. Степень предсказуемости знака напрямую связана с кросс-энтропией, которую можно оценить *a priori* методом box-counting. Для бинарного выхода (см. Рисунок 10):

$$I = H_{\max} - H(p) = 1 + p \log p + q \log q = 1 + \left[\left(\frac{1}{2} + \varepsilon \right) \ln \left(\frac{1}{2} + \varepsilon \right) + \left(\frac{1}{2} - \varepsilon \right) \ln \left(\frac{1}{2} - \varepsilon \right) \right] / \ln 2 \approx 2\varepsilon^2 / \ln 2$$

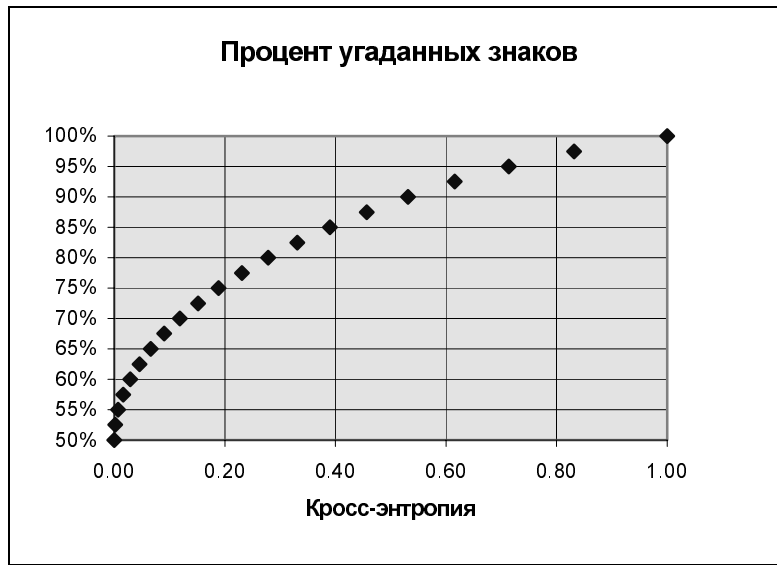


Рисунок 10. Доля правильно угаданных направлений изменений ряда как функция кросс-энтропии знака выхода при известных входах

В итоге получаем следующую оценку нормы прибыли при заданной величине предсказуемости знака I , выраженной в битах:

$$K_t = K_0 2^{aIt}.$$

То есть, для ряда с предсказуемостью I в принципе возможно удвоить капитал за $t = 1/(aI)$ вхождений в рынок. Так, например, измеренная выше предсказуемость временного ряда S&P500, равная $I = 0.17$ (см. Рисунок 8) предполагает удвоение капитала в среднем за $t = 1/(0.8 \cdot 0.17) \approx 8$ вхождений в рынок. Таким образом, даже небольшая предсказуемость знака изменения котировок способна обеспечить весьма заметную норму прибыли.

Подчеркнем, что оптимальная норма прибыли требует достаточно аккуратной игры, когда при каждом вхождении в рынок игрок рискует строго определенной долей капитала:

$$\frac{\langle \Delta K \rangle}{K} = \delta \langle |x| \rangle = (p - q) \langle |x| \rangle^2 / \langle x^2 \rangle = 2a\varepsilon \approx 1.6\varepsilon,$$

где ΔK - типичная при данной *волатильности* рынка $\langle |x| \rangle$ величина выигрыша или проигрыша.⁷ Как меньшие, так и большие значения ставок уменьшают прибыль. Причем, чересчур рискованная игра может привести к проигрышу при любой предсказательной способности. Этот факт иллюстрирует Рисунок 11.

