

НЕЙРОКОМП'ЮТЕРНІ ТЕХНОЛОГІЇ ТА МЕРЕЖІ

3.1. Поняття та можливості нейрокомп'ютерних технологій.

Впровадження нових наукоємких технологій в комерційній сфері - досить непроста справа, що вимагає, окрім грошей і часу, ще і деякої зміни психології. Проте, практика показує, що ці вкладення окупаються і виводять компанію на якісно новий рівень. В такій діяльності досить часто доводиться вирішувати різні задачі, причому їх постановка неформальна, а рішення неоднозначне. Відомо, що дохід компанії так чи інакше залежить від якості вирішення цих завдань. Навіть якщо строгий алгоритмічний підхід неможливий, а отримати точне рішення принципово не можливо, існують інші ефективні способи рішення. Важливе місце серед них займають нейрокомп'ютерні технології та нейронні мережі [8, 22, 32, 35].

Нейронні мережі - це адаптивні системи для обробки та інтелектуального аналізу даних, які є математичною структурою, що імітує деякі аспекти роботи людського мозку і демонструє такі його можливості, як здібність до неформального навчання, узагальнення і кластеризації неklasифікованої інформації, здатність самостійно будувати прогнози на основі спостереження часових рядів. Головною їх відмінністю від інших методів є те, що нейромережі в принципі не потребують заздалегідь відомої моделі, а будують її самі лише на основі інформації, яку отримали. Саме тому нейронні мережі увійшли до практики усюди, де потрібно вирішувати задачі прогнозування, класифікації, управління - іншими словами, в області людської діяльності, де є задачі, що погано алгоритмізуються, для вирішення яких необхідні або постійна робота групи кваліфікованих експертів, або адаптивні системи автоматизації, якими і є нейронні мережі [39, 42, 45, 52].

Нейронна мережа приймає вхідну інформацію і аналізує її способом, аналогічним тому, що використовує наш мозок. Під час аналізу мережа вивчає (набуває досвіду і знання) і видає вихідну інформацію на основі придбаного раніше досвіду. Основне завдання аналітика, що використовує нейронні мережі для вирішення якої-небудь проблеми, - створити найбільш ефективну архітектуру нейронної мережі, тобто правильно вибрати вигляд нейронної мережі, алгоритм її навчання, кількість нейронів і види зв'язків між ними. Ця робота не має формалізованих процедур, вона вимагає глибокого розуміння різних видів архітектури нейронних мереж, включає багато дослідницької і аналітичної роботи, і може зайняти досить багато часу.

Для неформалізованих завдань нейромережеві моделі можуть на порядок перевершувати традиційні методи рішення. Але використання нейронних мереж доцільне, якщо:

- накопичені достатні об'єми даних про попередню поведінку системи;
- не існує традиційних методів або алгоритмів, що задовільно вирішують проблему;
- дані частково спотворені, частково суперечливі або не повні і тому традиційні методи видають незадовільний результат.

Нейронні мережі найкраще проявляють себе там, де є велика кількість вхідних даних, між якими існують неявні взаємозв'язки і закономірності. В цьому випадку нейромережі допоможуть автоматично врахувати різні нелінійні залежності, приховані в даних. Це особливо важливо в системах підтримки прийняття рішень і системах прогнозування. Нейромережі є незамінними при аналізі даних, зокрема, для попереднього аналізу або відбору, виявлення «випадних фактів» або грубих помилок людини, що приймає рішення. Доцільно використовувати нейромережеві методи в задачах з неповною або «зашумленою» інформацією, особливо в задачах, де рішення можна знайти інтуїтивно, і при цьому традиційні математичні моделі не дають бажаного результату.

Методи нейронних мереж можуть використовуватися незалежно або ж служити прекрасним доповненням до традиційних методів статистичного аналізу, більшість з яких пов'язані з побудовою моделей, заснованих на тих або інших припущеннях і теоретичних висновках (наприклад, що досліджуєма залежність є лінійною або що деяка змінна має нормальний розподіл). Нейромережевий підхід не пов'язаний з такими припущеннями - він однаково придатний для лінійних і складних нелінійних залежностей, особливо ж ефективний в розвідувальному аналізі даних, коли ставиться мета з'ясувати, чи є залежності між змінними. При цьому дані можуть бути неповними, суперечливими і навіть свідомо спотвореними. Якщо між вхідними і вихідними даними існує якийсь зв'язок, що навіть не виявляється традиційними кореляційними методами, то нейронна мережа здатна автоматично налаштуватися на його із заданою мірою точності. Крім того, сучасні нейронні мережі володіють додатковими можливостями: вони дозволяють оцінювати порівняльну важливість різних видів вхідної інформації, зменшувати її об'єм без втрати істотних даних, розпізнавати симптоми наближення критичних ситуацій і так далі.

Штучні нейронні мережі породжені біологією, оскільки вони складаються з елементів, функціональні можливості яких аналогічні більшості елементарних функцій біологічного нейрона. Ці елементи потім організовуються за способом, який може відповідати (або не відповідати) анатомії мозку. Не дивлячись на таку поверхневу схожість, штучні нейронні мережі демонструють деякі властивості мозку. Наприклад, вони вивчаються на основі досвіду, узагальнюють попередні прецеденти на нові випадки і витягують істотні властивості з інформації, що поступає та містить зайві дані. Проте, навіть найоптимістичніший їх захисник не передбачить, що в недалекому майбутньому штучні нейронні мережі дублюватимуть функції людського мозку. Реальний «інтелект», що демонструється найскладнішими нейронними мережами, знаходиться нижчим за рівень дощового черв'яка, і ентузіазм має бути помірний відповідно до сучасних реалій. Але рівним чином

було б невірним ігнорувати дивну схожість у функціонуванні деяких нейронних мереж з людським мозком. Ці можливості, як би вони не були обмежені сьогодні, наводять на думку, що глибоке проникнення в людський інтелект, а також безліч революційних застосувань, можуть бути не за горами.

Для повного розуміння сутності нейрокомп'ютерних технологій розглянемо як мозок обробляє інформацію. Мозок побудований з кліток двох типів: гліальних і нейронів. І хоча роль глії в його роботі досить значна, більшість дослідників вважають, що в основному розуміння роботи мозку може бути досягнуте при вивченні нейронів, об'єднаних в єдину зв'язану мережу. Ця парадигма і використовується при побудові, вивченні і вживанні штучних нейронних мереж. Слідуює, проте, відмітити, що є і інші точки зору. Зокрема, такі учені як Пенроуз і Хамерофф вважають, що головні події відбуваються не в нейронній мережі, а в самих клітинах, а саме в їх цитоскелетоні, в так званих мікротрубочках. Згідно їх точці зору, і пам'ять, і навіть свідомість визначаються конформаційними змінами білків у внутріклітинних структурах і пов'язаними з ними квантовими ефектами.

Кількість нейронів в мозку оцінюється величиною 10^{10} - 10^{11} . Типові нейрони мають тіло клітини (сому), безліч коротких розгалужених відростків - дендритів і єдиний довгий і тонкий відросток - аксон. На кінці аксон також розгалужується і утворює контакти з дендритами інших нейронів - синапси (рис. 3.1).

Внутріклітинний простір нейрона має негативний електричний потенціал по відношенню до позаклітинного (-70 mV), тобто клітина в цілому поляризована. Поляризація виникає за рахунок вибіркової проникності клітинної мембрани для іонів натрію і калія, що приводить до різниці їх концентрацій усередині і поза кліткою. Проте, якщо зовні досить сильно змінити потенціал мембрани одного нейрона (передавача) поблизу виходу аксона з його клітинного тіла, то проникність мембрани міняється і перерозподіл іонів у внутріклітинному і позаклітинному просторі аксона приводить до поширення по ньому хвилі короткочасної депольоризації.

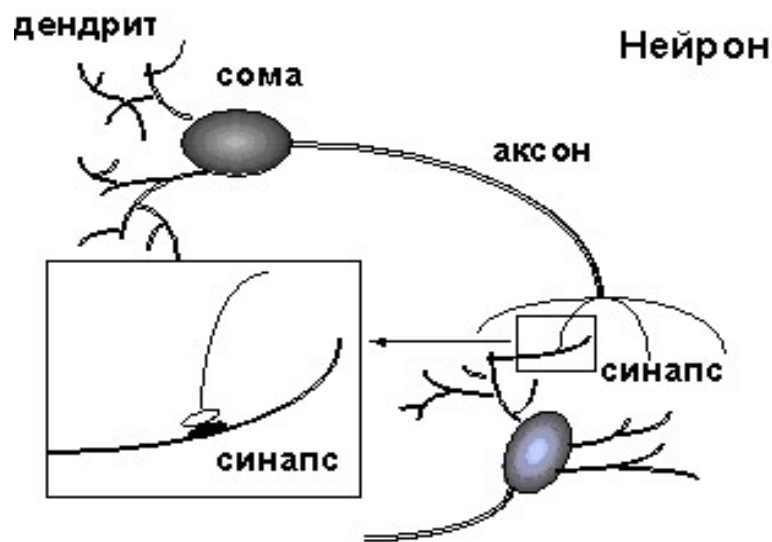


Рис. 3.1. Схема нейрона та міжнейронної взаємодії.

Електричний імпульс, поширившись по всіх галузженнях закінчення аксона з швидкістю близько 100 м/с, досягає синапсів, розташованих в місцях його контакту з дендритом або сомою інших кліток. Під впливом цього імпульсу в синапсах виділяються спеціальні хімічні речовини - нейромедіатори, які, перетинаючи синаптичну щілину, взаємодіють з мембраною нейрона-приймача і змінюють її потенціал. Таким чином дія передається від одного нейрона до інших. Відмітимо, що ця дія може бути як збуджуючою - сприяючою подальшій генерації хвилі деполяризації в нейроні-приймачі, так і такою, що інгібує - що перешкоджає такій генерації. Тип дії визначається хімічною природою нейромедіатора, що виділяється в синапсі. Після генерації імпульсу нейрон деякий час (період рефрактерності) не може активуватися. Тому частота, з якою нейрон може генерувати імпульси обмежується приблизно 100 Гц. Кожен з нейронів встановлює синаптичні зв'язки в середньому з 10^4 іншими нейронами. Тому число зв'язків в мозку оцінюється в $10^{14} - 10^{15}$. До цих пір невідомо, яким кодом користується нервова система для передачі взаємодії. Можливо, він є бінарним, і значення мають вказані стани нейронів. Можливо, важлива частота електричної активності нейронів, що кодує інтенсивність сигналу. Наприклад, у нейронів кори ця частота може бути

пропорційна вірогідності деякої події. Нарешті, інформація може міститися не в імпульсних процесах, а в повільніших змінах потенціалу мембрани, які не завжди активують клітку (тобто не перевищують порогу активації). Проте при будь-якому припущенні модель мережі взаємодіючих нейронів виявляється виключно багатою і такою, що володіє властивостями, які можна зіставити з реальними можливостями мозку [22, 32, 35].

Нескладно побудувати математичну модель вищеописанного процесу (рис. 3.2).

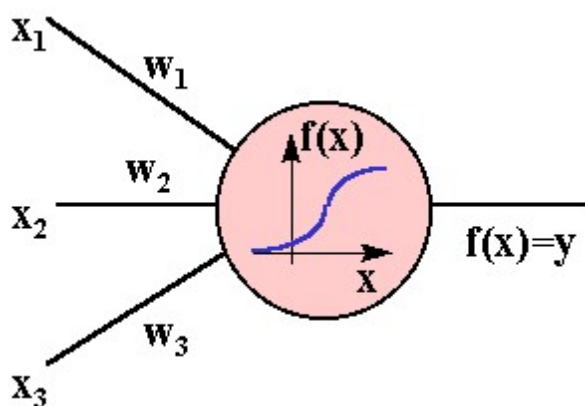


Рис. 3.2. Модель нейрона.

На рисунку зображена модель нейрона з трьома входами (дендритами), причому синапси цих дендритів мають ваги w_1, w_2, w_3 . Хай до синапсів поступають імпульси сили x_1, x_2, x_3 відповідно, тоді після проходження синапсів і дендритів до нейрона поступають імпульси w_1x_1, w_2x_2, w_3x_3 . Нейрон перетворює отриманий сумарний імпульс $x = w_1x_1 + w_2x_2 + w_3x_3$ відповідно до деякої передавальної функції $f(x)$. Сила вихідного імпульсу дорівнює $y = f(x) = f(w_1x_1 + w_2x_2 + w_3x_3)$. Таким чином, нейрон повністю описується своїми вагами w_k і передавальною функцією $f(x)$. Отримавши набір чисел (вектор) x_k як входи, нейрон видає деяке число y на виході.

