

ИНФОРМАЦИОННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ФИНАНСОВОГО ПРОГНОЗИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ КЛАССИФИКАТОРА СО СЛУЧАЙНЫМИ ПОДПРОСТРАНСТВАМИ

Abstract: This article describes the technical forecasting procedure, which was conducted using the feedforward neural network classifier. In order to predict the price move for the next day two cases were considered: the set of price increment values was divided on two and five categories correspondingly. The error rate and training speed dependencies on the number of input parameters were investigated. The results, obtained in these experiments, were structured according to the actual price change. The probabilistic analysis allows to estimate the quantity of useful information provided by the classifier. Some joint distributions for input parameters and required class numbers were investigated. The small values of the obtained correlation coefficients show the complexity of the forecasting problem.

Key words: neural network, financial forecasting, quantity of information, correlation coefficient, joint distribution, random subspace classifier.

Анотація: У роботі описано експеримент по технічному прогнозуванню ціни акцій на наступний день з використанням нейромережевого класифікатора з випадковими підпросторами, розглянуті випадки розділення множини значень приросту ціни на дві та п'ять категорій. Досліджені динаміка числа помилок прогнозування та швидкість навчання нейронної мережі в залежності від кількості вхідних параметрів, проведено структурне дослідження отриманих результатів. Ймовірнісний аналіз даних експеримента дозволяє оцінити кількість корисної інформації, що надається класифікатором. Розглянуті сумісні розподіли між вхідними параметрами та номерами класів, які обумовлені дійсною зміною ціни. Малі значення отриманих коефіцієнтів кореляції демонструють складність поставленої задачі.

Ключові слова: нейронна мережа, фінансове прогнозування, кількість інформації, коефіцієнт кореляції, сумісний розподіл, класифікатор з випадковими підпросторами.

Аннотация: В работе описан эксперимент по техническому прогнозированию цены акций на следующий день с использованием нейросетевого классификатора со случайными подпространствами, рассмотрены случаи разделения множества значений приращения цены на две и пять категорий. Исследованы динамика числа ошибок прогнозирования и скорость обучения нейронной сети в зависимости от количества входных параметров, проведено структурное исследование полученных результатов. Вероятностный анализ эксперимента позволяет оценить количество полезной информации, предоставляемой классификатором. Рассмотрены совместные распределения между входными параметрами и номерами классов, которые определяются реальным изменением цены. Малые значения полученных коэффициентов корреляции демонстрируют сложность поставленной задачи.

Ключевые слова: нейронная сеть, финансовое прогнозирование, количество информации, коэффициент корреляции, совместное распределение, классификатор со случайными подпространствами.

1. Введение

Использование нейронных сетей для решения задач финансового прогнозирования становится все более популярным среди специалистов, занимающихся исследованием рынков. Во многих работах показано преимущество нейросетевого подхода по сравнению с традиционным статистическим анализом данных. Используемый в работе классификатор со случайными подпространствами (Random Subspace Classifier, RSC) является высокопроизводительным и универсальным нейросетевым классификатором. Использование бинарных нейронов, а также архитектура прямого распространения позволяют как программную, так и аппаратную реализацию данной нейронной сети. Изначально классификатор был предложен в работе [1]. Точная модель сети, используемая в данной статье, описана в [2]. Кроме того, в [2] проведен анализ статистического поведения нейронной сети, даны рекомендации по выбору ее параметров, оценка минимального различного расстояния. Классификатор со случайными подпространствами использует схему грубого кодирования (coarse coding scheme) для преобразования исходного вещественного вектора в

бинарный вектор большой размерности. Интересный сравнительный анализ схем грубого кодирования приведен в [3]. Разделение полученных бинарных образов производится при помощи перцептрона. Классификатор использует идею метода опорных векторов (support vector machine, SVM), которая позволяет провести линейную разделяющую поверхность в пространстве характеристик (feature space), в то время как в исходном пространстве разделяющая поверхность, как правило, нелинейна. С другой стороны можно провести аналогию с нейронной сетью на основе радиально-базисных функций. Достаточно широкий обзор работ по использованию нейронных сетей для различных задач финансового анализа и прогнозирования представлен в [4, 5].

2. Использование классификатора со случайными подпространствами для выдачи рекомендаций по техническим характеристикам

Традиционным для прогнозирования поведения фондового рынка является использование следующих пяти показателей за каждый бизнес-день. *Open* – цена акций данного типа на момент открытия торгов; *High* – самая высокая цена сделки за текущий день; *Low* – самая низкая цена за день; *Close* – цена акций на момент закрытия торгов; *Volume* – объем торгов, т.е. количество акций, которые изменили владельца за текущий день. Следует отметить, что данная информация, очевидно, не является полной касательно финансового состояния компании. Тем не менее рассмотрим возможности прогнозирования на основе только этих данных. Данный подход часто называют техническим прогнозированием.

В приведенной ниже табл. 1 показана часть истории цен (historical quotes) для компании IBM. Для обучения классификатора со случайными подпространствами использовались данные за 1998, 1999 и 2000 годы, всего 756 бизнес-дней. Данные за 2001 год использовались в качестве тестовой выборки, всего 248 дней. Динамика цен представлена ниже, на рис. 1. Временной диапазон используемых данных был выбран таким образом, чтобы нормализацию цен можно было опустить вследствие того, что диапазон изменения цен не является значительным.

Таблица 1. Ежедневные ценовые показатели и данные объема торгов, используемые для генерации обучающей и тестовой выборки

Date	Open	High	Low	Close	Volume
02/01/1998	52,25	52,875	52,03	52,81	5275600
05/01/1998	52,78	53,405	52,50	53,22	10041800
06/01/1998	52,685	53,595	52,53	52,625	7117600
07/01/1998	52,50	52,50	51,31	52,125	8625600
08/01/1998	51,75	52,75	51,47	52,095	8129400
...
27/12/2001	122,75	123,93	122,64	123,50	3826200
28/12/2001	123,5	123,88	122,35	122,90	4213900
31/12/2001	122,75	123,24	120,95	120,96	3512700

На вход классификатора подавались данные как за текущий день, так и за несколько предшествующих ему, т.е. было использовано окно обратного просмотра (looking back window).

Например, если использовалась k -дневная история, то на вход классификатора подавалось $5 * k$ входных параметров.

Далее в работе рассматривается классификация на две и пять категорий. Наиболее простая задача технического прогнозирования состоит в предсказании изменения цены на следующий день. В качестве определяющей целесообразно выбрать цену $Close$, так как она меньше всего подвержена случайным факторам, спекуляциям, новостям и т.д.

Для классификации на две категории каждому входному вектору присваивался класс Buy , если цена закрытия на следующий день была выше, чем за текущий, т.е. при выполнении условия $Close(i + 1) > Close(i)$, и класс $Sell$, если $Close(i + 1) \leq Close(i)$. Здесь i обозначает текущий день, а $i + 1$ – последующий. В случае классификации на пять категорий в качестве выхода классификатора использовались следующие значения: $Strong Buy$, Buy , $Hold$, $Sell$ и $Strong Sell$. Такой способ оценки акций, как правило, используется для аналитических рекомендаций, выдаваемых крупными финансовыми компаниями. Таким образом, каждому входному вектору присваивался класс в соответствии со следующими условиями:

$Strong Buy$, если $Close(i) \cdot 1.015 < Close(i + 1)$;

Buy , если $Close(i) \cdot 1.005 < Close(i + 1) \leq Close(i) \cdot 1.015$;

$Hold$, если $Close(i) \cdot 0.995 \leq Close(i + 1) \leq Close(i) \cdot 1.005$;

$Sell$, если $Close(i) \cdot 0.985 \leq Close(i + 1) < Close(i) \cdot 0.995$;

$Strong Sell$, если $Close(i + 1) < Close(i) \cdot 0.985$.

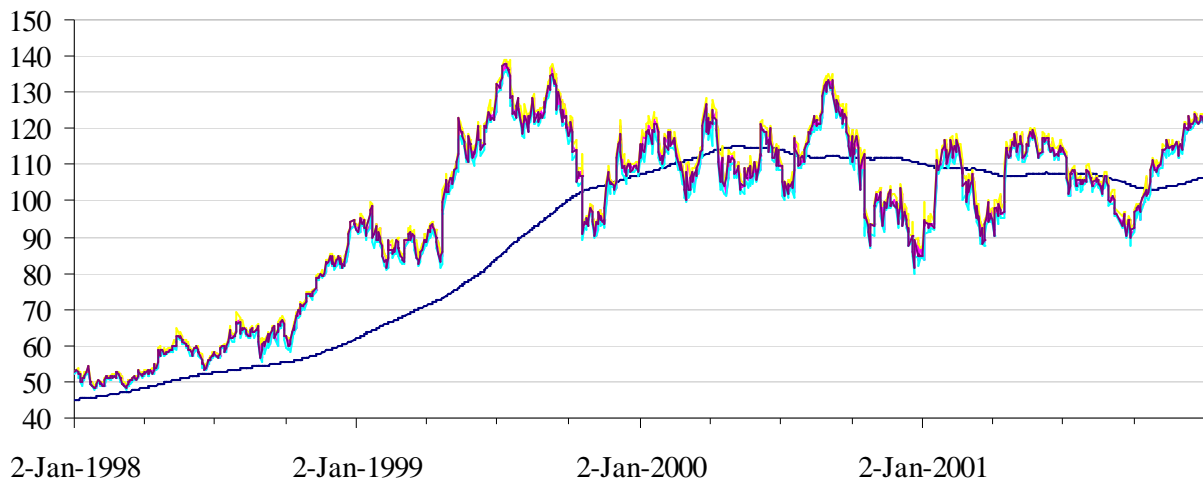


Рис. 1. Динамика цены акций компании IBM, 1998-2001 гг., скользящее среднее по цене закрытия за 256 дней

Коэффициенты были выбраны таким образом, чтобы обеспечить приблизительно равное представительство для каждого из классов. Это условие обеспечивает максимальное количество информации, которое выдается нейросетевым классификатором. В терминах нечеткой логики выход классификатора в данном случае эквивалентен лингвистической переменной. Существуют множество вариантов выбора критериев для выдачи рекомендаций, полезность того или иного критерия зависит от стратегии, выбранной участником торгов. Известно, что максимальное количество категорий, которыми человек легко может оперировать, составляет от пяти до семи.