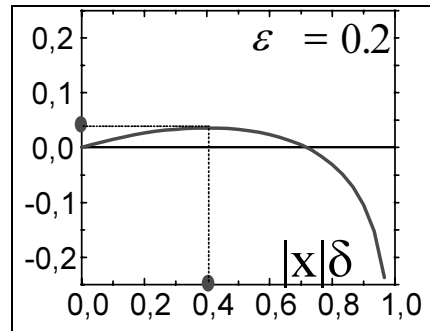


Рисунок 11. Зависимость средней нормы прибыли от выбора доли капитала "на кону"

Поэтому приведенные выше оценки дают представление лишь о верхнем пределе нормы прибыли.⁸ Более тщательный анализ с учетом влияния флуктуаций, выходит за рамки нашего изложения. Качественно понятно, однако, что выбор оптимального размера контрактов требует оценки точности предсказаний на каждом шаге.

Выбор функционала ошибки

Если принять, что целью предсказаний финансовых временных рядов является максимизация прибыли, логично настраивать нейросеть именно на этот конечный результат. Например, при игре по описанной выше схеме для обучения нейросети можно выбрать следующую функцию ошибки обучения, усредненную по всем примерам из обучающей выборки:

$$E = -\left\langle \ln \left[1 + x_i \delta_i \operatorname{sgn}(y_i) \right] \right\rangle.$$

⁷ Это отношение, в принципе, может быть не малым даже при очень маленькой волатильности - за счет механизмов залоговой торговли, практикуемых на биржах

⁸ Помимо явного пренебрежения комиссиями, вызывает сомнения само предположение о постоянном перевложении капитала. В более реалистичной игре на фиксированном капитале экспоненциальный рост сменяется линейным.

Здесь доля капитала в игре введена в качестве дополнительного выхода сети, настраиваемого в процессе обучения. При таком подходе, первый нейрон, y_i , с функцией активации $f = \tanh(\cdot)$ даст вероятность возрастания или убывания курса, в то время как второй выход сети δ_i даст рекомендованную долю капитала в игре на данном шаге.

Поскольку, однако, в соответствии с предыдущим анализом, эта доля должна быть пропорциональна степени уверенности предсказания, можно заменить два выхода сети - одним, положив $\delta_i = \delta |y_i|$, и ограничиться оптимизацией всего одного глобального параметра δ минимизирующего ошибку:

$$E = -\left\langle \ln[1 + \delta x_i y_i] \right\rangle$$

Тем самым, появляется возможность регулировать ставку в соответствии с уровнем риска, предсказываемым сетью. Игра с переменными ставками приносит большую прибыль, чем игра с фиксированными ставками. Действительно, если зафиксировать ставку, определив ее по *средней* предсказуемости, то скорость роста капитала будет пропорциональна $\langle \varepsilon \rangle^2$, тогда как если определять оптимальную ставку на каждом шаге, то - пропорциональна $\langle \varepsilon^2 \rangle \geq \langle \varepsilon \rangle^2$.

Использование комитетов сетей

Из-за случайности в выборе начальных значений синаптических весов, предсказания сетей, обученных на одной и той же выборке, будут, вообще говоря, различаться. Этот недостаток (элемент неопределенности) можно превратить в достоинство, организовав *комитет* нейро-экспертов, состоящий из различных нейросетей. Разброс в предсказаниях экспертов даст представление о степени уверенности этих предсказаний, что можно использовать для правильного выбора стратегии игры.

Легко показать, что среднее значений комитета должно давать лучшие предсказания, чем средний эксперт из этого же комитета. Пусть ошибка i -ого эксперта для значения входа x равна $\varepsilon_i(x)$. Средняя ошибка комитета всегда меньше среднеквадратичной ошибки отдельных экспертов в силу неравенства Коши:

$$\left(\frac{1}{L} \sum_{i=1}^L \varepsilon_i \right)^2 \leq \frac{1}{L} \sum_{i=1}^L \varepsilon_i^2$$

Причем, снижение ошибки может быть довольно заметным. Так, если ошибки отдельных экспертов не коррелируют друг с другом, т.е. $\langle \varepsilon_i \varepsilon_j \rangle = \langle \varepsilon_i^2 \rangle \delta_{ij}$, то среднеквадратичная ошибка комитета из L экспертов в $\sqrt{L}$ раз меньше, чем средняя индивидуальная ошибка одного эксперта!