Нелінійна функція $f(x)$ також називається активаційною і може мати різний вигляд, як показано на малюнку 3.3. Однією з найбільш поширених є

нелінійна функція з насиченням, так звана логістична функція або сигмоїд (функція S-образного вигляду):

$$f(x) = \frac{1}{1 + e^{-\alpha x}}$$

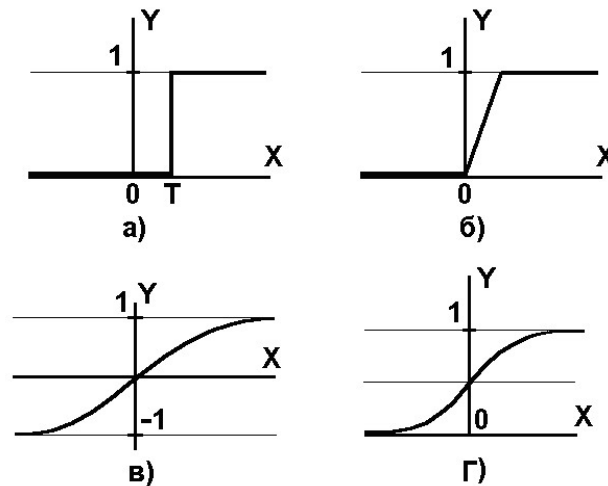


Рис. 3.3. Вигляд активаційної функції: а) функція одиничного стрибка; б) лінійний поріг (гистерезис); в) сигмоїд – гіперболічний тангенс; г) сигмоїд – функція з насиченням.

При зменшенні α сигмоїд стає пологішим і при $\alpha = 0$ вироджується в горизонтальну лінію на рівні 0.5, при збільшенні α сигмоїд наближається по зовнішньому вигляду до функції одиничного стрибка з порогом T в точці $x = 0$. З виразу для сигмоїда очевидно, що вихідне значення нейрона лежить в діапазоні $[0,1]$.

Нейрокомп'ютерні технології, як ми бачимо, зовсім не покликані замінити існуючі традиційні обчислювальні машини. Вони лише заповнюють ті можливості, для яких не вдається побудувати формальних алгоритмічних схем. Подібно до того, як в людському мозку ліва і права півкуля працюють в діалозі і спільно, сучасні інформаційні системи повинні використовувати симбіоз традиційних комп'ютерів і нейротехнологій для повноцінного і продуктивного інтелектуального аналізу інформації (рис. 3.4).

Як відомо, існує дві парадигми обробки інформації – «логічна» і «образна». Перша домінує в існуючих комп'ютерах, друга - лежить в основі роботи мозку, хоча людський мозок відрізняє від мозку інших тварин наявністю обох компонент мислення. Для повноцінного існування в навколишньому світі цінні обидва способи обробки інформації. І вони рано чи пізно виникають - як в ході біологічної еволюції, так і в процесі еволюції комп'ютерів, але, що характерно, - в різній послідовності. Біологічна еволюція йшла «від образів - до логіки». Комп'ютери ж, навпаки, почавши з логіки, лише через декілька десятиліть починають освоювати розпізнавання образів.

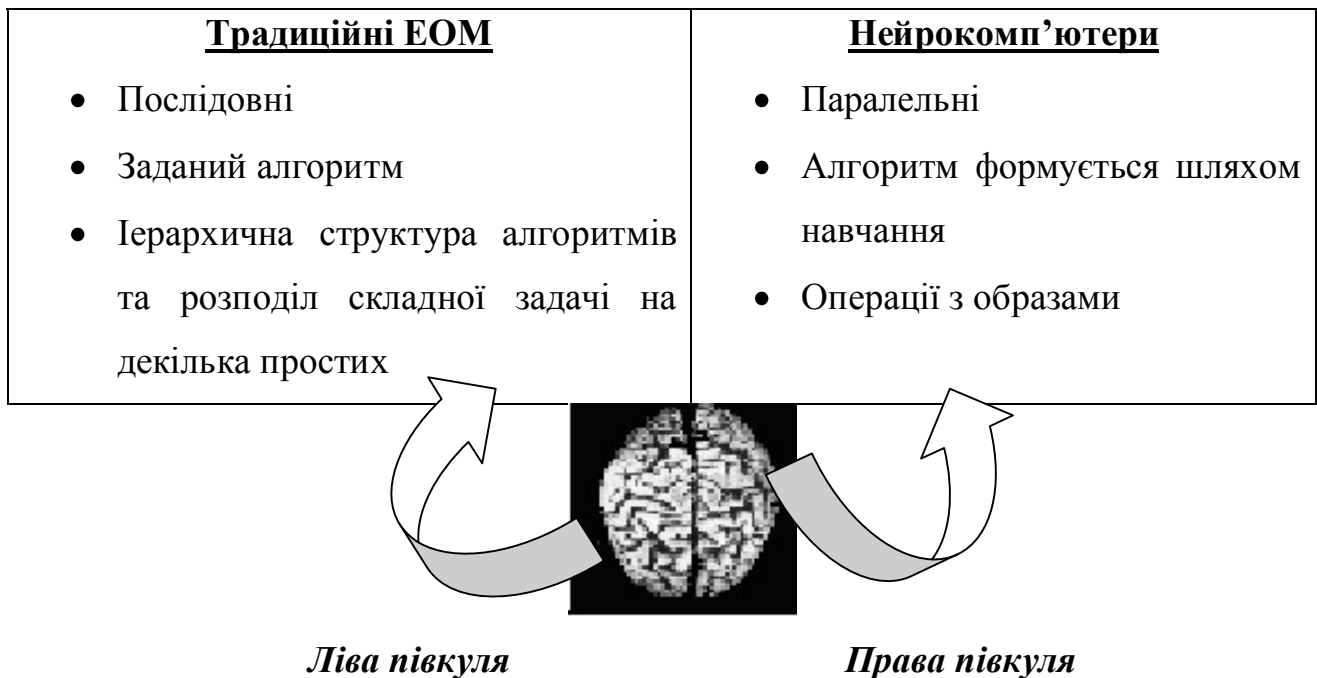


Рис. 3.4. Симбіоз традиційних та нейрокомп'ютерів як симбіоз правої та лівої півкулі головного мозку.

Вузьким місцем, що знижує ефективність людино-машинного симбіозу, є нездатність сучасних ЕОМ оперувати сенсорною або, більш загально образною інформацією. Людина, мозок якої орієнтований саме на такого роду інформацію, є зараз єдиною сполучною ланкою між світом абстрактних символів, що переробляються комп'ютерами, і зовнішнім світом. Нездатність комп'ютерів бачити, чути і відчувати не дозволяє їм звільнити людей від їх

теперішніх обов'язків «універсальних маніпуляторів» і контролерів при машинному виробництві. Відсутність сенсорного сприйняття миру комп'ютерами робить доступні їм моделі світу безпорадними. Між тим, вже зараз вартість комп'ютерної обробки інформації і вартість людського мислення майже порівнялися. У мозку людини близько 10^{10} нейронів, з яких одночасно активізоване приблизно 10^8 , працюючих з характерною частотою 10^2 Гц. Приймаючи в якості зарплати «білого комірця» \$30,000 в рік отримуємо оцінку вартості обробки інформації людиною:

- *Людина:* $(10^{10} \text{ оп/с} * 3 * 10^7 \text{ с/рік}) / (30\,000 \text{ \$/рік}) = 10^{13} \text{ оп/\$}$.

Обчислення на сучасному персональному комп'ютері продуктивністю 10^7 оп/с при амортизації близько \$300 в рік стоять лише на порядок менше:

- *Універсальний процесор:* $(10^7 \text{ оп/с} * 3 * 10^7 \text{ с/рік}) / (300 \text{ \$/рік}) = 10^{12} \text{ оп/\$}$.

Спеціалізовані процесори, як правило, дають додатковий вигравш у вартості обчислень приблизно на порядок:

- *Спеціалізований процесор:* $(10^8 \text{ оп/с} * 3 * 10^7 \text{ с/рік}) / (300 \text{ \$/рік}) = 10^{13} \text{ оп/\$}$, приблизно порівнюючись з людиною при такому способі зіставлення.

Отже, стає економічно доцільним перекласти всі рутинні людські функції на комп'ютери: все, що може бути формалізоване негайно перетворюється на програмні продукти і включається у виробничий процес. Проте, нарощування темпів комп'ютеризації натрапляє на обмежені можливості сучасних комп'ютерів в розв'язанні неалгоритмізованих задач - обробці образів. Це і є те вузьке місце, яке зараз різко звужує можливі області застосування комп'ютерів і, відповідно, - інтелектуальних аналітичних систем. Штучні нейромережі покликані «розшити» це вузьке місце, забезпечивши комп'ютерам здатність оперувати образною інформацією.

Нейрокомп'ютинг є новою парадигмою обчислювальних систем. Основне завдання нейрокомп'ютерів - обробка образів, заснована на навчанні, - те ж, що і у біологічних нейросистем. Подібно до біологічних, штучні нейромережі націлені на паралельну обробку образів. У новій схемотехніці, як і в мозку, відсутні загальні шини, немає розділення на активний процесор і

пасивну пам'ять. Обчислення, як і навчання, розподілені по всіх активних елементах - нейронах, кожний з яких є елементарний процесор образів, оскільки проводить хоч і просту операцію, але відразу над великою кількістю входів. Як обчислення, так і навчання повністю паралельні. У цьому сила природних нейрокомп'ютерів. Це дає можливість вирішувати задачі, непосильні навіть наймогутнішим суперкомп'ютерам, не дивлячись на багато кратну різницю в швидкодії елементної бази.

На думку Anil K. Jain з Мічиганського університету і фахівців Дослідницького центру IBM Jianchang Mao і K. M. Mohiuddin, список таких задач можна класифікувати таким чином [32, 35, 52].

Класифікація образів. Завдання полягає у вказівці приналежності вхідного образу (наприклад, мовного сигналу або рукописного символу), представленого вектором ознак, одному або декільком заздалегідь певним класам. До відомих додатків відносяться розпізнавання букв, розпізнавання мови, класифікація сигналу електрокардіограми, класифікація кліток крові, забезпечення діяльності біометричних сканерів і т.п.

Кластеризація/категоризація. При рішенні задачі кластеризації, яка відома також як класифікація образів «без вчителя» де відсутня навчальна вибірка з мітками класів. Алгоритм кластеризації заснований на подібності образів і розміщує близькі образи в один кластер. Відомі випадки застосування кластеризації для витягання знань, стиснення даних і дослідження властивостей даних.

Апроксимація функцій. Припустимо, що є навчальна вибірка пар даних вхід-вихід, яка генерується невідомою функцією, залежною від деякого аргументу, спотвореною шумом. Завдання апроксимації полягає в знаходженні оцінки невідомої функції. Апроксимація функцій необхідна при рішенні численних інженерних і наукових задач моделювання. Типовим прикладом є шумозаглушення при прийомі сигналу різної природи, незалежно від переданої інформації.

Прогноз. Хай задані n дискретних відліків $\{y(t_1), y(t_2), \dots, y(t_n)\}$ в послідовні моменти часу t_1, t_2, \dots, t_n . Завдання полягає в прогнозі значення $y(t_{n+1})$ в деякий майбутній момент часу t_{n+1} . Прогнози мають значний вплив на ухвалення рішень в бізнесі, науці і техніці. Прогноз цін на фондовій біржі і прогноз погоди є типовими додатками техніки прогнозу.

Оптимізація. Численні проблеми в математиці, статистиці, техніці, науці, медицині і економіці можуть розглядатися як проблеми оптимізації. Завданням алгоритму оптимізації є знаходження такого рішення, яке задовольняє системі обмежень і максимізує або мінімізує цільову функцію. Задача комівояжера, задача про призначення, транспортна задача є класичними прикладами задач оптимізації.