Таким образом, использование большего количества классов было бы сложным для восприятия, а также неоправданным с точки зрения возможностей нейронной сети. С другой стороны, использование только двух или трех категорий привело бы к чрезмерному упрощению задачи.

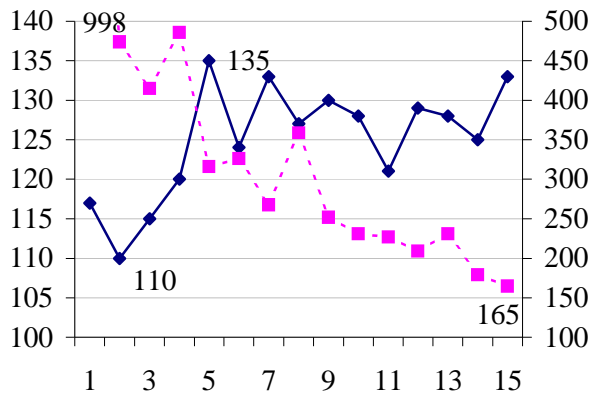


Рис. 2. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории входных данных ($m = 2$, параметр подпространств, $\eta = 3$)

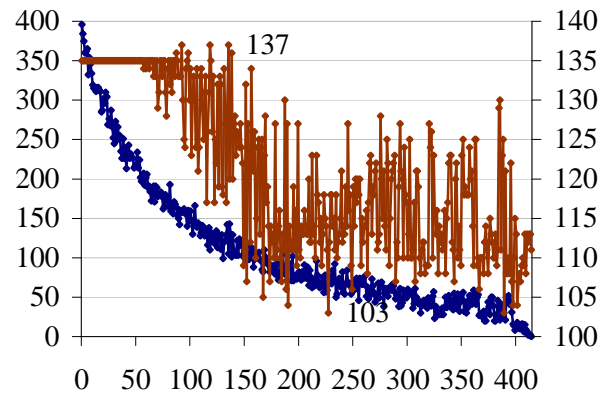


Рис. 3. Количество ошибок распознавания на обучающей выборке (левая ось) и количество ошибок прогнозирования на тестовой выборке (изменяющаяся зависимость, правая ось) в зависимости от номера эпохи ($m = 2$, параметр подпространств, $\eta = 3$)

Для приведенного выше эксперимента были использованы следующие параметры нейронной сети: количество нейронов среднего слоя классификатора со случайными подпространствами составляло 32768, размерность подпространства каждого нейрона скрытого слоя – 3, ширина накрывающего отрезка по одному измерению – 1,0. Изменения в конфигурации сети для последующих экспериментов будут специально оговариваться. Зависимость числа ошибок прогнозирования от количества дней в истории в случае классификации на две категории показана на рис. 2, количество категорий здесь обозначается как m . На рис. 3 представлены динамика уменьшения количества операций обучения (number of reinforcements) на обучающей выборке и динамика количества ошибок прогнозирования на тестовой выборке в зависимости от номера эпохи обучения. Для построения зависимости была использована двухдневная история.

Следует отметить, что случай классификации на две категории является достаточно важным, так как предоставляет непосредственную информацию для принятия решения о покупке или продаже данного типа акций. Кроме того, этот критерий позволяет сравнивать полезность прогноза с инструментами, которые прогнозируют вещественное значение цены. Знак изменения ценового показателя эквивалентен спецификации одной из двух категорий. Наименьшее количество ошибок получено для двухдневной истории. Для получения более статистически устойчивых результатов случаи с использованием истории длиной до 3-х дней рассмотрены ниже. Относительно неплохие результаты, полученные для этих условий, объясняют популярность использования метода японских свечей (candlesticks) [6].

Таблица 2. Статистика уровня ошибок прогнозирования в зависимости от числа дней в истории (однородная структура классификатора, две категории, параметр подпространств $\eta = 3$)

История	Количество ошибок					Среднее
	1	2	3	4	5	
1 день	117	109	113	112	110	112,2
2 дня	110	113	112	112	112	111,8
3 дня	115	116	115	123	114	116,6

Как показано в табл. 2, среднее число ошибок при использовании двухдневной истории составляет 111,8. Так как тестовая выборка содержит всего 248 векторов, то усредненная вероятность успеха составляет 54,9%. Лучший результат для 110 ошибок соответствует вероятности 56%. Таким образом, число ошибок распознавания увеличивается при увеличении длины истории, а время обучения уменьшается. Это можно объяснить тем, что классификатор интерпретирует случайные флуктуации как некоторую закономерность или распределение, несмотря на то, что при увеличении количества дней в истории классификатор получает больше информации. Принято считать, что задача прогнозирования направления изменения цены является одной из самых сложных, особенно в случае, когда в качестве входной информации используются только технические данные по одному типу акций. Одной из причин является то, что этот индикатор непосредственно влияет на инвестиционные решения и таким образом большинство инвесторов пытаются использовать каждую зависимость, которую они находят в динамике цены акций. Достаточно сложно получить вероятность успеха, существенно превышающую 50%.

Типичные результаты для традиционных сетей прямого распространения представлены в работах [7–9]. Классификация на две категории аналогична выделению линейной зависимости в истории цен. Как показано в [4], эти линейные зависимости слишком слабы для получения любого статистически значимого результата. Полезнее может оказаться выделение нелинейных зависимостей. В частности, более подробный нелинейный анализ обеспечивает классификация на пять категорий.

Предположим, что существует некоторое распределение вероятности изменения цены закрытия на следующий день: $r = (Close(i+1) - Close(i)) / Close(i)$ (англ. stock return). Даже если вероятность того, что цена вырастет, равна вероятности того, что цена уменьшится, нелинейные зависимости можно использовать. Пример распределения плотности вероятности показан на рис. 4.

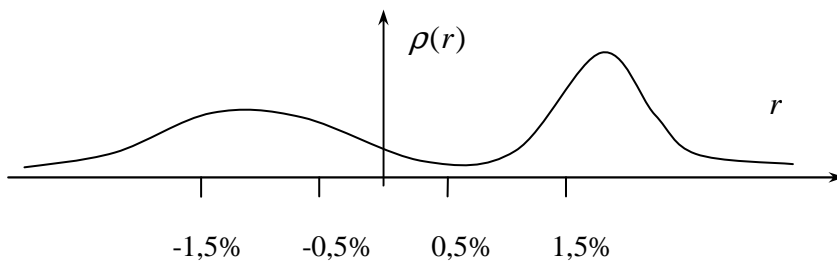


Рис. 4. Пример распределения плотности вероятности изменения цены закрытия

Вероятность того, что цена акций возрастет, равна $\int_0^{+\infty} \rho(r) dr$, соответственно вероятность уменьшения цены равна $\int_{-\infty}^0 \rho(r) dr$.

Средний выигрыш длинной

позиции пропорционален выражению $\int_{-\infty}^{+\infty} r \rho(r) dr$, которое является положительным для приведенного здесь распределения. В данном случае при классификации на пять категорий классификатор должен выдавать рекомендацию *Strong Buy*. Несмотря на предположение о том, что вероятности увеличения и уменьшения цены равны, такой вероятностный паттерн на протяжении нескольких, следующих друг за другом, дней обеспечивает тренд на повышение.

третьей независимой выборки. Кроме того, при последовательной верификации динамики уровня ошибок (cross-validation) уменьшение числа ошибок на одном тестовом множестве не гарантирует уменьшения на другом. Программная реализация данного алгоритма предполагает хранение состояния сети для наилучшего результата, достигнутого на тестовой выборке. Как можно заключить, корреляция между пиками и впадинами не является очевидной.

При проведении данного эксперимента, после завершения обучения, только около 45% синаптических весов имели ненулевые значения, максимальное значение было 360. Количество синаптических весов, имеющих некоторое определенное значение, имело распределение, близкое к гиперболическому. Нулевые веса получены в основном благодаря тому, что не все нейроны скрытого слоя бывают активными хотя бы раз при предъявлении всей обучающей выборки. Этот недостаток можно устранить, используя различные методы повышения информативности структуры классификатора. Для каждого входного вектора следует максимизировать детализацию представления, что может быть достигнуто при использовании тех же вычислительных ресурсов.