$$E_L^2 \equiv \left\langle \left(\frac{1}{L} \sum_{i=1}^L \varepsilon_i \right)^2 \right\rangle = \frac{1}{L^2} \left\langle \sum_{i=1}^L \varepsilon_i^2 + \sum_{i \neq j} \varepsilon_i \varepsilon_j \right\rangle = \frac{1}{L^2} \sum_{i=1}^L \langle \varepsilon_i^2 \rangle = \frac{1}{L} E_1^2$$

Поэтому, в предсказаниях всегда лучше опираться на средние значения всего комитета. Иллюстрацией этого факта служит Рисунок 12.

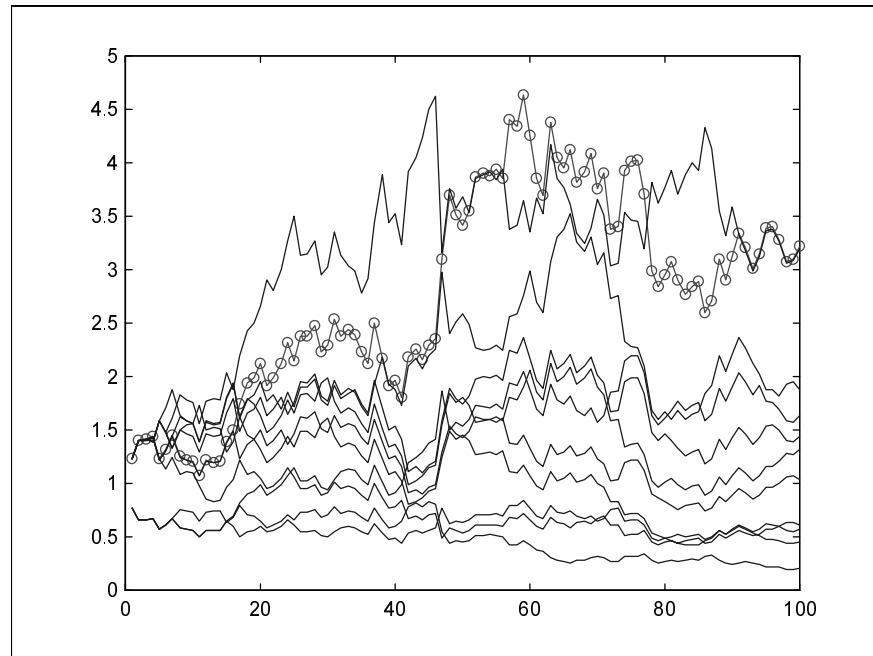


Рисунок 12. Норма прибыли на последних 100 значениях ряда sp500 при предсказании комитетом из 10 сетей. Выигрыш комитета (кружки) выше, чем выигрыш среднего эксперта. Счет угаданных знаков для комитета 59:41

Как видно из приведенного выше рисунка, в данном случае выигрыш комитета даже выше, чем выигрыш *каждого* из экспертов. Таким образом, метод комитетов может существенно повысить качество прогнозирования. Обратите внимание на абсолютное значение нормы прибыли: капитал комитета возрос в 3.25 раза при 100 вхождениях в рынок (Естественно, эта норма будет ниже при учете транзакционных издержек).

Предсказания получены при обучении сети на 30 последовательных экспоненциальных скользящих средних ($EMA_1 \dots EMA_{30}$) ряда приращений индекса. Предсказывался знак приращения на следующем шаге.

В этом эксперименте ставка была зафиксирована на уровне $\langle |x| \rangle \delta = 0.16$, близком к оптимальному при данной точности предсказаний (59 угаданных знаков против 41 ошибки) т.е. $p - q \approx 0.18$. На следующем же рисунке приведены результаты более рискованной игры *по тем же предсказаниям*, а именно - с $\langle |x| \rangle \delta = 0.2$.