Пам'ять, що адресується за змістом (асоціативна пам'ять). У моделі обчислень фон Неймана звернення до пам'яті доступно тільки за допомогою адреси, яка не залежить від змісту пам'яті. Більш того, якщо допущена помилка в обчисленні адреси, то може бути знайдена абсолютно інша інформація. Асоціативна пам'ять доступна по вказівці заданого змісту. Вміст пам'яті може бути викликаний навіть по частковому входу або спотвореному змісту. Асоціативна пам'ять може чудово знайти застосування при створенні мультимедійних інформаційних баз даних.

Управління. Розглянемо динамічну систему, задану сукупністю векторів вхідної дії, поточного стану та вихідного вектора системи. Класичною постановкою задачі є розрахунок управляючої дії, щоб в конкретний момент часу система знаходилася в певній бажаній точці.

Слід відзначити, що в загальному випадку найчастіше зустрічаються ситуації, що є комплексом визначених раніше.

Розглянемо парадигми нейрокомп'ютінга, тобто, родові риси, об'єднуючі принципи роботи і навчання всіх нейрокомп'ютерів. Головне, що їх об'єднує - націленість на обробку образів. Сформулюємо ці парадигми в концентрованому вигляді безвідносно до біологічних прототипів, як способи обробки даних.

- Коннекціонізм.

Відмінною рисою нейромереж є глобальність зв'язків. Базові елементи штучних нейромереж - формальні нейрони - спочатку націлені на роботу з широкосмуговою інформацією. Кожен нейрон нейромережі, як правило, пов'язаний зі всіма нейронами попереднього шару обробки даних (рис. 3.5), що ілюструє найбільш широко поширену в сучасних додатках архітектуру багат шарового персептрона. У цьому основна відмінність формальних нейронів від базових елементів послідовних ЕОМ - логічних вентилів, що мають лише два входи.

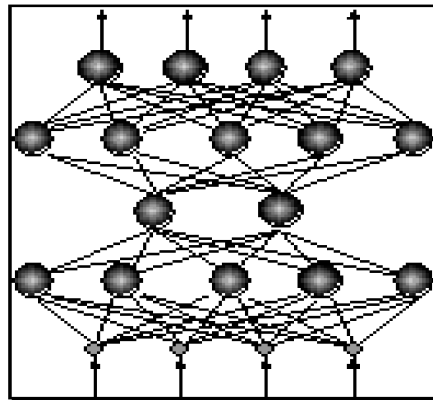


Рис. 3.5. Глобальний зв'язок в штучних нейромережах.

У результаті, універсальні процесори мають складну архітектуру, засновану на ієрархії модулів, кожний з яких виконує строго певну функцію. Навпаки, архітектура нейромереж проста і універсальна. Спеціалізація зв'язків виникає на етапі їх навчання під впливом конкретних даних.

Типовий формальний нейрон проводить просту операцію - зважує значення своїх входів з своїми ж вагами, що локально зберігаються, і проводить над їх сумою нелінійне перетворення (рис. 3.6):

$$y = f(u), \quad u = u_0 + \sum_i u_i x_i$$

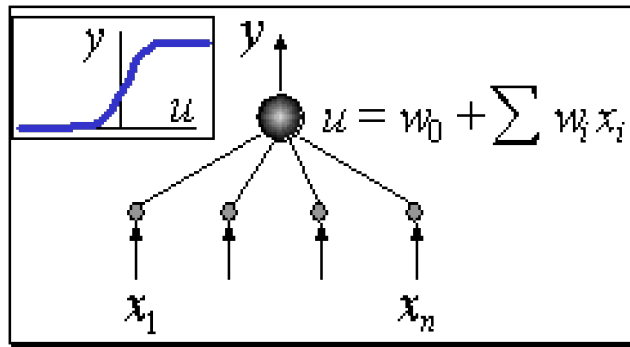


Рис. 3.6. Нелінійна операція над лінійною комбінацією входів.

Нелінійність вихідної функції активації принципова. Якби нейрони були лінійними елементами, то будь-яка послідовність нейронів також проводила б лінійне перетворення, і вся нейромережа була б еквівалентна одному нейрону (або одному шару нейронів - у разі декількох виходів). Нелінійність руйнує лінійну суперпозицію і призводить до того, що можливості нейромережі істотно вищі за можливості окремих нейронів.

- Локальність і паралелізм обчислень.

Масовий паралелізм нейрообчислювань, необхідний для ефективної обробки образів, забезпечується локальністю обробки інформації в нейромережах. Кожен нейрон реагує лише на локальну інформацію, що поступає до нього в даний момент від пов'язаних з ним таких же нейронів, без апеляції до загального плану обчислень, звичайною для універсальних ЕОМ. Таким чином, нейромережеві алгоритми локальні, і нейрони здатні функціонувати паралельно.

- Програмування: навчання, засноване на даних.

Відсутність глобального плану обчислень в нейромережах припускає і особливий характер їх програмування. Воно також носить локальний характер: кожен нейрон змінює свої параметри - синаптичні ваги - відповідно до локальної інформації, що поступає до нього, про ефективність роботи всієї мережі як цілого. Режим розповсюдження такої інформації по мережі і відповідної до неї адаптації нейронів носить характер навчання. Такий спосіб програмування дозволяє ефективно врахувати специфіку потрібного від мережі способу обробки даних, бо алгоритм не задається наперед, а породжується

самими даними - прикладами, на яких мережа навчається. Саме таким чином в процесі самонавчання біологічні нейромережі виробили такі ефективні алгоритми обробки сенсорної інформації.

Характерною особливістю нейромереж є їх здібність до узагальнення, що дозволяє навчати мережу на нікчемній частці всіх можливих ситуацій, з якими їй, можливо, доведеться зіткнутися в процесі функціонування. У цьому їх різюча відмінність від звичайних ЕОМ, програма яких повинна наперед передбачати їх поведінку у всіх можливих ситуаціях. Ця ж їх здатність дозволяє кардинально здешевити процес розробки додатків.

- Універсальність навчальних алгоритмів.

Привабливою рисою нейрокомп'ютінга є єдиний принцип навчання нейромереж - мінімізація емпіричної помилки. Функція помилки, що оцінює дану конфігурацію мережі, задається ззовні - залежно від того, яку мету переслідує навчання. Але далі мережа починає поступово модифікувати свою конфігурацію - стан всіх своїх синаптичних ваг - так, щоб мінімізувати цю помилку. У результаті, в процесі навчання мережа все краще справляється з покладеним на неї завданням.

Не вдаючись до математичних тонкощів, образно цей процес можна уявити собі як пошук мінімуму функції помилки $E(w)$, залежної від набору всіх синаптичних ваг мережі w (рис. 3.7).

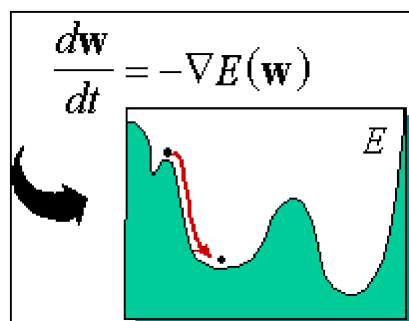


Рис. 3.7. Навчання мережі як задача оптимізації.

Базовою ідеєю всіх алгоритмів навчання є облік локального градієнта в просторі конфігурацій для вибору траєкторії якнайшвидшого спуску по функції помилки. Функція помилки, проте, може мати безліч локальних мінімумів, що представляють суб-оптимальні рішення. Тому градієнтні методи зазвичай доповнюються елементами стохастичної оптимізації, щоб запобігти застряганню конфігурації мережі в таких локальних мінімумах. Ідеальний метод навчання повинен знайти глобальний оптимум конфігурації мережі.

Вписавши появу нейрокомп'ютинга в загальний процес еволюції комп'ютерів, ми дістаємо можливість заглянути в найближче майбутнє - екстраполюючи сьогоденні тенденції.

1. «Розумні» нейрочіпи.

Сьогоднішній нейрокомп'ютинг проходить «обкатку», в основному, в програмному продукті для задач асоціативної обробки даних, рідко використовуючи при цьому свій «паралельний» потенціал. Епоха паралельного нейрокомп'ютинга почнеться з виходом на ринок широкого асортименту апаратних засобів - спеціалізованих нейрочіпів для обробки зображень, мови, аналітики та сенсорної інформації. Вже сьогодні існують, наприклад, дверні системи, що розпізнають господаря по вигляду, голосу, і мабуть, запаху в сукупності. Системи життєзабезпечення житла стануть адаптивними і зможуть навчатися. Всі побутові прилади порозумнішають і придбають здатність вгадувати, що від них вимагається саме в даний момент. Сенсорні датчики придбають здатність реагувати, а регулюючі системи - відчувати. Розумні контролери, що розпізнають потенційно небезпечні ситуації і що уміють приймати адекватні превентивні рішення, набудуть поширення в складних електричних і теплових мережах. На них ґрунтуватимуться системи регулювання транспортними потоками і потоками даних в комп'ютерних мережах і стільниковому зв'язку.

2. Операційні системи нової архітектури.

Проте, найбільші зміни торкнуться самих комп'ютерів. На думку Біла Гейтса, глави відомої компанії Microsoft, висловленому їм на зборах ради

директорів, через 10 років 90% операційних систем буде зайнято вирішенням задач розпізнавання образів. Таким чином, при проектуванні майбутніх поколінь комп'ютерів нейрокомп'ютинг висувається на перший план. Можна навіть уявити собі зразковий сценарій проникнення нейросистем в комп'ютери майбутнього, пов'язаний з розвитком глобальної мережі Internet. Зараз саме вона направляє розвиток комп'ютерних систем, поступово перетворюючи розрізнену мережу персоналок, робочих станцій і мейнфреймів в єдиний світовий мережевий комп'ютер з необмеженими інформаційними ресурсами. І подібно до того, як епоха великих відкриттів XV століття стимулювала розвиток астрономії і точної механіки для вдосконалення навігаційних приладів, освоєння нового інформаційного океану вимагає розвитку нових засобів навігації - асоціативного пошуку, створення адаптивних і автономних агентів. Новий інтерфейс користувача ґрунтуватиметься на агентах, що представляють його інтереси в мережі. Цей новий вигляд програмного забезпечення, що отримав назву agentware, претендує на центральне місце в майбутній системі людино-машинного спілкування. Тим часом, перші екземпляри agentware вже з'явилися на ринку, і що характерно, багато з них засновані на технології нейромереж. Це, мабуть, сьогодні найкоротша дорога до створення легко навчаємих автономних електронних секретарів. Природно передбачити, що саме на цьому напрямі, через його стратегічну важливість, в найближчому майбутньому буде досягнутий найбільший прогрес.

Розглянемо сучасний стан нейрокомп'ютингу та нейротехнологій.

- Елементна база нейрокомп'ютерів.

Практично всі нейрокомп'ютери, що діють, використовують традиційну елементну базу: мікроелектронні СБІС. Сотні мільярдів доларів, які вже вкладені в розвиток цієї технології, дають їй вирішальну перевагу перед іншими альтернативами, такими, як, наприклад, оптичні обчислення. Сучасна електроніка спирається, в основному, на цифрову обробку сигналів, стійку до перешкод і технологічних відхилень в параметрах базових елементів. Цифрова схемотехніка надає нейро-конструкторам найбільш багатий інструментарій.

Тому недивно, що найбільшого поширення набули саме цифрові нейрокомп'ютери. Це по суті - спеціалізовані матричні прискорювачі, що використовують матричний, пошаровий характер обробки сигналів в нейромережах. Широко використовуються стандартні процесори обробки сигналів (DSP - Digital Signal Processors), оптимізовані під такі операції. Прикладом сучасного DSP-процесора, пристосованого для прискорення нейро-обчислень є продукт Texas Instruments TMS320C80 продуктивністю 2 млрд. операцій в секунду. Цей кристал включає п'ять процесорів і реалізує відразу дві технології - DSP і RISC (рис. 3.8).

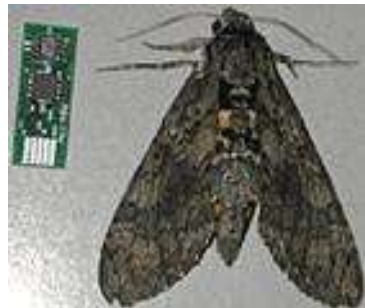


Рис. 3.8. Сучасні цифрові нейрокомп'ютерні елементи.