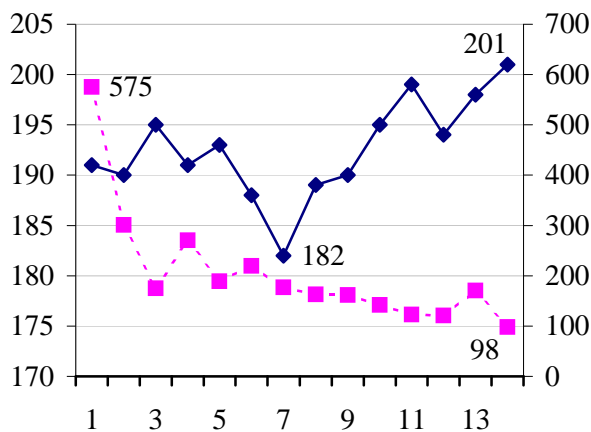


Рис. 7. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох полного обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории ($m = 5$, параметр подпространств, $\eta = 5$)

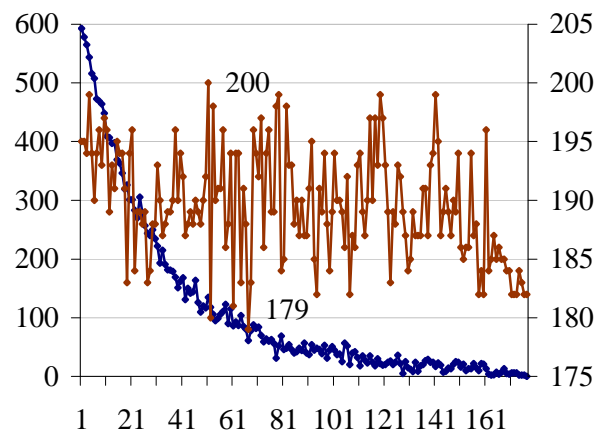


Рис. 8. Количество ошибок распознавания на обучающей выборке (левая ось) и количество ошибок на тестовой выборке (изменчивая зависимость, правая ось) в зависимости от номера эпохи, ($m = 5$, параметр подпространств, $\eta = 5$)

Выше, на рис. 7, приведена зависимость числа ошибок классификации в случае пяти категорий при другом значении параметра $\eta = 5$. Как и в предыдущих экспериментах, число групп нейронов составляло 32768, ширина накрывающего отрезка – 1,0. Проведенные исследования зависимости числа ошибок от параметра подпространств показывают, что результаты достаточно близки для значений $\eta = 2,3,4,5$. Тем не менее сравнение зависимостей представляет собой достаточный интерес. Кроме того, наблюдается общий локальный минимум для длины истории в семь дней. Как было продемонстрировано в [2], увеличение параметра подпространств позволяет строить более сложные разделяющие поверхности с многомерной структурой. Недостатком этого является уменьшение детализации представления данных в скрытом слое классификатора. Данные количества ошибок прогнозирования и количества эпох обучения обладают некоторой дисперсией. Как видно из полученных зависимостей, статистическую устойчивость для одного эксперимента нельзя считать достаточной. Тем не менее характер зависимостей определить

можно, причем он подтверждается при проведении повторных экспериментов. Динамика процесса обучения при использовании семидневной истории и параметре $\eta = 5$ представлена на рис. 8. Зависимости количества ошибок на обучающей и тестовой выборках от номера эпохи обучения для многих условий эксперимента достаточно похожи.

Как можно заключить из приведенных зависимостей (классификация на пять категорий), оптимальные результаты по количеству ошибок, с учетом времени обучения, достигаются при использовании данных за последние 1–7 дней. Таким образом, полученные выше результаты показывают, что рассматриваемая задача достаточно сложна. В частности, ценовые колебания имеют значительную случайную составляющую, которая зачастую на протяжении одного дня больше, чем тренд изменения цены за несколько дней. Как показано в работе далее, корреляционные зависимости будущего изменения цены от входных данных достаточно малы.

При количестве дней меньше семи разброс результатов можно объяснить различным количеством входной информации и большим статистическим разбросом данных. Соответственно, увеличение количества ошибок при количестве дней истории больше семи обуславливается тем, что на большем объеме данных классификатор находит больше случайных зависимостей, чем закономерных. Это подтверждается тем, что обучение классификатора в этом случае занимает меньшее количество эпох. Например, связывание цены и объема торгов за достаточно удаленные дни, очевидно, не может выявить какие-либо закономерности. В то же время количество таких комбинаций при случайной генерации структуры классификатора комбинаторно возрастает с увеличением количества дней в истории. Выявление полезных закономерностей во входных данных возможно, например, при применении генетических алгоритмов по выбору оптимальной структуры классификатора, при применении селекции структуры классификатора по какому-либо из критериев либо при удалении некоторых элементов сети (network pruning) [10].

В проведенных экспериментах производилось полное обучение классификатора, то есть количество правильных ответов на обучающей выборке составляло 100%. Подобная возможность полного обучения является особенностью классификатора со случайными подпространствами и не является типичной для нейронных сетей в целом. Таким образом, после завершения обучения данная сеть является переученной (network is overfitted). Имеется в виду, что недостатком сети может быть большая чувствительность к случайным флуктуациям во входных данных, которая, возможно, приводит к искажению предполагаемой закономерности. Положительным аспектом является то, что данный классификатор может формировать сложные разделяющие поверхности и адаптироваться к задачам со сложной и мелкой пространственной структурой.

3. Анализ структуры результатов прогнозирования

Очевидно, что количество ошибок классификации в случае использования, например, двузначной рекомендации существенно ниже. В то же время большое количество ошибок при разделении на пять классов не означает малую полезность этой рекомендации. Например, рекомендация *Buy* будет достаточно полезной, если фактическое изменение цены на следующий день будет соответствовать рекомендации *Strong Buy*.

На рис. 9 приведены гистограммы ошибок классификации, по оси абсцисс отложена величина ошибки, в частности, значению 0 соответствуют правильные рекомендации, положительным значениям соответствуют завышенные рекомендации, а отрицательным – заниженные. Таким образом, количество правильных рекомендаций существенно превышает количество ошибочных рекомендаций какого-либо определенного типа, в то же время общее число неправильных рекомендаций достаточно велико.

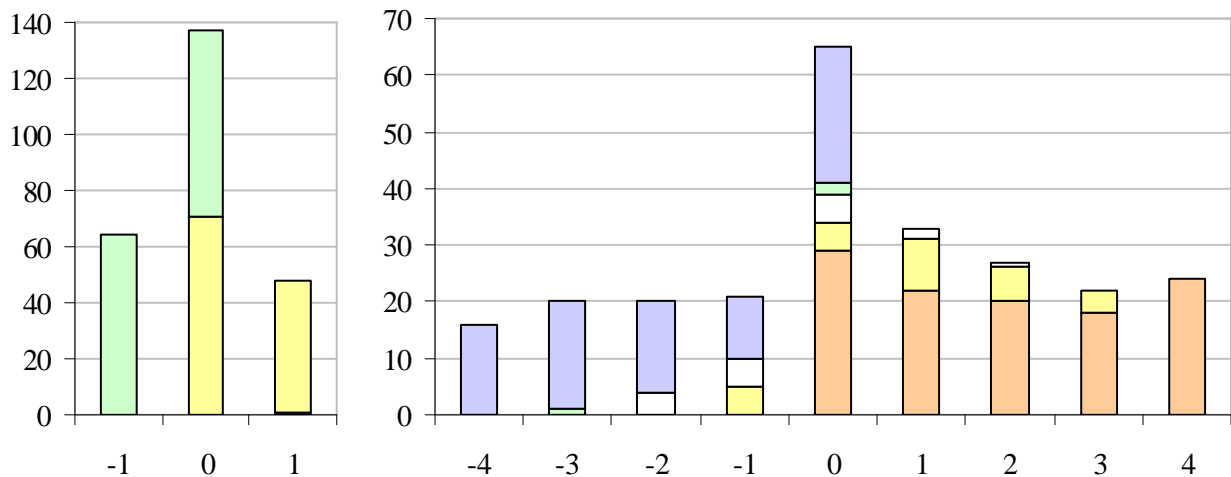


Рис. 9. Гистограмма ошибок прогнозирования на тестовой выборке в зависимости от значения, выдаваемого классификатором (классификация на две и на пять категорий, параметр, $\eta = 3$)

Распределение ошибок классификации в зависимости от рекомендации, выдаваемой классификатором, представлено в табл. 4 и 5. Альтернативным вариантом является рассмотрение числа ошибок в зависимости от класса, определяемого реальным изменением цены. Верхняя строка таблиц содержит величину ошибки. Следует отметить, что частота встречаемости категорий *Buy*, *Hold* и *Sell* существенно меньше, чем для *Strong Buy* и *Strong Sell*. Это смещение, в основном, определяется представительством соответствующих категорий в обучающем множестве. Таким образом, классификатор “увеличивает” области пространства, принадлежащие тем категориям, которые имеют большее количество представителей в обучающей выборке. Соответственно правильный выбор множителей в правилах определения категорий играет существенную роль.

При случайном распределении категорий математическое ожидание числа ошибок определенной величины равно $L_T \cdot (m - |e|) / m^2$, где e – ошибка прогнозирования; L_T – длина

тестовой выборки, в частности, $\sum_{e=1-m}^{m-1} (m - |e|) = m + 2 \sum_{e=1}^{m-1} (m - e) = m^2$. Таким образом, операция

прогнозирования искажает случайное распределение категорий для максимизации среднего столбца гистограммы за счет остальных. Распределение для пяти категорий на рис. 9 демонстрирует, что выделение центрального столбца гистограммы происходит в большей степени за счет соседних столбцов. В то же время оптимальным является равномерное уменьшение столбцов, соответствующих ненулевому значению ошибки. Последнее может быть достигнуто, например, при использовании модифицированной процедуры обучения.