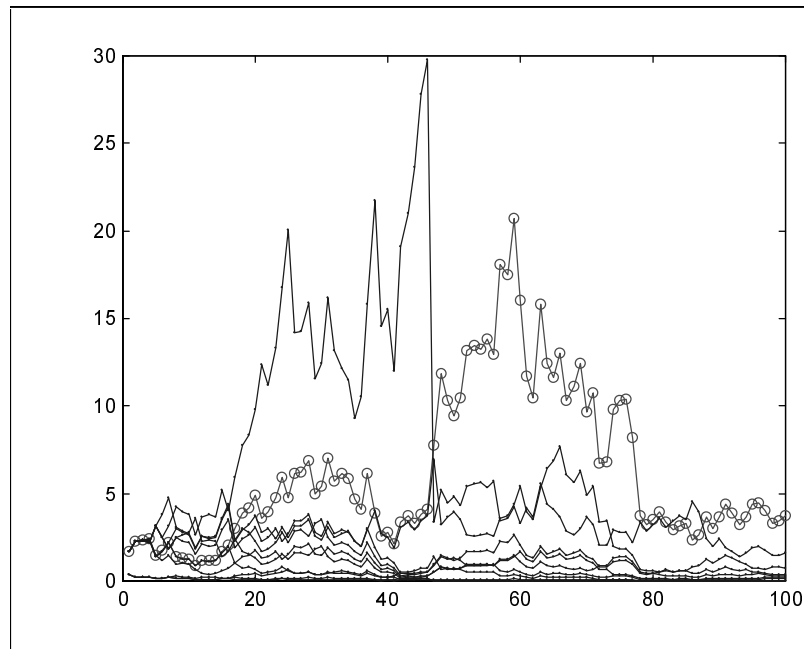


Рисунок 13. Норма прибыли на последних 100 значениях ряда sr500 при тех же предсказаниях комитета из 10 сетей, но по более рискованной стратегии.

Выигрыш комитета в целом остался на прежнем уровне (чуть увеличился), поскольку данное значение риска так же близко к оптимуму, как и предыдущее. Однако, для большинства сетей, предсказания которых хуже, чем у комитета в целом, такие ставки оказались слишком рискованными, что привело к практически полному их разорению.

Приведенные выше примеры показывают как важно уметь правильно оценить качество предсказания и как можно использовать эту оценку для увеличения прибыльности от одних и тех же предсказаний.

Можно пойти еще дальше и вместо *среднего* использовать *взвешенное* мнение сетей-экспертов. Веса выбираются адаптивно максимизируя предсказательную способность комитета на обучающей выборке. В итоге, хуже обученные сети из комитета вносят меньший вклад и не портят предсказания.

Возможности этого пути иллюстрирует приведенное ниже сравнение предсказаний двух типов комитетов из 25 экспертов (см. Рисунок 14 и Рисунок 15). Предсказания проводились по одной и той же схеме: в качестве входов использовались экспоненциальные скользящие средние приращений ряда с периодами равными первым 10 числам Фибоначчи. По результатам 100 экспериментов взвешенное предсказание дает в среднем превышение правильно угаданных знаков над ошибочным равное примерно 15, тогда как среднее - около 12. Заметим, что общее число повышений курса над понижением за указанный период как раз равно 12. Следовательно, учет общей тенденции к повышению в виде тривиального постоянного предсказания знака "+" дает такой же результат для процента угаданных знаков, что и взвешенное мнение 25 экспертов.

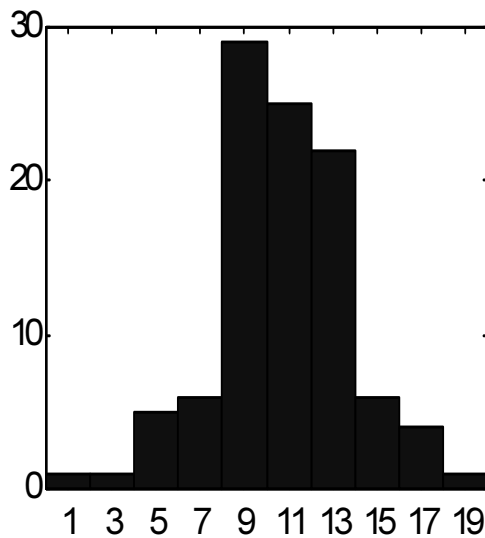


Рисунок 14. Гистограмма сумм угаданных знаков при средних предсказаниях 25 экспертов. Среднее по 100 комитетам = 11.7 при стандартном отклонении 3.2