Проте, сама природа нейромережевої обробки інформації - аналогова, і додаткового виграшу в швидкості обчислень (по деяких оцінках 10^3 - 10^4) та щільності обчислювальних елементів можна добитися, використовуючи спеціалізовану аналогову елементну базу. Найбільш перспективні, мабуть аналогові мікросхеми з локальними зв'язками між елементами (т.з. клітинні нейромережі, CNN - Cellular Neural Networks), наприклад силіконова ретина фірми Synaptics. З іншого боку, розробка аналогових чіпів з використанням нетрадиційних рішень схемотехнік вимагає додаткових і чималих витрат. В даний час ці роботи розгорнуті широким фронтом, наприклад, в рамках проекту SCX-1 (Silicon Cortex - кремнієва кора) (рис. 3.9).

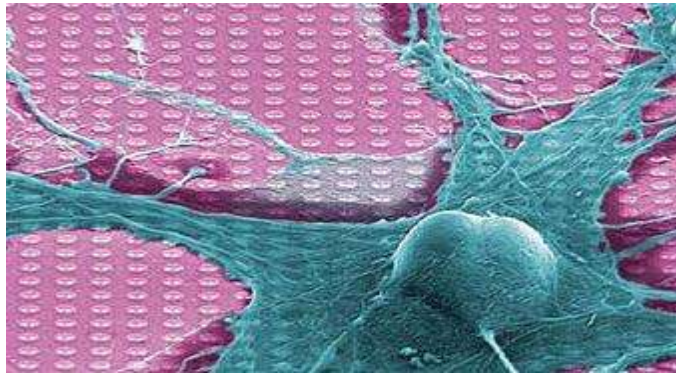


Рис. 3.9. Сучасні аналогові нейрокомп'ютерні елементи.

Переваги обох підходів намагаються поєднати гібридні мікросхеми, що мають цифровий інтерфейс з апаратурою, але виконують найбільш масові операції аналоговим способом.

- Архітектура нейрокомп'ютерів.

Перевага нейрокомп'ютинга полягає в можливості організувати масові паралельні обчислення. Тому базові процесорні елементи зазвичай поєднуються в обчислювальні комплекси: якомога більше - на одному чіпі, а що не помістилося - в мультипроцесорні плати. Ці плати потім або вставляють в персональні комп'ютери і робочі станції як нейро-прискорювачі, або збирають в повномасштабні нейрокомп'ютери.

- Порівняння вартості звичайних і нейрообчислень.

Продуктивність сучасних персональних комп'ютерів складає приблизно 10^7 операцій. Отже, при вартості всього на порядок більше звичайних РС, нейро-прискорювач в декілька сотень разів перевершує їх в швидкості. Таким чином, питома вартість сучасних нейрообчислень приблизно на порядок нижча, ніж в традиційних комп'ютерів. Це всього лише наслідок спеціалізації матричних процесорів (DSP), що мають ту ж елементну базу, що і універсальні мікропроцесори. Проте, вигреш на один порядок у вартості обчислень рідко коли здатний стати вирішальним аргументом для використання спеціалізованої апаратури, зв'язаної з додатковими витратами, у тому числі на навчання персоналу. Тому реально нейрокомп'ютери використовуються в

спеціалізованих системах, коли потрібно навчати і постійно перенавчати сотні нейромереж, об'єднаних в єдині інформаційні комплекси, або в системах реального часу, де швидкість обробки даних критична.

- Нейроемулятори.

Більшість же прикладних систем нейромережевого аналізу даних використовують емуляцію нейромереж на звичайних комп'ютерах, зокрема на РС. Такі програми отримали назву нейроемуляторів. Доступність і збільшені обчислювальні можливості сучасних комп'ютерів привели до широкого поширення програм, що використовують принципи нейромережевої обробки даних, але виконуються на послідовних комп'ютерах. Цей підхід не використовує переваг властивого нейрообчисленням паралелізму, орієнтуючись виключно на здатність нейромереж вирішувати завдання, що не формалізуються .

Переваги таких «віртуальних» нейрокомп'ютерів для відносно невеликих завдань очевидні. По-перше, не треба витратитися на нову апаратуру, якщо можна завантажити вже наявні комп'ютери загального призначення. По-друге, користувач не повинен освоювати особливості програмування на спецпроцесорах і способи їх поєднання з базовим комп'ютером. Нарешті, універсальні ЕОМ не накладають жодних обмежень на структуру мереж і способи їх навчання, тоді як спецпроцесори частенько мають обмежений набір «захитих» в них функцій активації і досягають пікової продуктивності лише на певному колі задач.

- Готові нейропакети.

Це закінчені незалежні програмні продукти, призначені для широкого класу задач, в основному - для передбачень і статистичної обробки даних. Більшість з нейропакетів, що є на ринку, має дружній інтерфейс користувача, що не вимагає знайомства з мовами програмування. Безліч нейроемуляторів початкового рівня можна знайти в Internet як shareware або freeware. Це, зазвичай, багатошарові перцептрони з одним або декількома правилами навчання. Виключення складає повністю професійний Штутгартський

симулятор з великим набором можливостей, що працює, правда, лише на UNIX-машинах. Комерційні пакети відрізняються від вільно поширюваних великим набором засобів імпорту і передобробки даних, додатковими можливостями по аналізу значущості входів і оптимізації структури мережі. Як правило, такі пакети (BrainMaker Professional, NeuroForecaster, Лора-IQ300) мають власний вбудований блок передобробки даних, хоча інколи для цієї мети зручніше використовувати стандартні електронні таблиці. Так, нейро-продукти групи нейрокомп'ютинга ФІАН вбудовується безпосередньо в Microsoft Excel як спеціалізовані функції обробки даних. При цьому всю передобробку даних і візуалізацію результатів можна проводити стандартними засобами Excel, який, крім того, має багатий і розширюваний набір конверторів для імпорту і експорту даних. Такі пакети націлені на вирішення інформаційних задач в діалоговому режимі - при безпосередній участі користувача.

- Інструменти розробки нейрододатків.

Головне, що відрізняє цей клас програмного забезпечення - здатність генерувати «відчуждані» нейромережеві продукти, тобто генерувати програмний код, що використовує навчені нейромережі для аналізу даних. Такий код може бути вбудований як підсистема в будь-які скільки завгодно складні інформаційні комплекси. Прикладами подібних систем, здатних генерувати вихідні тексти програм є NeuralWorks Professional II Plus фірми NeuralWare і Neural Bench. Остання цікава, окрім іншого, тим, що може генерувати коди на багатьох мовах, включаючи Java. Такі Java-апплеты можуть використовуватися для організації різного роду сервісів в глобальних і локальних мережах. Зручним інструментом розробки складних нейросистем є MATLAB з нейромережевим інструментарієм, що додається до нього. MATLAB надає зручне середовище для синтезу нейромережевих методик з іншими методами обробки даних (wavelet-анализ, статистика, фінансовий анализ і так далі).

- Готові рішення на основі нейромереж.

В таких системах нейромережі заховані від користувача в надрах готових автоматизованих комплексів, призначених для вирішення конкретних аналітичних завдань. Наприклад, вже згадуваний продукт Falcon вбудовується в банківську автоматизовану систему обслуговування платежів за пластиковими картками. У іншому випадку це буде автоматизована система управління заводом або реактором. Кінцевого користувача, як правило, не цікавить спосіб досягнення результату, йому важлива лише якість продукту. Оскільки багато таких готових рішень володіють унікальними можливостями і забезпечують реальні конкурентні переваги, їх ціна може бути досить висока. Іспанська компанія SEMP займається підвищенням ефективності обслуговування кредитних карт VISA, що емітуються іспанськими банками. Кількість подібних транзакцій - від 500000 до 1000000 в день. Нейромережева система, розроблена для неї вченими з Мадридського Інституту Інженерії Знань (Instituto de Ingenieria del Conocimiento), зменшила вірогідність несанкціонованого використання карт на 30-40% для основних каналів шахрайства.

- Нейромережевий консалтинг.

Опис ринку нейро-продуктів буде не повним без згадки про нейро-консалтинг. Замість того, щоб продавати готові програми або інструменти для їх розробки, можна торгувати і послугами. Деякі завдання, наприклад такі, як передбачення ринкових часових рядів, є настільки складними, що доступні лише справжнім професіоналам. Не кожна компанія може дозволити собі витрати, що асоціюються з передовими науковими розробками. Тому набувають популярності фірми, єдиною продукцією яких є передбачення ринків. При великому числі клієнтів ціна таких передбачень може бути вельми помірною. Прикладом тут може служити Prediction Company, заснована в 1991 році фізиками Дойном Фармером і Норманом Паккардом, - фахівцями в області динамічного хаосу. Продукція компанії користується великим успіхом серед Швейцарських банків, що скуповують прогнози для гри на фондових і валютних ринках.

- Ринок нейропродукції.

Об'єм ринку нейропродукції, структуру зростає стрімкими темпами: по різних оцінках - від 30% до 50% в рік, переваливши недавно за декілька мільярдів доларів. Таке ж зростання спостерігалось на початку 80-х років на ринку персональних комп'ютерів. Мільярдні доходи тоді з'явилися сигналом для вступу на цей ринок гіганта комп'ютерної індустрії - ІВМ. Всі ми прекрасно пам'ятаємо як в результаті виниклої конкуренції за гроші кінцевого користувача перетворився весь комп'ютерний світ. Нейрокомп'ютери і їх програмні емулятори, природно, цікаві не самі по собі, а як інструмент вирішення практичних завдань. Лише в цьому випадку нейропродукція володітиме споживчою вартістю і мати відповідний об'єм ринку.

Є багато вражаючих демонстрацій можливостей штучних нейронних мереж. У деяких областях, таких як виявлення фальсифікацій і оцінка ризику, вони стали безперечними лідерами серед використовуваних методів. Їх застосування в системах прогнозування і системах маркетингових досліджень постійно росте. Розглянемо деякі найбільш характерні приклади практичного використання нейронних мереж [41, 53, 70, 77].

Контроль операцій з кредитними картками. Здібності нейромереж до класифікації застосовуються для відстежування операцій з краденими кредитними картами і підробленими чеками. Спеціалізована система Falcon фірми HNC дозволяє по частоті операцій і характеру покупок виділити підозрілі операції і сигналізувати про це. Ця система використовує запатентовану уніфіковану технологію підтримки ухвалення рішення (Unified Decisions Technology), яка комбінує розширену базу даних правил обробки транзакцій, статистичний аналіз і нейронну мережу. Система Falcon також містить спеціальний компонент, що забезпечує можливість досвідченим фахівцям включати в базу даних системи правил, що дозволяють з високою мірою достовірності визначати потенційні випадки шахрайства з банківськими картами по географічному місцю розташування або поштовому індексу отримувача карти. Практичне використання цієї системи показало підвищення показників якості виявлення на 20–60% при значному зниженні помилкових

спрацьовувань. Завдяки цій системі втрати банків від таких операцій помітно зменшилися. В даний час Falcon контролює більше 260 мільйонів рахунків 16 найбільших емітентів кредитних карт.

Аналогічна система, яка розроблена фірмою ІТС, використовується для обробки операцій з кредитними картами Visa. У 2005 році за допомогою цієї системи запобігли нелегальні операції на суму більше 100 млн. доларів.

Сімейство систем PRISM, розроблене компанією Nestor, засноване на використанні нейромереж, експертних систем і статистичних методів для виявлення в реальному режимі часу шахрайства з кредитними і дебетовими картами, а також для виявлення інших типів шахрайства при здійсненні фінансових або торгівельних операцій. Нейромережа, використовувана в системах сімейства PRISM, була навчена на основі більш ніж напівмільйона транзакцій з різними типами карт. По деяких оцінках, використання систем сімейства PRISM дозволяє зменшити число шахрайств на 50 %.

Нейромережі на фінансових ринках. Американський Citibank використовує нейромережеві передбачення з 1990 року. У 2002 році, за свідченням журналу The Economist, автоматичний ділінг показував прибутковість 25% річних, що набагато перевищує показники більшості брокерів. Chemical Bank використовує нейро-систему фірми Neural Data для попередньої обробки транзакцій на валютних біржах 23 країн, фільтруючи «підозрілі» операції. Fidelity of Boston використовує нейромережі при управлінні портфелями з сумарним об'ємом 3 мільярда доларів. Повністю автоматизовані системи ведення портфелів з використанням нейромереж застосовують, наприклад, Deere & Co - на суму \$100 млн і LBS Capital - на суму \$400 млн. У останньому випадку експертна система об'єднується з приблизно 900 нейромережами.