Таблица 4. Распределение ошибок прогнозирования по рекомендации классификатора (пять категорий, семидневная история, параметр подпространств, $\eta = 3$)

Ошибка	-4	-3	-2	-1	0	1	2	3	4	
Всего	16	20	20	21	65	33	27	22	24	
Strong Buy	0	0	0	0	29	22	20	18	24	113
Buy	0	0	0	5	5	9	6	4	0	29
Hold	0	0	4	5	5	2	1	0	0	17
Sell	0	1	0	0	2	0	0	0	0	3
Strong Sell	16	19	16	11	24	0	0	0	0	86

Таблица 5. Распределение ошибок прогнозирования по рекомендации классификатора (пять категорий, семидневная история, параметр подпространств, $\eta = 5$)

Ошибка	-4	-3	-2	-1	0	1	2	3	4	
Всего	16	25	24	24	66	32	22	23	16	
Strong Buy	0	0	0	0	22	17	18	23	16	96
Buy	0	0	0	1	0	0	0	0	0	1
Hold	0	0	7	1	12	6	4	0	0	30
Sell	0	7	3	6	6	9	0	0	0	31
Strong Sell	16	18	14	16	26	0	0	0	0	90

Для сравнения результатов прогнозирования с исходными данными распределение категорий для обучающей и тестовой выборок представлено в табл. 6. Как видно, представители классов распределены относительно равномерно.

Таблица 6. Распределение векторов для обучающей выборки (верхние две строки) и тестовой выборки (нижние две строки) в зависимости от требуемой категории

Класс	Strong Sell	Sell	Hold	Buy	Strong Buy
Число	178	116	152	120	190
Процент	23,54%	15,34%	20,11%	15,87%	25,13%
Число	53	39	50	51	55
Процент	21,37%	15,72%	20,16%	20,56%	22,18%

Также была исследована возможность получения двузначной рекомендации по результатам классификации на большее число категорий. В проведенных экспериментах при классификации на пять и восемь категорий полученные результаты не давали статистически значимой зависимости для случая классификации на две категории, несмотря на то, что некоторая регулярность в случае нескольких категорий действительно была получена. Аналогию можно провести с коэффициентами гармоник при разложении функции в ряд Фурье. Значение коэффициента при одной гармонике не дает информации о величине других коэффициентов.

Сравним количество полезной информации, которое выдается классификатором в экспериментах для двух и пяти категорий соответственно. Определения теории информации приведены в работе [11]. Пусть опыт α_2 заключается в определении двузначной рекомендации классификатора, опыт β_2 состоит в определении одной из двух категорий по реальному изменению цены. Таким образом, опыты α_2 и β_2 могут иметь только два исхода: *Buy* и *Sell*.

Количество информации, которое содержится в опыте α_2 , относительно опыта β_2 определяется следующим выражением:

$$I(\alpha_2, \beta_2) = H(\beta_2) - H(\beta_2 | \alpha_2),$$

где $H(\beta_2)$ – энтропия (степень неопределенности) опыта β_2 , а $H(\beta_2 | \alpha_2)$ – средняя условная энтропия опыта β_2 при условии выполнения опыта α_2 . Значение энтропии $H(\beta_2)$ вычисляется в соответствии с определением $H = -\sum_i p_i \cdot \log(p_i)$:

$$H(\beta_2) = -\mathbf{P}\{\beta_2 = Buy\} \cdot \log(\mathbf{P}\{\beta_2 = Buy\}) - \mathbf{P}\{\beta_2 = Sell\} \cdot \log(\mathbf{P}\{\beta_2 = Sell\}).$$

В теории информации традиционно используются логарифмы по основанию 2. Используя данные проведенного выше эксперимента,

$$\mathbf{P}\{\beta_2 = Buy\} \approx (64 + 71) / 248 \approx 0,544355; \quad \mathbf{P}\{\beta_2 = Sell\} \approx (47 + 66) / 248 \approx 0,455645.$$

Следовательно, для полной энтропии получаем значение $H(\beta_2) \approx 0,994316$ бита. Далее, средняя условная энтропия определяется следующим выражением:

$$H(\beta_2 | \alpha_2) = \mathbf{P}\{\alpha_2 = Buy\} \cdot H(\beta_2 | \alpha_2 = Buy) + \mathbf{P}\{\alpha_2 = Sell\} \cdot H(\beta_2 | \alpha_2 = Sell).$$

Для распределения рекомендаций классификатора нетрудно получить значения

$$\mathbf{P}\{\alpha_2 = Buy\} \approx (47 + 71) / 248 \approx 0,475806; \quad \mathbf{P}\{\alpha_2 = Sell\} \approx (64 + 66) / 248 \approx 0,524194.$$

Условные энтропии определяются следующими выражениями:

$$\begin{aligned} H(\beta_2 | \alpha_2 = Buy) &= -\mathbf{P}\{\beta_2 = Buy | \alpha_2 = Buy\} \log(\mathbf{P}\{\beta_2 = Buy | \alpha_2 = Buy\}) - \\ &\quad - \mathbf{P}\{\beta_2 = Sell | \alpha_2 = Buy\} \log(\mathbf{P}\{\beta_2 = Sell | \alpha_2 = Buy\}); \\ H(\beta_2 | \alpha_2 = Sell) &= -\mathbf{P}\{\beta_2 = Buy | \alpha_2 = Sell\} \log(\mathbf{P}\{\beta_2 = Buy | \alpha_2 = Sell\}) - \\ &\quad - \mathbf{P}\{\beta_2 = Sell | \alpha_2 = Sell\} \log(\mathbf{P}\{\beta_2 = Sell | \alpha_2 = Sell\}). \end{aligned}$$

Воспользуемся следующими оценками условных вероятностей:

$$\begin{aligned} \mathbf{P}\{\beta_2 = Buy | \alpha_2 = Buy\} &\approx 71 / (47 + 71) \approx 0,601695; \\ \mathbf{P}\{\beta_2 = Sell | \alpha_2 = Buy\} &\approx 47 / (47 + 71) \approx 0,398305; \\ \mathbf{P}\{\beta_2 = Buy | \alpha_2 = Sell\} &\approx 64 / (64 + 66) \approx 0,492308; \\ \mathbf{P}\{\beta_2 = Sell | \alpha_2 = Sell\} &\approx 66 / (64 + 66) \approx 0,507692. \end{aligned}$$

Таким образом, $H(\beta_2 | \alpha_2 = Buy) \approx 0,969950$, $H(\beta_2 | \alpha_2 = Sell) \approx 0,999829$. Средняя условная энтропия $H(\beta_2 | \alpha_2) \approx 0,985612$. Следовательно, полезное количество информации составляет $I(\alpha_2, \beta_2) = H(\beta_2) - H(\beta_2 | \alpha_2) \approx 0,008704$ бита.

Рассматриваемое количество информации можно получить иным способом, воспользовавшись известным соотношением

$$H(\alpha\beta) = H(\alpha) + H(\beta | \alpha) = H(\beta) + H(\alpha | \beta),$$

где $\alpha\beta$ представляет собой совместный опыт, определенный на декартовом произведении множества исходов опытов α и β соответственно. В частности,

$$\begin{aligned}
 H(\alpha\beta) &= -\sum_i \sum_j \mathbf{P}\{\alpha = i, \beta = j\} \log(\mathbf{P}\{\alpha = i, \beta = j\}) = \\
 &= -\sum_i \sum_j \mathbf{P}\{\beta = j | \alpha = i\} \cdot \mathbf{P}\{\alpha = i\} \cdot \log(\mathbf{P}\{\beta = j | \alpha = i\} \cdot \mathbf{P}\{\alpha = i\}) = \\
 &= -\sum_i \sum_j \mathbf{P}\{\beta = j | \alpha = i\} \cdot \mathbf{P}\{\alpha = i\} \cdot \log(\mathbf{P}\{\beta = j | \alpha = i\}) - \\
 &\quad - \sum_i \sum_j \mathbf{P}\{\beta = j | \alpha = i\} \cdot \mathbf{P}\{\alpha = i\} \cdot \log(\mathbf{P}\{\alpha = i\}) = \\
 &= -\sum_i \mathbf{P}\{\alpha = i\} \sum_j \mathbf{P}\{\beta = j | \alpha = i\} \cdot \log(\mathbf{P}\{\beta = j | \alpha = i\}) - \\
 &\quad - \sum_i \mathbf{P}\{\alpha = i\} \cdot \log(\mathbf{P}\{\alpha = i\}) \cdot \sum_j \mathbf{P}\{\beta = j | \alpha = i\} = \\
 &= \sum_i \mathbf{P}\{\alpha = i\} \cdot H(\beta | \alpha = i) - \sum_i \mathbf{P}\{\alpha = i\} \cdot \log(\mathbf{P}\{\alpha = i\}) = H(\beta | \alpha) + H(\alpha).
 \end{aligned}$$

Таким образом, можно воспользоваться следующим соотношением:

$$I(\alpha, \beta) = H(\beta) - H(\beta | \alpha) = H(\alpha) + H(\beta) - H(\alpha\beta) = H(\alpha) - H(\alpha | \beta) = I(\beta, \alpha).$$

Далее получим требуемые энтропии $H(\alpha_2)$ и $H(\alpha_2\beta_2)$:

$$H(\alpha_2) = -\mathbf{P}\{\alpha_2 = Buy\} \cdot \log(\mathbf{P}\{\alpha_2 = Buy\}) - \mathbf{P}\{\alpha_2 = Sell\} \cdot \log(\mathbf{P}\{\alpha_2 = Sell\});$$

$$\mathbf{P}\{\alpha_2 = Buy\} \approx (47+71)/248 \approx 0,475806;$$

$$\mathbf{P}\{\alpha_2 = Sell\} \approx (64+66)/248 \approx 0,524194;$$

$$\mathbf{P}\{\alpha_2 = Buy, \beta_2 = Buy\} \approx 71/248 \approx 0,286290;$$

$$\mathbf{P}\{\alpha_2 = Buy, \beta_2 = Sell\} \approx 47/248 \approx 0,189516;$$

$$\mathbf{P}\{\alpha_2 = Sell, \beta_2 = Buy\} \approx 64/248 \approx 0,258065;$$

$$\mathbf{P}\{\alpha_2 = Sell, \beta_2 = Sell\} \approx 66/248 \approx 0,266129;$$

$$H(\alpha_2) \approx 0,998310, \quad H(\alpha_2\beta_2) \approx 1,983923;$$

$$I(\alpha_2, \beta_2) = H(\alpha_2) + H(\beta_2) - H(\alpha_2\beta_2) \approx 0,008703 \text{ бита.}$$