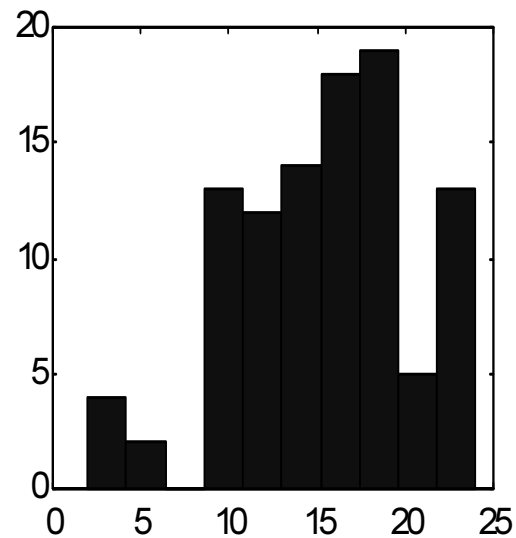


Рисунок 15. Гистограмма сумм угаданных знаков при взвешенных предсказаниях тех же 25 экспертов. Среднее по 100 комитетам = 15.2 при стандартном отклонении 4.9

Возможная норма прибыли нейросетевых предсказаний

До сих пор результаты численных экспериментов формулировались нами в виде процента угаданных знаков. Зададимся теперь вопросом о *реально* достижимой норме прибыли при игре с помощью нейросетей. Полученные выше без учета влияния флуктуаций верхние границы нормы прибыли вряд ли достижимы на практике, тем более, что до сих пор мы не учитывали транзакционных издержек, которые могут свести на нет достигнутую степень предсказуемости.

Действительно, учет комиссионных приводит к появлению отрицательного члена в показателе экспоненты:

$$K_t = K_0 2^{(1.6\varepsilon^2 - \gamma)t}.$$

Причем, в отличие от степени предсказуемости ε , комиссия γ входит не квадратично, а линейно. Так, в приведенном выше примере типичные значения предсказуемости $\varepsilon = 0.08$ не смогут "пересилить" комиссию свыше $\gamma = 0.1$.

Чтобы дать читателю представление о реальных возможностях нейросей в этой области, приведем результаты автоматического нейросетевого трейдинга на трех финансовых инструментах, с различными характерными временами: значения индекса S&P500 с *месячными* интервалами между отсчетами, *дневные* котировки немецкой марки DM/\$ и *часовые* отсчеты фьючерсов на акции Лукойл на Российской бирже. Статистика предсказаний набиралась на 50 различных нейросистемах (содержащих комитеты из 50 нейросетей каждая). Сами ряды и результаты по предсказанию знаков на тестовой выборке из 100 последних значений каждого ряда приведены на следующем рисунке.