Компанія Alela Corp. займається прогнозуванням зміни біржових індексів. Для передбачення знаків зміни індексів застосовується нейронна мережа, використовуюча РБФ (радіальні базисні функції). На сайті компанії можна безкоштовно отримати прогнози зміни індексів Dow Jones, S&P500 і

Merval, а також переконатися, що доля вірних передбачень складає не менше 80%.

Аналіз страхових позовів. *Фірмою Neural Innovation Ltd. створена нейромережева система Claim Fraud Analyser, що дозволяє миттєво виявляти підозрілі страхові позови, що відносяться до пошкоджених автомобілів. На входи системи подаються такі параметри, як вік і досвід водія, вартість автомобіля, наявність подібних випадків у минулому та інші. В результаті обробки видається число - вірогідність того, що даний позов пов'язаний з шахрайством. Така система дозволяє не лише заощадити за рахунок виявлення фальсифікацій, але і поліпшити стосунки з клієнтами за рахунок швидшого задоволення чесних позовів.*

Аналіз споживчого ринку. Кілька років тому фірма IBM Consulting виконала замовлення на створення нейромережевої системи, що прогнозує властивості споживчого ринку. Замовник - один з найбільших виробників харчових продуктів, що має величезні ринки збуту. Одним з основних маркетингових механізмів замовника є поширення купонів, що дають право покупки певного товару із знижкою. Так як витрати на розсилку купонів досить великі, вирішальним чинником є ефективність розсилки, тобто доля клієнтів, що скористалися знижкою. Для підвищення ефективності купонної системи важливо було провести попередню сегментацію ринку, а потім адресувати клієнтам кожного сегменту саме ті купони, якими вони з більшою вірогідністю скористаються. В термінах інтелектуального аналізу даних потрібно було вирішити задачу кластеризації, що і було успішно зроблено за допомогою мереж Кохонена. На другому етапі для споживачів кожного з кластерів підбиралися відповідні комерційні пропозиції, а потім будувався прогноз об'єму продажів для кожного сегменту.

Інший варіант рішення цієї ж задачі вибрала компанія GoalAssist Corporation, виконуючи замовлення крупної маркетингової фірми. Потрібно було дослідити стратегію заохочувальних товарів (коли, наприклад, надіславши 5 етикеток від чіпсів, клієнт отримує безкоштовно футболку) для певної

компанії, яка торгувала харчовими продуктами. Звичайні методи прогнозування відгуку споживачів виявилися в даному випадку недостатньо точні. В результаті попит на футболки виявився дуже великий і багатьом покупцям довелося довго чекати здобуття призу, тоді як інші дарунки залишилися незатребуваними. Аби підвищити точність прогнозування, було вирішено використовувати історичні дані і нейронні мережі. Компанія GoalAssist Corporation побудувала дві нейромережі для вирішення цього завдання. Перша з них - це мережа з адаптивною архітектурою пакету NeuroShell Classifier компанії Ward Systems Group, на вході якої подавалися різні параметри товарів і рекламної політики. За допомогою цієї мережі, призначеної спеціально для класифікації, було отримано розділення входів на 4 класи, що характеризують відгук споживачів. Ті ж вході разом з відповіддю першої мережі подавалися далі на вхід пакету NeuroShell Predictor, який також містить складну мережу пристосовану для задач кількісного прогнозування. Середня помилка передбачень складала всього біля 4%. Побудова цієї моделі зайняла близько 120 годин, також був потрібен час на передобробку вхідних даних. Експерти GoalAssist Corp. вважають, що ця модель і далі успішно застосовуватиметься для вирішення маркетингових задач.

Компанія Neural Innovation Ltd. використовує при роботі з маркетинговими компаніями конкретну стратегію прямої розсилки. Спочатку розсилається 25% від загального числа пропозицій і збирається інформація про відгуки споживачів. Потім ця інформація поступає на вхід нейрокомп'ютера, який здійснює пошук оптимального сегменту споживчого ринку для даного товару. Останні 75% пропозицій розсилаються у вказаний сегмент. При цьому ефективність розсилки істотно зростає.

Dr. Al Behrens, співробітник Northern Natural Gas в штаті Небраска, навчив нейронну мережу, яка передбачає зміну цін на газ в наступному місяці з середньою точністю 97 %. Щомісячна ціна інколи прив'язується до індексації цін в друкарських виданнях (Inside FERC і Natural Gas Week), включаючи недавню ринкову діяльність компанії, сезонні чинники, погоду і так далі.

Відома корпорація Microsoft також використовує програмні продукти на основі нейромереж в своїй маркетинговій політиці. Щороку Microsoft надсилає більше 40 млн. рекламних пропозицій про покупку своїх продуктів більш ніж 8,5 млн. зареєстрованим користувачам. Велика частина направлена на модернізацію ПО або покупку додаткових пакетів. Мета використання ПО на основі нейронних мереж очевидна - збільшення обсягу продажу. Такими засобами було досягнуте збільшення попиту з 4,9 % до 8,2 %, а відмінність у витратах на рекламу знизилася на 35 %.

Виробництво. Нейронна мережа, застосована на заводі Intel, здатна ідентифікувати брак на виробництві. Спочатку системі давали електричну випробувальну інформацію від готових чіпів і відповідних змінних управління виробничим процесом. Стосунки між цими двома параметрами були визначені числовим експериментом і моделюванням процесу CMOS. Чутлива поверхнева модель використовувалася для збереження результатів достатньої кількості числових експериментів. В результаті нейронна мережа була здатна забракувати нефункціонуючий чіп з точністю 99,5%.

При проведенні випробувань якості бетону використовується велика кількість методів. Одним з них є буріння у пошуках порожнини, що утворилася. Проте за допомогою нейромереж можливо перевірити весь матеріал, а також визначити глибину, на якій знаходиться порожнина. Шляхом подачі звукових хвиль і прийому відбитого сигналу, а потім обробкою нейронною мережею, фахівці з National Institute of Standards and Technology (NIST) здатні перевірити якість бетону при товщині матеріалу до півметра.

Ford Motors Company впровадила у себе нейросистему для діагностики двигунів після невдалих спроб побудувати експертну систему, оскільки хоча досвідчений механік і може діагностувати несправності він не в змозі описати алгоритм такого розпізнавання. На вхід нейро-системи подаються дані від 31 датчика. Нейромережа навчалася різним видам несправностей по 868 прикладам. Після повного циклу навчання якість діагностування несправностей

мережею досягла рівня кращих експертів, і значно перевершувала їх в швидкості.

Медична діагностика. Компанією «НейроПроект» створена система об'єктивної діагностики слуху у грудних дітей. Загальноприйнята методика об'єктивної діагностики полягає в тому, що в процесі обстеження реєструються «викликані потенціали» (відгуки мозку) у відповідь на звуковий подразник, що виявляються у вигляді сплесків на електроенцефалограмі. Для досить упевненої діагностики слуху дитини досвідченному експерту-аудиологу необхідно провести близько 2000 тестів, що займає близько години. Нейромережа здатна з тією ж достовірністю визначити рівень слуху вже за 200 спостереженнями протягом всього декількох хвилин, причому без участі кваліфікованого персоналу.

Ще одним прикладом використання нейронних мереж в програмах медичної діагностики служить пакет кардіодіагностики, розроблений фірмою RES Informatica спільно з Центром кардіологічних досліджень в Мілані. Причому для таких хвороб, як ішемія міокарду і артеріальна гіпертензія, досягається точність постановки діагнозу більш ніж 95 %. Окрім цього, одним з перспективних напрямів в дослідженнях є онлайн-постановка діагнозу. Пацієнт заповнює форму, вказуючи необхідні параметри, а система на основі навчальної вибірки і накопиченої бази даних визначає діагноз. За деякими даними, це дозволить збільшити продуктивність більш ніж в 1000 разів, здешевити дослідження не менше чим в 500 разів, обробити дані від практично необмеженої кількості хворих, при збільшенні діагностичної точності.

Системи безпеки в аеропортах. Американська фірма SAIC (Science Application International Corporation) використовувала нейронні мережі в своєму проекті TNA. TNA є ящиком вартістю \$750.000, який здатний виявляти пластикову вибухівку в запакованому багажі. TNA бомбардує багаж повільними нейтронами, що викликають вторинне гамма-випромінювання, спектр якого аналізується нейронною мережею. Система виявляє вибухівку з вірогідністю вище 97% і переглядає 10 місць багажу в хвилину.

Активна реклама в Internet. Нейромережевий продукт SelectCast фірми Aptex Software Inc. виявляє профілі інтересів користувачів Internet і пропонує їм відповідним чином відфільтровану рекламу. Після установки на серверах Excite і Infoseek, нейромережева реклама охопила біля третини всіх користувачів Internet. Згідно проведеним дослідженням, встановлено, що відгук на таку активну рекламу в середньому удвічі вище, ніж на звичайну рекламу, що розміщується в Мережі. А на окремі види реклами відгук зріс вп'ятеро. Відмітимо, що рекламний сектор Internet переживає зараз період бурхливого розвитку. Результати першого півріччя 2007 років свідчать про річний темп зростання 250%, що в грошовому вираженні складає \$400 млн.

Політичні технології. Компанія «НейроПроект» досить упевнено передбачає результати президентських виборів в США на підставі анкети з 12 питань. Причому, аналіз навченої нейромережі дозволив виявити п'ять ключових питань, відповіді на яких формують два головні чинники, що визначають успіх президентської кампанії.

Наукові дослідження. Доктор Henrik Lundstedt з Lund Observatory, Швеція, навчив нейронні мережі прогнозувати ефекти від сонячних спалахів, такі як обурення магнітних полів Землі. Як відомо, т.з. магнітні бурі роблять вплив не лише на стан і здоров'я людини, під їх вплив потрапляють і електростанції, трапляються збої в передачі радио- і телепередач, збої в роботі геологічного устаткування, супутників і так далі. Нейронна мережа, що враховує 37 відомих впливаючих чинників раз в чотири дні і аналізує їх зміни, здатна з високою мірою точності визначити «космічну погоду». Відомий випадок, коли стандартний метод прогнозування не визначив сильних магнітних бурь взагалі, а метод на основі нейронних мереж в той же час точно спрогнозував дві з трьох бурь.

Автомітований гіперзвуковий літак-розвідник. Названий LoFLYTE (Low-Observable Flight Test Experiment) реактивний безпілотний літак завдовжки 2,5 м був розроблений для NASA і Air Force фірмою Accurate Automation Corp. в рамках програми підтримки малого інноваційного бізнесу.

Це експериментальна розробка для дослідження нових принципів пілотування, включаючи нейронні мережі, що дозволяють автопілоту навчатися, копіюючи прийоми пілотування льотчика. З часом нейромережі переймають досвід управління, а швидкість обробки інформації дозволить швидко знаходити вихід в екстремальних і аварійних ситуаціях. LoFLYTE призначений для польотів з дуже високою швидкістю, коли швидкості реакції пілота може не вистачити для адекватного реагування на зміни режиму польоту.

3.2. Архітектура нейронних мереж.

Ідеї нейрокомп'ютинга з'явилися практично одночасно із зародженням послідовних ЕОМ. Основні роботи по нейро-обчисленнях з'явилася на два роки раніше знаменитої доповідної записки фон Неймана про принципи організації обчислень в послідовних універсальних ЕОМ. Вияв цікавості до штучних нейронних мереж був обумовлений роботами піонерів у цій справі - У. Маккалоха і У. Піттса. У 1943 році увагу громадськості привернула робота під назвою «Логічне числення ідей, що відносяться до нервової діяльності», в якій вони запропонували математичну модель нейрона і сформулювали принципи побудови штучних нейронних мереж, згідно розробленої ними моделі функціонування головного мозку. Багато учених з ентузіазмом почали пропонувати свої рішення і нову архітектуру нейронних мереж [32, 33, 52].

Паралельно з прогресом в нейроанатомії і нейрофізіології психологами були створені моделі людського навчання. Одній з таких моделей, що виявилася найбільш плідною, була модель Д. Хебба, який в 1949 році запропонував закон навчання, що виявився стартовою точкою для алгоритмів навчання штучних нейронних мереж. Доповнений сьогодні безліччю інших методів він продемонстрував вченим того часу, як мережа нейронів може навчатися.