Воспользуемся полученными результатами для определения полезного количества информации в случае пяти категорий. Ниже использованы данные распределения ошибок прогнозирования, приведенные в табл. 4. Обозначим выражение $-p \log(p)$ как $h(p)$, кроме того, определим неопределенность невозможного исхода $h(0) = 0$, в частности, $\lim_{p \rightarrow 0} (-p \cdot \log(p)) = 0$:

$$\begin{aligned}
 H(\alpha_5) &= h(\mathbf{P}\{\alpha_5 = Strong Buy\}) + h(\mathbf{P}\{\alpha_5 = Buy\}) + \\
 &\quad + h(\mathbf{P}\{\alpha_5 = Hold\}) + h(\mathbf{P}\{\alpha_5 = Sell\}) + h(\mathbf{P}\{\alpha_5 = Strong Sell\}) \approx \\
 &= h(96/248) + h(1/248) + h(30/248) + h(31/248) + h(90/248) \approx 1,729782 \text{ бита;}
 \end{aligned}$$

$$\begin{aligned}
H(\beta_5) &= h(\mathbf{P}\{\beta_5 = \text{Strong Buy}\}) + h(\mathbf{P}\{\beta_5 = \text{Buy}\}) + \\
&+ h(\mathbf{P}\{\beta_5 = \text{Hold}\}) + h(\mathbf{P}\{\beta_5 = \text{Sell}\}) + h(\mathbf{P}\{\beta_5 = \text{Strong Sell}\}) \approx \\
&= h(55/248) + h(51/248) + h(50/248) + h(39/248) + h(53/248) \approx 2,312375 \text{ бита.}
\end{aligned}$$

Энтропия совместного опыта $H(\alpha_5 \beta_5)$ выражается следующим образом:

$$\begin{aligned}
H(\alpha_5 \beta_5) &= h(\mathbf{P}\{\alpha_5 = \beta_5 = \text{Strong Buy}\}) + h(\mathbf{P}\{\alpha_5 = \text{Strong Buy}, \beta_5 = \text{Buy}\}) + \dots \\
&\dots + h(\mathbf{P}\{\alpha_5 = \text{Strong Sell}, \beta_5 = \text{Sell}\}) + h(\mathbf{P}\{\alpha_5 = \beta_5 = \text{Strong Sell}\}) \approx \\
&\approx h(29/248) + h(22/248) + \dots + h(11/248) + h(24/248) \approx 4,011429 \text{ бита;} \\
I(\alpha_5, \beta_5) &= H(\alpha_5) + H(\beta_5) - H(\alpha_5 \beta_5) \approx 0,030728 \text{ бита.}
\end{aligned}$$

Сравнивая количество информации, содержащееся в опыте α относительно опыта β , можно заключить, что классификация на пять категорий является гораздо более информативной по сравнению с классификацией на две категории. Кроме того, можно максимизировать полезное количество информации в зависимости от количества категорий. В то же время практическим критерием полезности того или иного подхода является стратегия участника торгов. Таким образом, получено дополнительное подтверждение того, что использование нелинейного анализа (классификация на пять категорий) предоставляет больше полезной информации.

Альтернативный способ оценки полезности процедуры классификации состоит в рассмотрении опыта γ , который может иметь два исхода: $\alpha = \beta$ или $\alpha \neq \beta$. Рассмотрим энтропию данного опыта:

$$\begin{aligned}
H(\gamma) &= -\mathbf{P}\{\alpha = \beta\} \cdot \log(\mathbf{P}\{\alpha = \beta\}) - \mathbf{P}\{\alpha \neq \beta\} \cdot \log(\mathbf{P}\{\alpha \neq \beta\}) = \\
&= -\sum_i \sum_{j=i} \mathbf{P}\{\alpha = i, \beta = j\} \cdot \log(\sum_k \sum_{l=k} \mathbf{P}\{\alpha = k, \beta = l\}) - \\
&\quad - \sum_i \sum_{j \neq i} \mathbf{P}\{\alpha = i, \beta = j\} \cdot \log(\sum_k \sum_{l \neq k} \mathbf{P}\{\alpha = k, \beta = l\}).
\end{aligned}$$

Оценим значение $H(\gamma)$ для проведенных выше экспериментов:

$$\begin{aligned}
H(\gamma_2) &= h(\mathbf{P}\{\alpha_2 = \beta_2\}) + h(\mathbf{P}\{\alpha_2 \neq \beta_2\}) \approx \\
&\approx h((71+66)/248) + h((47+64)/248) \approx 0,992057 \text{ бита;} \\
H(\gamma_5) &= h(\mathbf{P}\{\alpha_5 = \beta_5\}) + h(\mathbf{P}\{\alpha_5 \neq \beta_5\}) \approx \\
&\approx h(65/248) + h((248-65)/248) \approx 0,829894 \text{ бита.}
\end{aligned}$$

Так как $\mathbf{P}\{\alpha_2 = \beta_2\} \geq \mathbf{P}\{\alpha_2 \neq \beta_2\}$, то улучшение результатов классификации соответствует уменьшению $H(\gamma_2)$. В случае пяти классов $\mathbf{P}\{\alpha_5 = \beta_5\} \leq \mathbf{P}\{\alpha_5 \neq \beta_5\}$, следовательно, целесообразно увеличение $H(\gamma_5)$. Некоторые оценки $H(\gamma)$ можно получить, используя неравенство Йенсена $\sum_i p_i f(x_i) \leq f(\sum_i p_i x_i)$, которое справедливо при условии $\sum_i p_i = 1$ и выпуклости подграфика функции $f(x)$.

4. Выборочное удаление синаптических связей

Единственными параметрами классификатора, которые зависят от данных обучающей последовательности, являются веса синаптической матрицы. Учитывая, что входные данные имеют значительную случайную составляющую, можно предположить, что полезный вклад синаптических весов в принятие решения не является равномерным по отношению к величине весов. Например, большие веса могут обуславливать случайные флуктуации вблизи разделяющих поверхностей, что может существенным образом искажать полезные закономерности, сформированные в синаптической матрице.

Ниже, на рис. 10, приведена гистограмма значений весов синаптической матрицы. Полученная зависимость достаточно близка к гиперболической. После завершения процесса обучения матрица имеет 135126 нулевых весов и только 28714 ненулевых. Последнее можно объяснить тем, что обучающая выборка содержит только некоторое ограниченное множество комбинаций диапазонов входных параметров. Для приведенной ниже зависимости максимальное значение веса синаптической матрицы составляло 216. Следует упомянуть такие методы более полного использования синаптической матрицы: использование чувствительной структуры классификатора, использование генетического алгоритма формирования структуры, нормализация входных данных и т.п. Из средств обработки самой матрицы наиболее популярными являются удаление связей (weight elimination, pruning) и динамическое ослабление связей (weight decay).

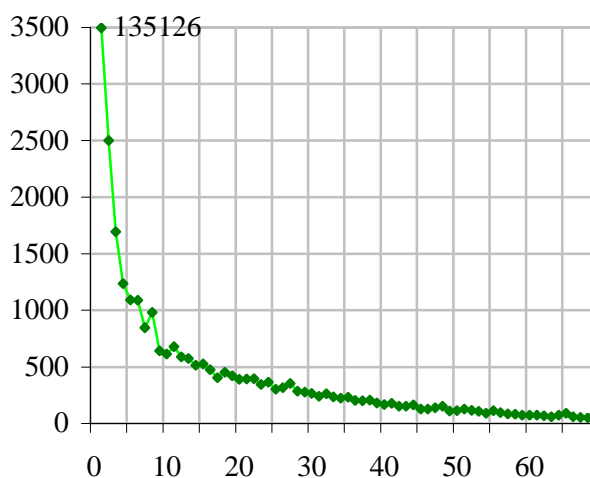


Рис. 10. Гистограмма значений весов синаптической матрицы после завершения процедуры обучения классификатора (пять категорий, семидневная история, параметр подпространств, $\eta = 5$)

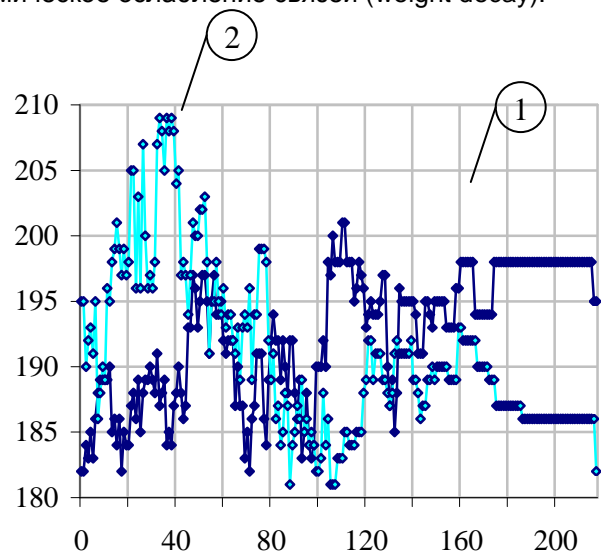


Рис. 11. Количество ошибок распознавания при обнулении малых (1) и больших (2) весов синаптической матрицы (пять категорий, семидневная история, $\eta = 5$)

Гиперболическое распределение значений синаптических весов позволяет сделать заключение, что полезность группы синаптических связей одной величины в среднем одинакова. Кроме того, устранение действия синаптических связей какого-либо определенного диапазона не дает улучшения результатов. Это подтверждают приведенные на рис. 11 две зависимости количества ошибок распознавания при обнулении некоторых синаптических весов в соответствии с их абсолютной величиной.