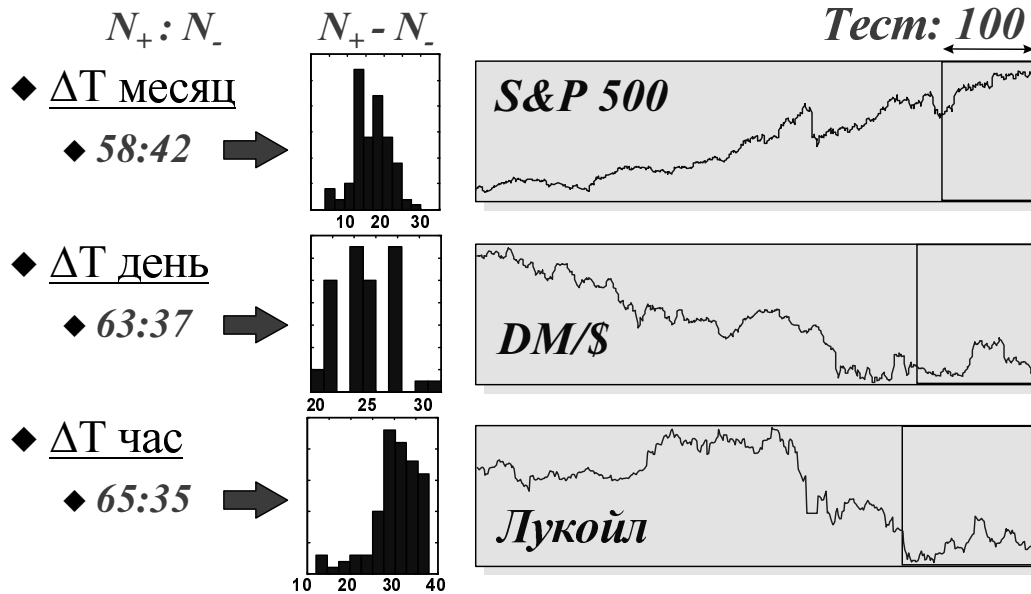


Рисунок 16. Средние значения и гистограммы количества правильно (N_+) и неправильно (N_-) угаданных знаков на тестовых выборках из 100 значений трех реальных финансовых инструментов.

Эти результаты подтверждают интуитивно понятную закономерность: ряды тем более предсказуемы, чем меньше времени проходит между его отсчетами. Действительно, чем больше временной масштаб между последовательными значениями ряда, тем больше внешней по отношению к его динамике информации доступно участникам рынка, и, соответственно меньше информации о будущем содержится в самом ряде.

Далее полученные выше предсказания использовались для игры на тестовой выборке. При этом, размер контракта на каждом шаге выбирался пропорциональным степени уверенности предсказания, а значение глобального параметра δ оптимизировалось по обучающей выборке. Кроме того, в зависимости от своих успехов, каждая сеть в комитете имела свой плавающий рейтинг, и в предсказаниях на каждом шаге использовалась лишь "лучшая" в данный момент половина сетей. Результаты таких нейро-трейдеров показаны на следующем рисунке (Рисунок 17).

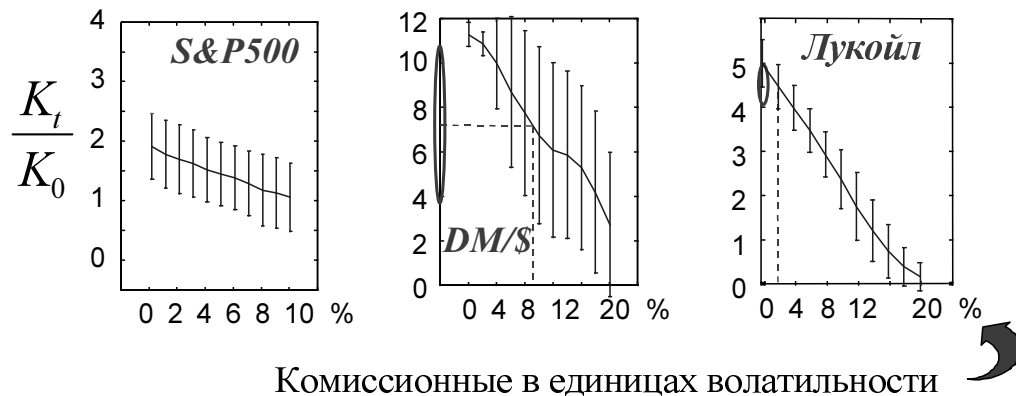


Рисунок 17. Статистика выигрышей по 50 реализациям в зависимости от величины комиссионных. Реалистичные значения комиссионных, показанные пунктиром, определяют область реально достижимых норм прибыли.