Перший експериментальний нейрокомп'ютер Snark був побудований Марвіном Мінським в 1951 році. Проте, він не був пристосований до вирішення практично цікавих задач, і перший успіх нейрокомп'ютинга пов'язують з розробкою іншого американця - Френка Розенблатта - перцептроном (від англійського perception - сприйняття). Перцептрон був вперше змодельований на універсальній ЕОМ IBM-704 в 1958 році, причому його навчання вимагало біля півгодини машинного часу. Апаратний варіант - Mark I Perceptron, був побудований в 1960 році і призначався для розпізнавання зорових образів. Його рецепторне поле складалося з 400 пікселів (матриця фотоприймачів 20x20), і він успішно справлявся з вирішенням ряду задач, зокрема міг розрізняти деякі букви (рис. 3.10).



Рис. 3.10. Френк Розенблатт з Mark I Perceptron.

Мінський, Розенблатт, Уїдроу та інші розробили мережі, що складаються з одного шару штучних нейронів. Часто звані перцептронами, вони були використані для такого широкого класу задач, як передбачення погоди, аналіз електрокардіограм і штучний зір. Протягом деякого часу здавалося, що ключ до інтелекту знайдений і відтворення людського мозку є лише питанням конструювання досить великої мережі. Але ця ілюзія скоро розсіялася. Мережі не могли вирішувати завдання, зовні вельми схожі з тими, які вони успішно вирішували. З цих нез'ясовних невдач почався період інтенсивного аналізу. Мінський, використовуючи точні математичні методи, строго довів ряд теорем,

що відносяться до функціонування мереж. Його дослідження привели до написання книги, в якій він разом з Пайпертом довів, що використовувані у той час одношарові мережі теоретично нездійсненні вирішити багато простих завдань. Мінський також не був оптимістичний відносно потенційно можливого прогресу.

Їх робота остудила науковий запал багатьох учених-ентузіастів практично на два десятиліття. Утворилося деяке затишшя, пов'язане з розчаруванням в можливостях нейронних мереж, викликаним не в останню чергу психологічним чинником, що виявляється в нездатності людини описати словами те, як він думає. І хоча в цей період було запропоновано багато заслугуючих уваги розробок і досліджень, науковий світ відносився до них скептично. Дослідження в цьому напрямі були згорнуті аж до 1983 року, коли вони, нарешті, отримали фінансування від Агентства Перспективних Військових Досліджень США, DARPA. Цей факт став сигналом на початок нового нейромережевого буму.

Інтерес широкої наукової громадськості до нейромереж прокинувся після теоретичної роботи фізика Джона Хопфілда (1982 р.), що запропонував модель асоціативної пам'яті в нейронних ансамблях. Холфілд і його багаточисельні послідовники збагатили теорію нейромереж багатьма ідеями з арсеналу фізики, такими як колективні взаємодії нейронів, енергія мережі, температура навчання і так далі. Проте справжній бум практичного вживання нейромереж почався після публікації в 1986 році Девідом Румельхартом із співавторами методу навчання багатшарового персептрона, названого ними методом зворотного поширення помилки (error back-propagation). В цей же час з'явилися перші комерційні проекти нейрокомп'ютерів (Mark III фірми TRW, США). Обмеження персептронів, про які писали Мінський і Пейперт, виявилися переборними, а можливості обчислювальної техніки - достатніми для вирішення широкого круга прикладних завдань. У 90-х роках продуктивність послідовних комп'ютерів зростає настільки, що це дозволило моделювати з їх допомогою роботу паралельних нейронних мереж з числом

нейронів від декількох сотень до десятків тисяч. Такі емулятори нейромереж здатні вирішувати багато важливих з практичної точки зору завдань. У свою чергу, нейромережеві програмні комплекси стануть тим носієм, який виведе на технологічну орбіту сьогодення паралельне нейромережеве hardware.

Подальший розвиток нейромережевих технологій підтримувався створенням спеціалізованих центрів нейрокомп'ютерів, проведенням великого числа міжнародних конференцій і форумів. Обсяг продажу комерційних нейромережевих продуктів в період з 1991 по 2007 р. виріс більш ніж в чотирнадцять разів. Зараз, з врахуванням переходу на нанотехнології, появою нових знань про діяльність людського мозку, з'явилися принципово нові архітектури, технологічні рішення, визначені напрями досліджень, з'явився широкий спектр завдань, які не під силу вирішувати алгоритмічно навіть на сучасних ПК.

Означення. Штучні нейронні мережі - вид математичних моделей, які будуються за принципом організації і функціонування їх біологічних аналогів, - мереж нервових клітин (нейронів) мозку. У основі їх побудови лежить ідея про те, що нейрони можна моделювати досить простими автоматами (штучними нейронами), а вся складність мозку, гнучкість його функціонування і інші найважливіші якості визначаються зв'язками між нейронами.

Відомі такі типи нейронних мереж:

- персептрон Розенблатта та багат шаровий персептрон;
- нейронні мережі Джордана, Елмана, Хеммінга;
- мережа Хопфілда та мережа Кохонена;
- когнітрон та неокогнітрон;
- хаотична нейронна мережа та осциляторна нейронна мережа;
- мережі зустрічного розповсюдження та мережі радіальних базисних функцій (RBF – мережі);
- мережі узагальненої регресії та імовірнісні мережі;
- сіамська нейронна мережа та мережа адаптивного резонансу.

Як вже було відмічено, світовий ринок пропонує більше кілька тисяч нейромережевих пакетів. Загальний об'єм ринку нейронних мереж до 2008 року перевищив \$10 млрд. І, практично, кожний розробник традиційних аналітичних пакетів сьогодні прагне включити нейронні мережі в нові версії своїх програм. У США нейронні мережі застосовуються в аналітичних комплексах кожного крупного банку. Наприклад, продаж одного тільки нейромережевого пакету «Brain Maker Pro» порівнянн з об'ємами продажів найпопулярнішого пакету технічного аналізу MetaStock.

У зв'язку з великою різноманітністю штучних нейронних мереж и специфікою їх реалізації важливим є проведення класифікацію їх архітектур для подальшого застосування в інтелектуальних системах аналізу даних.

- Нейронні мережі можуть бути *синхронні* і *асинхронні*.

У синхронних нейронних мережах в кожен момент часу свій стан міняє лише один нейрон. У асинхронних - стан міняється відразу у цілої групи нейронів, як правило, у всього шару.

- По типу базової архітектури - *шаровані* і *повнозв'язні* мережі.

Ключовим в шаруватих мережах є поняття прошарка. Шар - один або декілька нейронів, на входи яких подається один і той же загальний сигнал. Шаровані нейронні мережі - нейронні мережі, в яких нейрони розбиті на окремі групи (шари) так, що обробка інформації здійснюється пошарово. У шарованих мережах нейрони i -го шару отримують вхідні сигнали, перетворюють їх і через точки розгалуження передають нейронам $(i+1)$ шару. І так до k -го шару, який видає вихідні сигнали для інтерпретатора і користувача. Число нейронів в кожному шарі не пов'язане з кількістю нейронів в інших шарах і може бути довільним.

В рамках одного шару дані обробляються паралельно, а в масштабах всієї мережі обробка ведеться послідовно - від шару до шару. До шарованих нейронних мереж відносяться, наприклад, багатошарові перцептрони, мережі радіальних базисних функцій, когнітрон, некогнітрон, мережі асоціативної пам'яті. Проте сигнал не завжди подається на всі нейрони шару. У когнітроні,

наприклад, кожен нейрон поточного шару отримує сигнали тільки від близьких йому нейронів попереднього шару.

Шаровані мережі, у свою чергу, можуть бути *одношаровими* і *багатошаровими*. Одношарова мережа – мережа складається з одного шару. Багатошарова мережа – мережа має декілька шарів. У багатошаровій мережі перший шар називається вхідним, подальші - внутрішніми або прихованими, останній шар - вихідним. Таким чином, проміжні шари - це всі шари в багатошаровій нейронній мережі, окрім вхідного і вихідного. Вхідний шар мережі реалізує зв'язок з вхідними даними, вихідний - з вихідними.

Таким чином, нейрони можуть бути вхідними, вихідними і прихованими. Вхідний шар організований з вхідних нейронів (input neuron), які отримують дані і поширюють їх на входи нейронів прихованого шару мережі. Прихований нейрон (hidden neuron) - це нейрон, що знаходиться в прихованому шарі нейронної мережі. Вихідні нейрони (output neuron), з яких організований вихідний шар мережі, видає результати роботи нейронної мережі.

У *повнозв'язних мережах* кожен нейрон передає свій вихідний сигнал решті нейронів, включаючи самого себе. Вихідними сигналами мережі можуть бути все або деякі вихідні сигнали нейронів після декількох тактів функціонування мережі. Всі вхідні сигнали подаються всім нейронам.

- По спрямованості зв'язків.

Нейронні мережі бувають із *зворотними зв'язками* і *без зворотних зв'язків*.

Серед мереж без зворотних зв'язків розрізняють *мережі із зворотним розповсюдженням помилки* та *інші мережі*. Мережі першої групи характеризуються фіксованою структурою, ітераційним навчанням, корегуванням вагів по помилках. Перевагами мереж без зворотних зв'язків є простота їх реалізації і гарантоване отримання відповіді після проходження даних по шарах. Недоліком цього виду мереж вважається мінімізація розмірів мережі - нейрони багато разів беруть участь в обробці даних. Менший об'єм мережі полегшує процес навчання.

До мереж із зворотними зв'язками відносять мережі Хопфілда та мережі Кохонена. Перевагами мереж із зворотними зв'язками є складність навчання, викликана великим числом нейронів для алгоритмів одного і того ж рівня складності. Недоліки цього виду мереж - потрібні спеціальні умови, що гарантують сходимость обчислень.

- Мережі прямого розповсюдження і рекурентні мережі.

До мереж *прямого розповсюдження* зазвичай відносять перцептрони, мережу Back Propagation, мережу зустрічного розповсюдження та карти Кохонена. Всі зв'язки направлені строго від вхідних нейронів до вихідних.

Характерна особливість *рекурентних мереж* - наявність блоків динамічної затримки і зворотних зв'язків, що дозволяє їм обробляти динамічні моделі, наприклад, мережа Хопфілда або мережа Елмана - мережа, що складається з двох шарів, в якій прихований шар охоплений динамічним зворотним зв'язком, що дозволяє врахувати передісторію спостережуваних процесів і накопичити інформацію для вироблення правильної стратегії управління. Ці мережі застосовуються в системах управління рухомими об'єктами. Окремим випадком рекурентних мереж є двонаправлені мережі. У таких мережах між шарами існують зв'язки як в напрямі від вхідного шару до вихідному, так і в зворотному. Класичним прикладом є нейронна мережа Косько.

- Навчання з вчителем або без нього.

Перед використанням нейронної мережі її необхідно навчити. Процес навчання нейронній мережі полягає в підстроюванні її внутрішніх параметрів під конкретне завдання. Алгоритм роботи нейронної мережі є ітеративним, його кроки називають епохами або циклами.

Епоха - одна ітерація в процесі навчання, що включає пред'явлення всіх прикладів з навчальної множини і, можливо, перевірку якості навчання на контрольній множині. Процес навчання здійснюється на навчальній вибірці. Вона включає вхідні значення і відповідні їм вихідні значення набору даних. В ході навчання нейронна мережа знаходить деякі залежності вихідних полів від

вхідних. Таким чином, перед нами ставиться питання - які вхідні поля (ознаки) нам необхідно використовувати. Спочатку вибір здійснюється евристично, далі кількість входів може бути змінена. Складність може викликати питання про кількість спостережень в наборі даних. І хоча існують деякі правила, що описують зв'язок між необхідною кількістю спостережень і розміром мережі, їх вірність не доведена. Кількість необхідних спостережень залежить від складності вирішуваного завдання. При збільшенні кількості ознак кількість спостережень зростає нелінійно, ця проблема носить назву "Проклін розмірності". При недостатній кількості даних рекомендується використовувати лінійну модель.

Аналітик повинен визначити кількість шарів в мережі і кількість нейронів в кожному шарі. Далі необхідно призначити такі значення вагів і зміщень, які зможуть мінімізувати помилку рішення. Ваги і зміщення автоматично підстраюються так, щоб мінімізувати різницю між бажаними і отриманими на виході сигналами, яка називається помилка навчання.