Для первой зависимости (темная линия), представленной на рис. 11, специально обнулялись все веса синаптической матрицы, которые меньше значения, отложенного по оси абсцисс. Для второй зависимости (светлая линия) обнулялись все синаптические веса больше этого значения. В качестве исходной синаптической матрицы была использована матрица, полученная для условий последнего эксперимента.

Таким образом, удаление синаптических связей в соответствии с их величиной не дает гарантированного улучшения результатов. Вполне возможно, что выборочное удаление синаптических связей может оказаться полезным, если минимизация количества ошибок рассматривается для каждого единичного обнуления веса. Следует отметить, что в этом случае эффективность метода следует проверять при помощи третьей независимой выборки.

5. Прогнозирование с использованием производных технических показателей

Техническое прогнозирование существенным образом зависит от набора входных параметров, используемых на входе нейронной сети. Ниже представлены результаты для некоторых комбинаций технических показателей. Для зависимости, представленной на рис. 12, в качестве набора параметров, используемых за каждый день, рассматривались три цены: *High*, *Low* и *Close*. Как видно, полученный уровень ошибок уступает полученным ранее результатам. Следует отметить большое количество эпох, необходимых для полного обучения сети. Последнее объясняется отсутствием во входных данных достаточно изменчивой информации об объеме торгов. Кроме того, было проверено, что остальные комбинации из пяти базовых параметров дают еще большее количество ошибок распознавания.

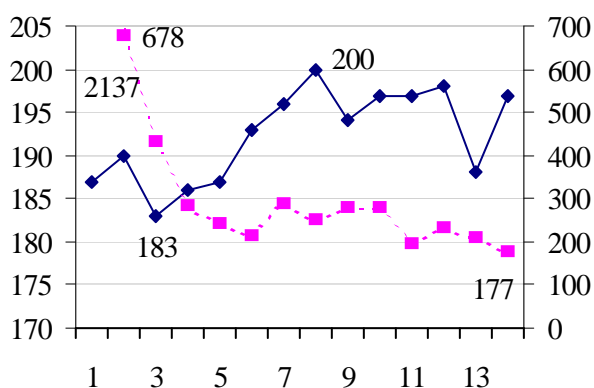


Рис. 12. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории (цены *High*, *Low* и *Close*, классификация на пять категорий, параметр подпространств, $\eta = 5$)

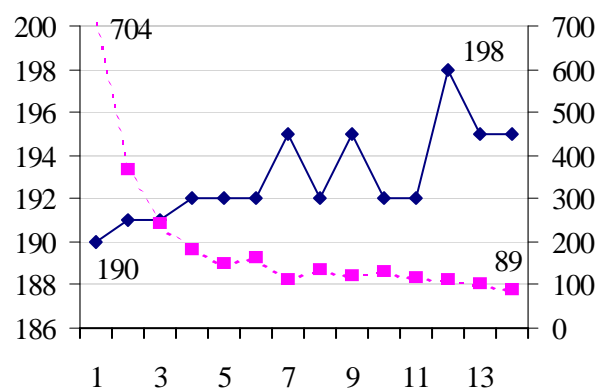


Рис. 13. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории (ценовые разности первого порядка и объем торгов, классификация на пять категорий, $\eta = 5$)

Другим подходом к выбору входных параметров является использование однодневных разниц для всех ценовых показателей. В терминологии временных рядов предлагается рассмотреть разности первого порядка (*first order differences*). В частности, для представленной на рис. 13 зависимости для каждого дня в истории рассматривались следующие параметры:

$$Open(i) - Open(i-1); High(i) - High(i-1); Low(i) - Low(i-1);$$

$$Close(i) - Close(i-1); Volume(i),$$

где i соответствует некоторому рассматриваемому дню.

Следует отметить, что более слабые результаты при использовании разностей первого порядка соответствуют большому содержанию шума во входных данных. На рис. 14 представлены зависимости для другого набора показателей:

$$Open(k) - Open(i-1); High(k) - High(i-1); Low(k) - Low(i-1);$$

$$Close(k) - Close(i-1); Volume(i),$$

где i соответствует некоторому дню в истории, а k – последнему дню из рассматриваемого диапазона. Уровень результатов достаточно близок к уровню ошибок прогнозирования, полученному для базовых параметров. Фактически данное преобразование можно рассматривать как преобразование задачи классификации из одного пространства в другое, что не влияет существенным образом на результаты прогнозирования. Различие может быть получено в основном благодаря тому, что классификатор формирует различные разделяющие поверхности, которые не переводятся друг в друга прямым или обратным преобразованием.

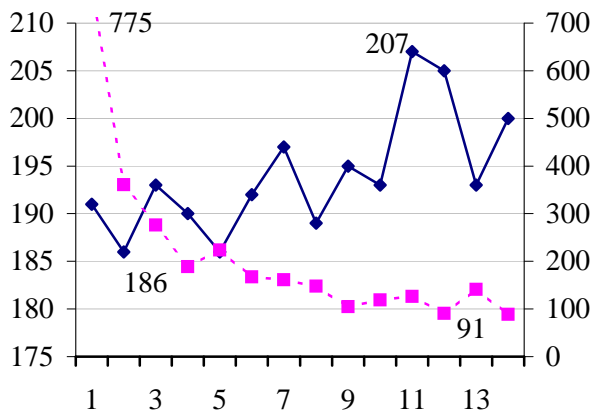


Рис. 14. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории (ценовые разности с последним днем и объем торгов, классификация на пять категорий, $\eta = 5$)

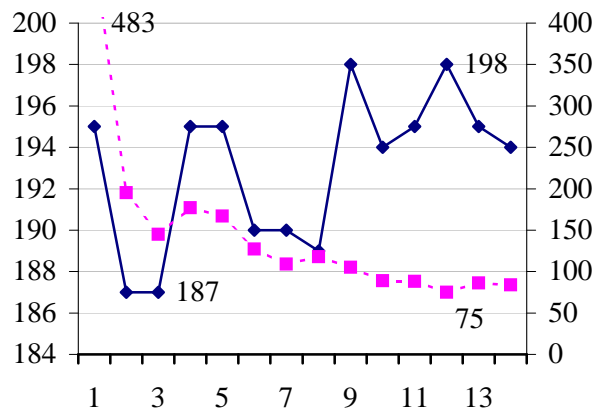


Рис. 15. Динамика количества ошибок на тестовой выборке (сплошная линия, левая ось) и количества эпох обучения (пунктирная линия, правая ось) в зависимости от количества дней в истории (параметры $RSI14$, $EMA9$, $EMA18$, $MACD9$, классификация на пять категорий, $\eta = 5$)

На рис. 15 представлен эксперимент с использованием технических показателей $RSI14$, $EMA9$, $EMA18$, $MACD9$. Алгоритмы вычисления этих показателей общеизвестны и изложены, например, в [12]. Уровень ошибок прогнозирования заметно выше, чем при использовании базового набора. В то же время подбор наиболее полезных параметров является достаточно сложной задачей. Например, только при анализе девяти технических показателей число возможных комбинаций составляет порядка 2^9 . Количество используемых технических показателей может быть гораздо больше [6, 12]. Кроме того, при рассмотрении возможности выбора отдельных параметров по времени число комбинаций возрастает еще больше.

6. Исследование совместного распределения входных данных

Типичным подходом к первичному анализу данных является исследование статистических закономерностей. В случае рассмотрения данных фондового рынка трудно выявить те или иные зависимости. Тем не менее анализ совместного распределения данных позволяет выделить полезную информацию, а также оценить сложность задачи. Ниже, на рис. 16, приведено совместное распределение цены закрытия и объема торгов.

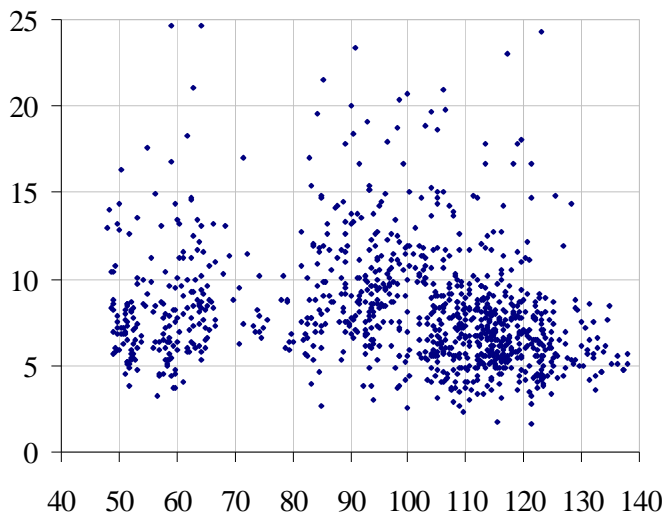


Рис. 16. Совместное распределение цены закрытия и объема торгов (в миллионах); использованы данные по акциям IBM за 1998-2001 гг.

Коэффициент корреляции для данного распределения оценивается следующим образом:

$$r(p, v) = \frac{\mathbf{M}(p - \mathbf{M}p)(v - \mathbf{M}v)}{\sqrt{\mathbf{D}p \cdot \mathbf{D}v}},$$

где p обозначает цену закрытия, а v – объем торгов. Средние значения и соответствующие дисперсии равны

$$\mathbf{M}p \approx 96,97; \quad \mathbf{M}v \approx 8380142;$$

$$\mathbf{M}(p - \mathbf{M}p)(v - \mathbf{M}v) \approx -14787565,5001;$$

$$\mathbf{D}p \approx 548,523977;$$

$$\mathbf{D}v \approx 1,884243e+13.$$

Следовательно, искомый коэффициент равен $r(p, v) \approx -0,145456$.