Итоговый выигрыш (как и сама стратегия игры), естественно, зависит от величины комиссионных. Эта зависимость и изображена приведенных выше графиках. Там, где реалистичные значения комиссионных в выбранных единицах измерений были известны авторам, они отмечены на рисунке. Уточним, что в этих экспериментах не учитывалась "квантованность" реальной игры, т.е. то, что величина сделок должна равняться целому числу типовых *контрактов*. Этот случай соответствует игре на большом капитале, когда типичные сделки содержат много контрактов. Кроме того, подразумевалась залоговая форма игры, т.е. норма прибыли исчислялась к *залоговому капиталу*, гораздо меньшему, чем масштабы самих контрактов.

Приведенные выше результаты свидетельствуют о перспективности нейросетевого трейдинга, по крайней мере на "коротких" временных масштабах. Более того, в силу самоподобия финансовых временных рядов (Peters, 1994), норма прибыли за *единицу времени* будет тем выше, чем меньше характерное время трейдинга. Таким образом, автоматические нейросетевые трейдеры оказываются наиболее эффективны при торговле в реальном времени, где как раз наиболее заметны их преимущества над обычными брокерами: неутрачиваемость, неподверженность эмоциям, потенциально гораздо более высокая скорость реагирования. Обученная нейросеть, подсоединенная к электронной системе торгов, может принимать решения еще до того, как брокер-человек успеет распознать изменения графика котировок на своем терминале.

Заключение

Подытожим результаты этой главы. Во-первых, мы показали, что (по крайней мере некоторые) рыночные временные ряды частично предсказуемы. Как и любой другой вид нейроанализа, предсказание временных рядов требует достаточно сложной и тщательной предобработки данных. Однако, работа с временными рядами имеет свою специфику, которую можно использовать для увеличения прибыли. Это касается как выбора входов (использование специальных способов представления данных), так и выбора выходов и использования специфических функционалов ошибки. Наконец, мы показали, насколько выгоднее может быть использование комитетов нейро-экспертов по сравнению с отдельными нейросетями, и представили данные о реальных нормах прибыли на нескольких реальных финансовых инструментах.

Литература

- Шарп, У.Ф., Александер Г.Дж., Бэйли, Дж. В. (1997). *Инвестиции*. Инфра-М.
- Abu-Mostafa, Y.S. (1995). "Financial market applications of learning from hints". In *Neural Networks in Capital Markets*. Apostolos-Paul Refenes (Ed.), Wiley, 221-232.
- Beltratti, A., Margarita, S., and Terna, P. (1995). *Neural Networks for Economic and Financial Modeling*. ITCP.
- Chorafas, D.N. (1994). *Chaos Theory in the Financial Markets*. Probus Publishing.
- Colby, R.W., Meyers, T.A. (1988). *The Encyclopedia of Technical Market Indicators*. IRWIN Professional Publishing.
- Ehlers, J.F. (1992). *MESA and Trading Market Cycles*. Wiley.
- Kaiser, G. (1995). *A Friendly Guide to Wavelets*. Birk.
- LeBeau, C., and Lucas, D.W. (1992). *Technical traders guide to computer analysis of futures market*. Business One Irwin.
- Peters, E.E. (1994). *Fractal Market Analysis*. Wiley.
- Pring, M.G. (1991). *Technical Analysis Explained*. McGraw Hill.
- Plummer, T. (1989). *Forecasting Financial Markets*. Kogan Page.
- Sauer, T., Yorke, J.A., and Casdagli, M. (1991). "Embedology". *Journal of Statistical Physics*. **65**, 579-616.
- Vemuri, V.R., and Rogers, R.D., eds. (1993). *Artificial Neural Networks. Forecasting Time Series*. IEEE Comp.Soc.Press.
- Weigend, A and Gershenfield, eds. (1994). *Times series prediction: Forecasting the future and understanding the past*. Addison-Wesley.
- Бэстенс, Д.-Э., Ван Ден Берг, В.-М., Вуд, Д. (1997). *Нейронные сети и финансовые рынки. Принятие решений в торговых операциях*. ТВП Научное издательство.