Помилка навчання для побудованої нейронної мережі обчислюється шляхом порівняння вихідних і цільових (бажаних) значень. З отриманих різниць формується функція помилок. Функція помилок - це цільова функція, що вимагає мінімізації в процесі керованого навчання нейронної мережі. За допомогою функції помилок можна оцінити якість роботи нейронної мережі під час навчання. Наприклад, часто використовується сума квадратів помилок. Від якості навчання нейронній мережі залежить її здатність вирішувати поставлені перед нею завдання.

Навчання з вчителем передбачає, що для кожного навчального вхідного прикладу потрібне знання правильної відповіді або функції оцінки якості відповіді. Таке навчання називають *керованим*. Нейронній мережі пред'являються значення вхідних і вихідних сигналів, а вона по певному алгоритму підстроює ваги синаптичеських зв'язків. В процесі навчання проводиться корегування вагів мережі за наслідками порівняння фактичних вихідних значень з вхідними, відомими наперед.

При *навчанні без вчителя* розкривається внутрішня структура даних або кореляції між зразками в наборі даних. Виходи нейронної мережі формуються самостійно, а ваги змінюються по алгоритму, що враховує тільки вхідні і похідні від них сигнали. Це навчання називають також *некерованим*. В результаті такого навчання об'єкти або приклади розподіляються по категоріях, самі категорії і їх кількість можуть бути наперед не відомі.

При навчанні нейронних мереж часто виникають серйозні труднощі, названі *проблемою перенавчання* (overfitting). Перенавчання, або надмірно близька підгонка - надмірно точна відповідність нейронної мережі конкретному набору навчальних прикладів, при якому мережа втрачає здібність до узагальнення.

Перенавчання виникає в разі дуже довгого навчання, недостатнього числа навчальних прикладів або переускладненої структури нейронної мережі. Перенавчання пов'язане з тим, що вибір навчальної (тренувальної) множини є випадковим. З перших кроків навчання відбувається зменшення помилки. На подальших кроках з метою зменшення помилки (цільовій функції) параметри підстроюються під особливості навчальної множини. Проте при цьому відбувається «підстроювання» не під загальні закономірності ряду, а під особливості його частини - навчальної підмножини. При цьому точність прогнозу зменшується. Один з варіантів боротьби з перенавчанням мережі - ділення навчальної вибірки на дві множини (навчальної і тестової). На навчальній множині відбувається навчання нейронній мережі. На тестовій множині здійснюється перевірка побудованої моделі. Ці множини не повинні перетинатися. З кожним кроком параметри моделі змінюються, проте постійне зменшення значення цільової функції відбувається саме на навчальній множині. При розбитті множини на дві ми можемо спостерігати зміну помилки прогнозу на тестовій множині паралельно із спостереженнями над навчальною множиною. Якась кількість кроків помилки прогнозу зменшується на обох множинах. Проте на певному кроці помилка на тестовій множині починає зростати, при цьому помилка на навчальній множині продовжує зменшуватися.

Цей момент вважається кінцем реального або справжнього навчання, з нього і починається перенавчання. Прогноз на тестовій множині є перевіркою працездатності побудованої моделі. Помилка на тестовій множині може бути помилкою прогнозу, якщо тестова множина максимально наближена до теперішнього моменту.

- Одношарові і багатошарові штучні нейронні мережі.

Хоча один нейрон і здатний виконувати прості процедури аналізу, але для серйозних нейронних обчислень необхідно сполучати нейрони в мережі. Проста мережа складається з групи нейронів, створюючих шар (рис. 3.11). Відзначимо, що вершини-круги зліва служать лише для розподілу вхідних сигналів. Вони не виконують яких-небудь обчислень і тому не вважатимуться шаром. Для більшої наочності позначимо їх кругами, щоб відрізнити їх від обчислюючих нейронів, позначених квадратами. Кожен елемент з множини входів X окремою вагою поєднан з кожним штучним нейроном. А кожен нейрон видає зважену суму входів в мережу. У штучних і біологічних мережах багато з'єднань можуть бути відсутніми, але тут вони показані всі для демонстрації загальної картини. Можуть існувати також з'єднання між виходами і входами елементів в шарі.

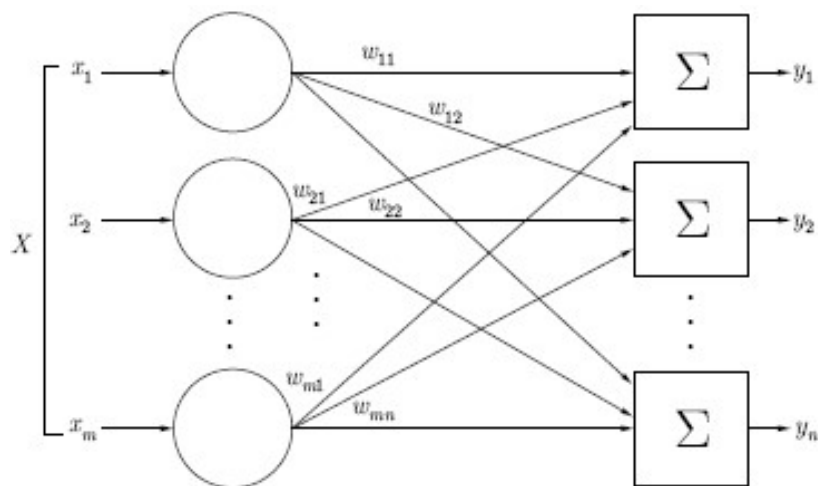


Рис. 3.11. Одношарова нейронна мережа.

Зручно вважати ваги елементами матриці W . Матриця має m рядків і n стовпців, де m - число входів, а n - число нейронів. Наприклад, $w_{2,3}$ - це вага,

що пов'язує другий вхід з третім нейроном. Таким чином, обчислення вихідного вектора N , компонентами якого є виходи OUT нейронів, зводиться до матричного множення $N = XW$, де N і X - вектори-рядки.

Більші і складніші багатошарові нейронні мережі володіють, як правило, і великими обчислювальними можливостями. Хоча створені мережі всіх конфігурацій, які тільки можна собі представити, пошарова організація нейронів копіює шаруваті структури певних відділів мозку. Виявилось, що такі багатошарові мережі володіють більшими можливостями ніж одношарові, і останніми роками були розроблені алгоритми для їх навчання. Багатошарові мережі можуть будуватися з каскадів шарів. Вихід одного шару є входом для подальшого шару (рис. 3.12).

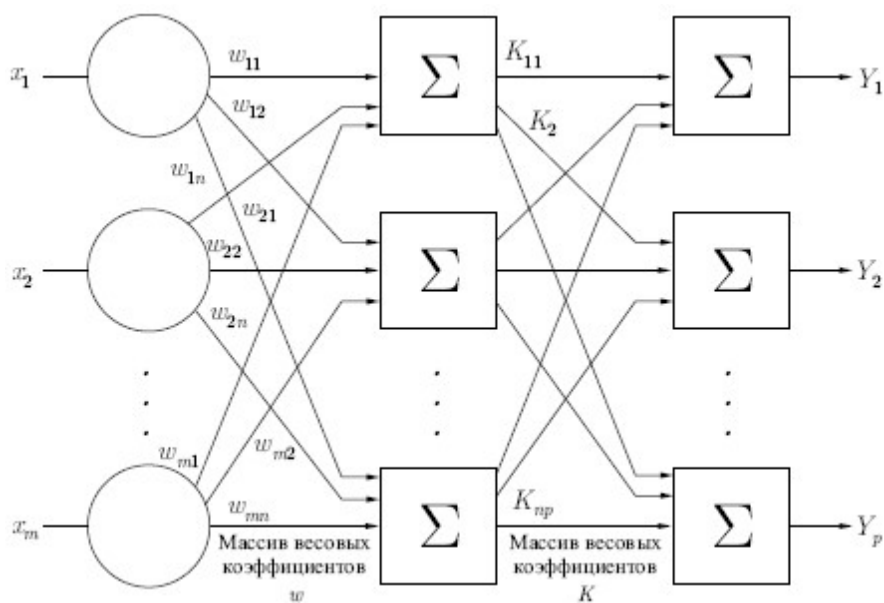


Рис. 3.12. Багатошарова нейронна мережа.

Багатошарові мережі не можуть привести до збільшення обчислювальної потужності в порівнянні з одношаровою мережею, якщо активаційна функція між шарами лінійна. Обчислення виходу шару полягає в множенні вхідного вектора на першу вагову матрицю з подальшим множенням (якщо відсутня нелінійна активаційна функція) результуючого вектора на другу вагову матрицю $OUT = (XW_1)W_2$. Оскільки множення матриць асоціативне, то

$(XW_1)W_2 = X(W_1W_2)$. Це показує, що двошарова лінійна мережа еквівалентна одному шару з ваговою матрицею, рівною множинню двох вагових матриць. Отже, будь-яка багатошарова лінійна мережа може бути замінена еквівалентною одношаровою мережею. Проте одношарові мережі вельми обмежені по своїх обчислювальних можливостях. Таким чином, для розширення можливостей мереж в порівнянні з одношаровою мережею необхідна нелінійна активаційна функція.

На жаль, немає загальноприйнятого способу підрахунку числа шарів в мережі. Багатошарова мережа складається, як показано на малюнку 3.12, з множини нейронів і вагів, що чередуються. Перший шар не приймається до уваги при підрахунку шарів, і мережа, подібна зображеною на малюнку 3.12, вважається двошаровою, оскільки тільки два шару виконують обчислення. Далі, ваги шару вважаються пов'язаними з наступними за ними нейронами. Отже, шар складається з множини вагів з наступними за ними нейронами, що підсумовують зважені сигнали.

Розглянемо найбільш поширені архітектури штучних нейронних мереж.

Перцептрон. Перцептрон (perceptron) - математична і комп'ютерна модель сприйняття інформації мозком (кібернетична модель мозку), запропонована Френком Розенблаттом і реалізована у вигляді електронної машини «Марк-1» в 1960 році. Перцептрон став однією з перших моделей нейромереж, а «Марк-1» - першим в світі нейрокомп'ютером. Не дивлячись на свою простоту, перцептрон здатний навчатися і вирішувати досить складні задачі (рис. 3.13).

Перцептрон складається з трьох типів елементів, а саме: сигнали, що поступають від сенсорів (S – елементи), передаються асоціативним елементам (A – елементи), а потім реагуючим елементам (R – елементи). Кожен сенсор може знаходитися в одному з двох станів - *спокою* або *збудження*, і лише в останньому випадку він передає одиничний сигнал в наступний шар, асоціативним елементам. А-елементи називаються асоціативними, тому що кожному такому елементу, як правило, відповідає цілий набір (асоціація) S-

елементів. А-елемент активізується, як тільки кількість сигналів від S-елементів на його вході перевищила деяку величину θ .

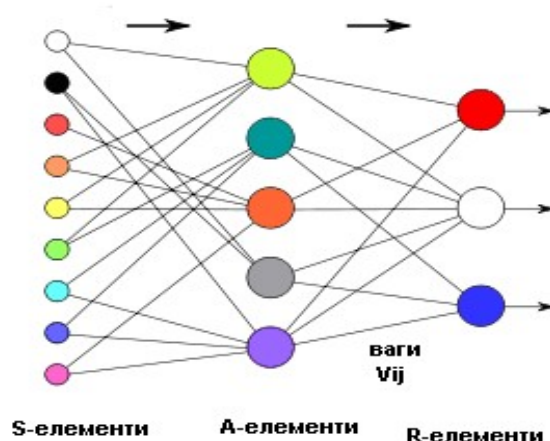


Рис. 3.13. Логічна схема персептрона.

Сигнали від А-елементів, що збудилися, у свою чергу, передаються в суматор R, причому сигнал від i -го асоціативного елемента передається з коефіцієнтом w_i . Цей коефіцієнт називається *вагою* А-R зв'язку. Так само як і А-елементи, R-елемент підраховує суму значень вхідних сигналів, помножених на ваги. R-елемент, а разом з ним і елементарний персептрон, видає «1», якщо лінійна форма перевищує поріг θ , інакше на виході буде «-1». Математично, функцію, що реалізується R-елементом, можна записати так:

$$f(x) = \text{sign}\left(\sum_{i=1}^n w_i x_i - \theta\right)$$

Таким чином, персептрони дозволяють створити набір «асоціацій» між вхідними стимулами і необхідною реакцією на виході. У біологічному плані це відповідає перетворенню, наприклад, зорової інформації у фізіологічну відповідь від рухових нейронів. Згідно сучасної термінології, персептрони можуть бути класифіковані як штучні нейронні мережі:

1. з одним прихованим шаром;
2. з пороговою передавальною функцією;
3. з прямим розповсюдженням сигналу.