Таким образом, очевидна зависимость, состоящая в уменьшении объема торгов при увеличении котировок акций. Последнее можно объяснить тем, что при увеличении цены количество акций, которое можно приобрести за одну и ту же сумму денег, уменьшается. С другой стороны, коэффициент корреляции цены акций и денежного объема $p \cdot v$ (dollar volume) имеет положительное значение $r(p, p \cdot v) \approx 0,288115$. Приведенный анализ корреляционных зависимостей является линейным, а, значит, ограниченным. Следует также отметить, что значения коэффициентов существенным образом зависят от распределения диапазона цен на протяжении времени.

На рис. 17 показано совместное распределение изменения цены закрытия и соответствующего значения объема торгов, т.е. разницы $\Delta p = \text{Close}(i) - \text{Close}(i-1)$ и параметра $v = \text{Volume}(i)$. Коэффициент корреляции данного распределения равен $r(\Delta p, v) = -0,062347$. Так как это распределение симметрично по отношению к нулевому изменению цены, нетрудно заметить, что объем торгов сам по себе представляет мало информации для классификации на две категории либо для построения линейной регрессии. В то же время нелинейная зависимость существует. Она может быть полезной, например, для классификации на пять категорий. Закономерность может быть описана следующим правилом: чем больше объем торгов, тем больше абсолютное изменение цены. Таким образом, как подтверждается в работах [7, 9], вероятность

корректной идентификации направления изменения цены на следующий день не зависима от абсолютной величины изменения. Если рассмотреть совместные распределения изменения цены *Close* и объема торгов за предыдущий либо за следующий день, то можно выделить такие аспекты: распределение также имеет практически нулевую корреляцию, кроме того, нелинейная зависимость существенно меньше, чем в приведенном случае. Два параметра можно считать независимыми либо считать, что они предоставляют независимую информацию для решения какой-либо задачи, если их нормальное распределение обладает малой корреляцией. Исходя из распределения, представленного на рис. 17, можно заключить, что некоторая компрессия данных объема торгов, например, логарифмирование, приблизит распределение к нормальному.

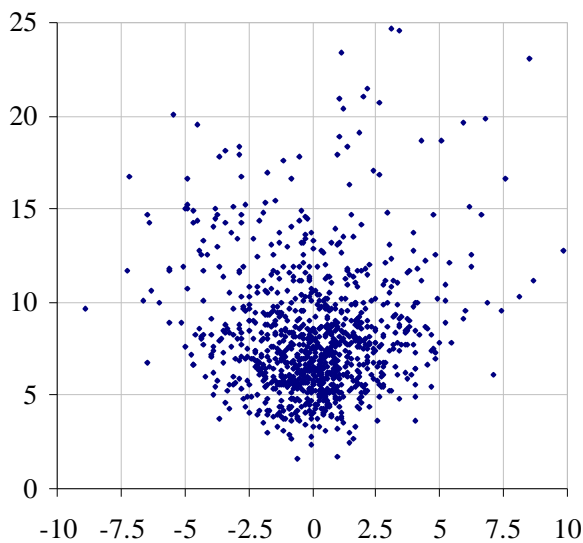


Рис. 17. Совместное распределение разницы цен закрытия (в долларах) и объема торгов (в миллионах акций). Использована история за 1998-2001 гг.

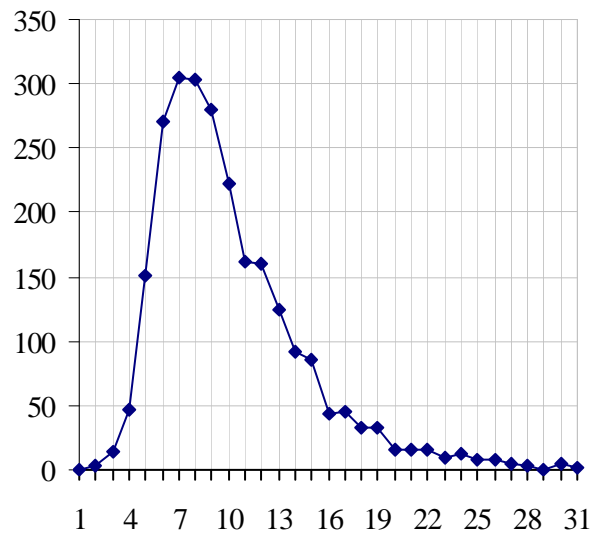


Рис. 18. Гистограмма объема торгов для компании IBM за десятилетие с 1993 по 2002 гг., в акциях, деноминированных на начало 2003 г.

Линейное распределение объема торгов показано на рис. 18. По оси абсцисс отложен объем торгов в миллионах акций, по оси ординат – количество дней, объем торгов, за которые попадает в соответствующий одномиллионный диапазон. Максимальное значение объема в рассмотренном множестве данных составляло 69'447'300. Следует обратить внимание, что распределение достаточно похоже на график плотности распределения χ^2 . Как известно, χ^2 распределением с m степенями свободы обладает сумма квадратов m независимых нормальных случайных величин с нулевым математическим ожиданием и единичной дисперсией. Это наблюдение хорошо согласуется с гипотезой случайного блуждания (random walk). Гипотеза предполагает, что будущее изменение цены не предсказуемо и является независимой случайной величиной. При моделировании случайного блуждания, как правило, рассматривают несмещенное нормальное распределение. В самом деле, реальные наблюдения за рынком показывают, что изменение цены акций, сопровождаемое тем или иным объемом торгов, обусловлено, как правило, новостями либо какими-то “внешними” факторами. Чем более существенной является новость, тем больший объем торгов имеет данный тип акций за текущий день. В гораздо меньшей степени объем обусловлен собственно динамикой цены акций.

Как уже упоминалось, благодаря значительной относительной дисперсии данных объема торгов, процесс обучения нейронной сети сходится существенно быстрее. Этот вывод был получен как при использовании классификатора со случайными подпространствами, так и в других работах, например, в [7].

Следует отметить, что анализ корреляционных зависимостей во входных данных не дает непосредственной информации касательно прогнозируемых значений ценовых показателей. Исследование возможностей прогнозирования можно провести, например, при рассмотрении плоскостей Пуанкаре временного ряда либо при исследовании корреляции с номером класса, к которому должен быть отнесен входной вектор. Плоскость Пуанкаре отображает данные временного ряда $x(t)$ как зависимость $x(t+1) = f(x(t))$, множество точек, нанесенных на данную плоскость (compass rose pattern), часто рассматриваются как диаграмма в полярных координатах для выявления регулярностей распределения.

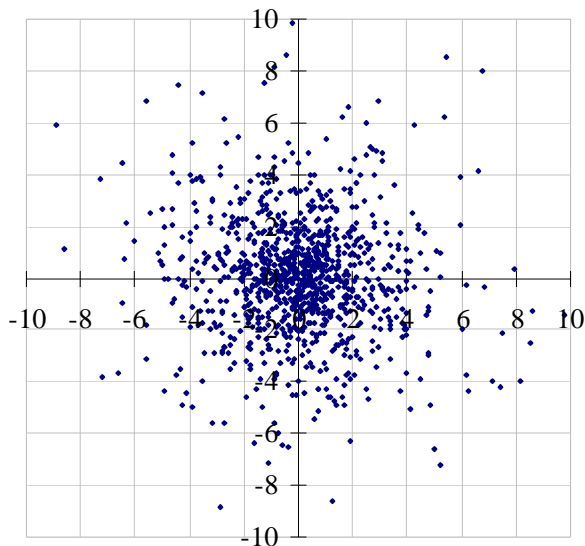


Рис. 19. Совместное распределение изменений цен закрытия с разницей в один день (в долларах), использована история по акциям IBM за 1998-2001 гг.

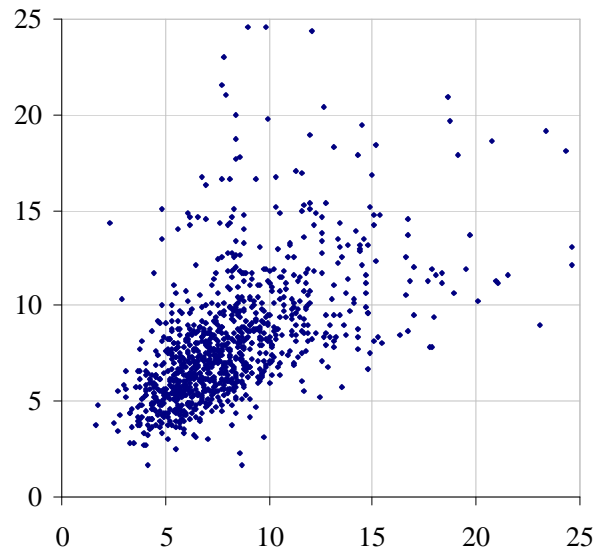


Рис. 20. Совместное распределение объемов торгов с разницей в один день (в миллионах), использована история по акциям IBM за 1998-2001 гг.

На рис. 19 показано совместное распределение изменений цен закрытия соответственно на текущий и последующий дни, по оси абсцисс отложено значение $Close(i) - Close(i-1)$, а по оси ординат – приращение $Close(i+1) - Close(i)$. Полученное распределение исключительно близко к нормальному. Таким образом, изменение цены $Close$ за текущий день практически не дает никакой информации касательно изменения цены на следующий день. Коэффициент корреляции для данного множества точек составляет $-0,065588$.

Автокорреляционное распределение для объема торгов показано на рис. 20. Очевидно, данному показателю свойственна некоторая инерционность, т.е. значению объема торгов определенной величины будет следовать значение, с большей вероятностью близкое к исходному. Это обусловлено вытянутостью рассматриваемого множества значений по большой диагонали. Коэффициент корреляции рассматриваемых случайных величин составляет $0,511376$.

В то же время данная регулярность не имеет непосредственной импликации на ценовую динамику. Значению коэффициента корреляции, которое достаточно близко к $1/2$, можно дать следующую интерпретацию: усредненное изменение объема торгов на следующий день в два раза меньше, чем средний разброс его значений.

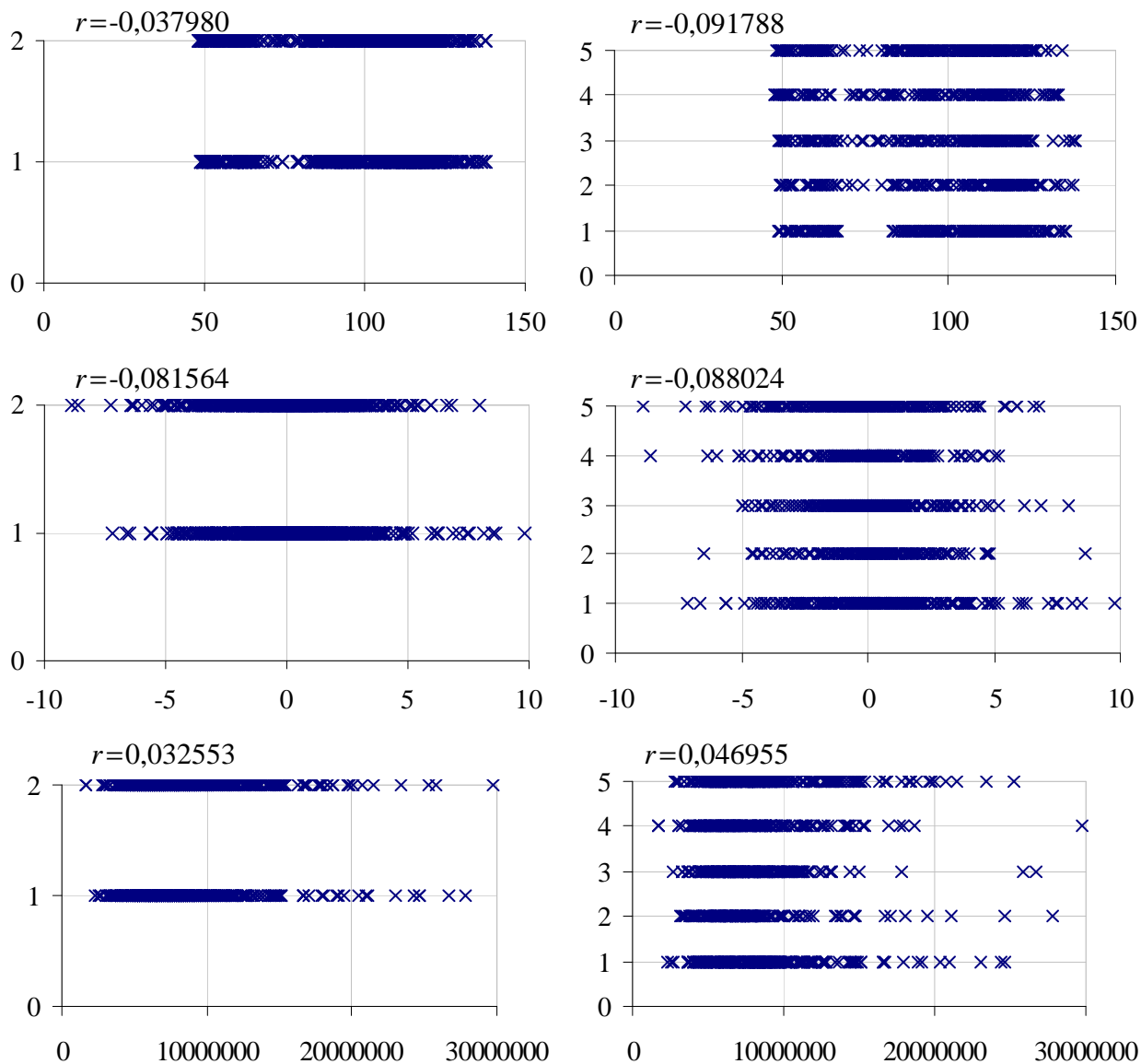


Рис. 21. Распределение цены закрытия, изменение цены закрытия и объема торгов совместно с номером класса в случае двух и пяти категорий соответственно, номер класса определяется изменением цены закрытия на следующий день, история по акциям IBM за 1998-2001 гг.

Некоторые аспекты динамики можно проследить, рассмотрев следующую логическую цепочку. Пусть, например, за текущий день цена акций существенно выросла. Как было получено выше, большему изменению цены чаще соответствует большее значение объема торгов. Далее можно заключить, что объем торгов на следующий день, вероятно, также будет выше среднего. Значит, на следующий день следует ожидать существенного изменения цены. На данном этапе знак изменения не известен. Принята следующая терминология: *Limit*-ордера, введенные на покупку, осуществляют поддержку цены (*support*), а *Limit*-ордера на продажу оказывают сопротивление (*resistance*) [6]. Так как за текущий день цена выросла значительно, то ввиду

инерционности участников торгов поддержка на следующий день ожидается сравнительно слабой. Следовательно, для обеспечения большего объема торгов необходим рост цены. Последнее заключение и обуславливает продолжение тренда. Несмотря на то, что вероятность выполнения данной цепочки мала, объем торгов играет существенную роль в развитии и ослаблении трендов. Кроме того, большое значение объема торгов часто служит усилителем сигнала покупки или продажи, полученного с использованием каких-либо других критериев.

Совместные распределения с номером класса и соответствующие коэффициенты корреляции представлены выше, на рис. 21. Очевидно, столь малая корреляционная зависимость не может быть использована для построения простых правил на основе какого-либо одного параметра.

7. Заключение

В работе показано, что задача финансового прогнозирования является достаточно сложной проблемой. Практически отсутствуют корреляционные зависимости будущего изменения цены и какого-либо одного из входных параметров. В то же время интересен тот факт, что достаточно успешное прогнозирование возможно с использованием нейронной сети. Таким образом, подтверждается, что классификатор со случайными подпространствами обладает способностью выявления существенно многомерных, сложных и нелинейных корреляционных зависимостей. Данная работа демонстрирует использование базовой модели классификатора и непосредственно доступной истории цен. Вероятность успешного прогнозирования может быть заметно улучшена при использовании адаптированной модели сети, а также благодаря предобработке входных данных [13]. Практическое значение также имеет рассмотрение расширенного набора параметров, например, возможно использование цены открытия для прогнозируемой торговой сессии [14].

СПИСОК ЛИТЕРАТУРЫ

1. Adaptive high performance classifier based on random threshold neurons / E.M. Kussul, T.N. Baidyk, V.V. Lukovich, D.A. Rachkovskij // *Cybernetics and Systems'94* / Ed. R. Trappl. –Singapore: World Scientific Publishing Co. Pte. Ltd., 1994. – P. 1687–1695.
2. Жора Д.В. Анализ функционирования классификатора со случайными порогами // *Кибернетика и системный анализ*. – 2003. – № 3. – С. 72–91; доступно: <http://rsc.netfirms.com/>.
3. Kussul E.M., Rachkovskij D.A., Wunsch D.C. The random subspace coarse coding scheme for real-valued vectors // *Proc. Int. Joint Conf. Neural Networks '99*. – 1999. – Vol. 1. – P. 450–455.
4. Trippi R.R., Turban E. *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, Rev. ed. Irwin. – 1996. – 821 p.
5. *Neural Networks in the Capital Markets*. /Ed. P.Refenes, A.P. Refenes. – John Wiley & Sons. – 1995. – 392 p.
6. Найман Э.Л. *Малая энциклопедия трейдера*. – 3-е изд. – Альпина Паблишер, 2002. – 384 с.
7. Wang X.H., Phua P.K.H., Lin W. Stock market prediction using neural networks: Does trading volume help in short-term prediction? // *Proc. Int. Joint Conf. Neural Networks '03*. – 2003. – Vol. 4. – P. 2433–2437.
8. Yao J.T., Tan C.L., Poh H.-L. Neural networks for technical analysis: a study on KLSI // *International Journal of Theoretical and Applied Finance*. – 1999. – Vol. 2. – № 2. – P. 221–241.
9. Castiglione F. Forecasting price increments using an artificial Neural Network // *Adv. Complex Systems*. – 2002. – Vol. 1. – P. 1–12.
10. Boger Z. Who is afraid of the BIG bad ANN? // *Proc. Int. Joint Conf. Neural Networks '02*. – 2002. – Vol. 3. – P. 2000–2005.
11. Яглом А.М., Яглом И.М. *Вероятность и информация*. – 3-е изд. – Москва: Наука, 1973. – 511 с.
12. Жора Д.В. Принципы построения программного агента по торговле ценными бумагами // *Проблемы программирования*. – 2004. – № 2–3. – С. 534–545.
13. Zhora D.V. Financial Forecasting using Random Subspace Classifier // *Proc. Int. Joint Conf. Neural Networks '04*. – 2004. – Vol. 4. – P. 2735–2740.
14. Lam K. P. Predictability of Intraday Stock Index // *Proc. Int. Joint Conf. Neural Networks '02*. – 2002. – Vol. 3. – P. 2156–2161.