Навчання елементарного персептрона полягає в зміні вагових коефіцієнтів w_i зв'язків А-R. Ваги зв'язків S-A (які можуть приймати значення

{-1; 0; +1}) і значення порогів А-елементів вибираються випадковим чином на самому початку і потім не змінюються.

Після навчання перцептрон готовий працювати в режимі *розпізнавання* або *узагальнення*. У цьому режимі перцептрону пред'являються не знайомі об'єкти, і перцептрон повинен встановити, до якого класу вони належать. Робота перцептрона полягає в наступному: при пред'явленні об'єкту А-елементи, що збудилися, передають сигнал R-елементу, рівний сумі відповідних коефіцієнтів w_i . Якщо ця сума позитивна, то ухвалюється рішення, що даний об'єкт належить до першого класу, а якщо вона негативна - то другому.

Важливою властивістю будь-якої нейронної мережі є здібність до навчання. Розенблатт намагався класифікувати різні алгоритми навчання перцептрона, називаючи їх системами підкріплення. При цьому слід розрізняти поняття *представляємості* і *навчання*. Поняття *представляємості* відноситься до здатності перцептрона (або іншої мережі) моделювати певну функцію. Навчання ж вимагає наявності систематичної процедури настройки вагів мережі для реалізації цієї функції.

Навчання з вчителем. Класичний метод навчання перцептрона - це *метод корекції помилки*. Він є таким видом навчання з вчителем, при якому вага зв'язку не змінюється до тих пір, поки поточна реакція перцептрона залишається правильною. При появі неправильної реакції вага змінюється на одиницю, а знак (+/-) визначається протилежним від знаку помилки.

Допустимо, ми хочемо навчити перцептрон розділяти два класи об'єктів так, щоб при пред'явленні об'єктів першого класу вихід перцептрона був позитивний (+1), а при пред'явленні об'єктів другого класу - негативним (-1). Для цього виконаємо наступний алгоритм.

1. Випадковим чином вибираємо пороги для А-елементів і встановлюємо зв'язки S-A (далі вони змінюватися не будуть).
2. Початкові коефіцієнти w_i вважаємо рівними нулю.
3. Пред'являємо навчальну вибірку з вказівкою класу, до якого вона належить.

- Показуємо перцептрону об'єкт першого класу. При цьому деякі А-елементи збудяться. Коефіцієнти w_i , відповідні цим збудженим елементам, збільшуємо на 1.

- Пред'являємо об'єкт другого класу і коефіцієнти w_i тих А-елементів, які збудяться при цьому показі, зменшуємо на 1.

4. Обидві частини кроку 3 здійснимо для всієї навчальної вибірки. В результаті навчання сформується значення вагів зв'язків w_i .

Теорема збіжності перцептрона показує, що елементарний перцептрон, який навчається по такому алгоритму, незалежно від початкового стану вагових коефіцієнтів і послідовності появи стимулів завжди приведе до досягнення рішення за кінцевий проміжок часу.

Навчання без вчителя. Окрім класичного методу навчання перцептрона Розенблатт також ввів поняття про навчання без вчителя, запропонувавши спосіб навчання, названий альфа-система підкріплення. Ця система підкріплення, при якій ваги всіх активних зв'язків c_{ij} , які ведуть до елемента u_j , змінюються на однакову величину r , а ваги неактивних зв'язків за цей час не змінюються. Потім, з розробкою поняття багатошарового перцептрона, альфа-система була модифікована, і її стали називати *дельта-правило*. Модифікація була проведена з метою зробити функцію навчання такою, що диференціюється, що у свою чергу, потрібно для застосування *методу градієнтного спуску*, завдяки якому можливе навчання більш одного шару.

Метод зворотного розповсюдження помилки. Для навчання багатошарових мереж був запропонований градієнтний алгоритм навчання з вчителем, який проводить сигнал помилки, обчислений виходами перцептрона, до його входів, шар за шаром. Зараз це найпопулярніший метод навчання багатошарових перцептронів. Його перевага в тому, що він може навчити всі шари нейронної мережі, і його легко прорахувати локально. Проте цей метод є дуже довготривалим, до того ж, для його застосування потрібно, щоб передавальна функція нейронів була такою, що диференціюється. При цьому в

персептронах довелося відмовитися від бінарного сигналу, і користуватися на вході безперервними значеннями.

Алгоритм навчання мережі на основі метода зворотного розповсюдження помилки вимагає виконання наступних операцій:

1. Вибрати чергову навчальну пару з навчальної множини та подати вхідний вектор на вхід мережі.
2. Обчислити вихід мережі.
3. Обчислити різницю між виходом мережі і необхідним виходом (цільовим вектором навчальної пари).
4. Підкоригувати ваги мережі так, щоб мінімізувати помилку.

Повторювати кроки з 1 по 4 для кожного вектора навчальної множини до тих пір, поки помилка на всій множині не досягне прийняттого рівня.

Розенблатт виділив два фундаментальні обмеження для тришарових персептронів: відсутність у них здібності до узагальнення своїх характеристик на нові стимули або нові ситуації, а також нездатність аналізувати складні ситуації в зовнішньому середовищі шляхом розчленовування їх на простіші.

В той же час персептрони демонструють великі можливості для свого застосування, зокрема, в задачах класифікації, кластеризації, прогнозування та апроксимації, стиснення даних та асоціативної пам'яті.

Задачі класифікації. Як образи можуть виступати різні за своєю природою об'єкти: економічні показники, зображення, зразки звуків і т.д. При навчанні мережі пропонуються різні зразки образів з вказівкою того, до якого класу вони відносяться. Зразок, як правило, представляється як вектор значень ознак. При цьому сукупність всіх ознак повинна однозначно визначати клас, до якого відноситься зразок. У випадку, якщо ознак недостатньо, мережа може співвіднести один і той же зразок з декількома класами, що невірно. Після закінчення навчання мережі їй можна пред'являти невідомі раніше образи і отримувати відповідь про приналежність до певного класу.

Топологія такої мережі характеризується тим, що кількість нейронів у вихідному шарі, як правило, рівна кількості визначуваних класів. При цьому

встановлюється відповідність між виходом нейронної мережі і класом, який він представляє. Коли мережі пред'являється якийсь образ, на одному з її виходів повинна з'явитися ознака того, що образ належить цьому класу, в той же час на інших виходах повинна бути ознака того, що образ даному класу не належить. Якщо на двох або більш виходах є ознака приналежності до класу, вважається що мережа «не упевнена» в своїй відповіді.

Задачі кластеризації. Під кластеризацією розуміється розбиття множини вхідних сигналів на класи, при тому, що ні кількість, ні ознаки класів наперед невідомі. Після навчання така мережа здатна визначати, до якого класу відноситься вхідний сигнал. Мережа також може сигналізувати про те, що вхідний сигнал не відноситься ні до одного з виділених класів - це є ознакою нових, відсутніх в навчальній вибірці, даних. Таким чином, подібна мережа може виявляти нові, невідомі раніше класи сигналів. Відповідність між класами, виділеними мережею, і класами, що існують в предметній області, встановлюється людиною.

Прогнозування і апроксимація. Здібності нейронної мережі до прогнозування безпосередньо виходять з її здібності до узагальнення і виділення прихованих залежностей між вхідними і вихідними даними. Після навчання мережа здатна передбачити майбутнє значення якоїсь послідовності на основі декількох попередніх значень або деяких існуючих зараз чинників. Слід зазначити, що прогнозування можливе тільки тоді, коли попередні зміни дійсно якоюсь мірою зумовлюють майбутні. Наприклад, прогнозування котирувань акцій на основі котирувань за минулий тиждень може виявитися успішним, тоді як прогнозування результатів завтрашньої лотереї на основі даних за останні 50 років майже напевно не дасть ніяких результатів.

Стиснення даних і асоціативна пам'ять. Здібність нейромереж до виявлення взаємозв'язків між різними параметрами дає можливість виразити дані великої розмірності компактніше, якщо дані тісно взаємозв'язані один з одним. Зворотний процес - відновлення початкового набору даних з частини інформації – називається асоціативною пам'яттю. Асоціативна пам'ять дозволяє

також відновлювати початковий образ із зашумлених вхідних даних. Рішення задачі гетероасоціативної пам'яті дозволяє реалізувати пам'ять, що адресується по змісту.

Багат шаровий перцептрон. Багат шаровий перцептрон Розенблатта - перцептрон з додатковими шарами A - елементів, розташованими між S і R елементами. Оскільки елементарний перцептрон вже володів двома шарами зв'язків і трьома шарами елементів (нейронів), то такий перцептрон не вважався багат шаровим, і багат шаровість зумовлювалась тільки за наявності мінімум чотирьох шарів елементів. Інша важлива відмінність полягає в тому, що не обов'язково всі зв'язки мають бути навчані, частина з них може бути випадково вибранна і фіксована.

Нейронна мережа Ворда. Штучна нейронна мережа, топологія якої характеризується тим, що внутрішні (приховані) шари нейронів розбиті на блоки (рис. 3.14):

1. Нейрони вхідного шару.
2. Нейрони блоку прихованого шару.
3. Нейрони вихідного шару.

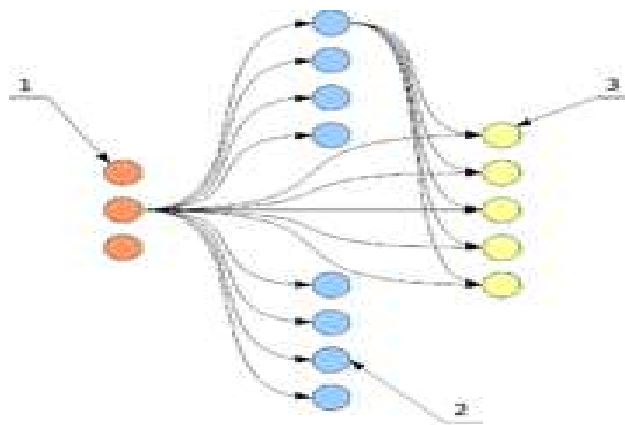


Рис. 3.14. Нейронна мережа Ворда с двома прихованими шарами.

Нейронні мережі Ворда розрізняються кількістю блоків прихованого шару і наявністю або відсутністю обхідних з'єднань. Розбиття прихованих шарів на блоки дозволяє використовувати різні передавальні функції для різних блоків

прихованого шару. Таким чином, одні і ті ж сигнали, отримані від вхідного шару, зважуються і обробляються паралельно з використанням декількох способів, а отриманий результат потім обробляється нейронами вихідного шару. Застосування різних методів обробки для одного і того ж набору даних дозволяє сказати, що нейронна мережа аналізує дані з різних аспектів. Практика показує, що мережа показує дуже добрі результати при рішенні задач прогнозування і розпізнавання образів. Для нейронів вхідного шару, як правило, встановлюється лінійна функція активації. Функція активації для нейронів з блоків прихованого і вихідного шару визначається експериментально.

Для навчання нейронної мережі Ворда можна застосовувати метод зворотного розповсюдження помилки.

Нейронна мережа Джордана. Мережа Джордана - це вид мереж, який виходить з багат шарового персептрона, якщо на його вхід подати крім вхідного вектора вихідний із затримкою на один або декілька тактів. В перших рекурентних мережах головною ідеєю було дати мережі бачити свій вихідний образ на попередньому кроці. У такої мережі тільки частина сенсорів приймає сигнали з навколишнього світу, на інші сенсори приходять вихідний образ з попереднього моменту часу. Розглянемо проходження послідовності сигналів через мережу. Сигнал поступає на групу сенсорів сполучених із зовнішнім світом (INPUT) і проходить в прихований шар (HIDDEN). Перетворений прихованим шаром сигнал піде на вихідний шар (OUTPUT) і вийде з мережі, а його копія потрапить на затримку. Далі в мережу, на сенсори, що сприймають зовнішні сигнали, поступає другий образ, а на контекстну групу сенсорів (CONTEXT) - вихідний образ з попереднього кроку із затримки. Далі зі всіх сенсорів сигнал піде в прихований шар, потім на вихідний (рис. 3.15).

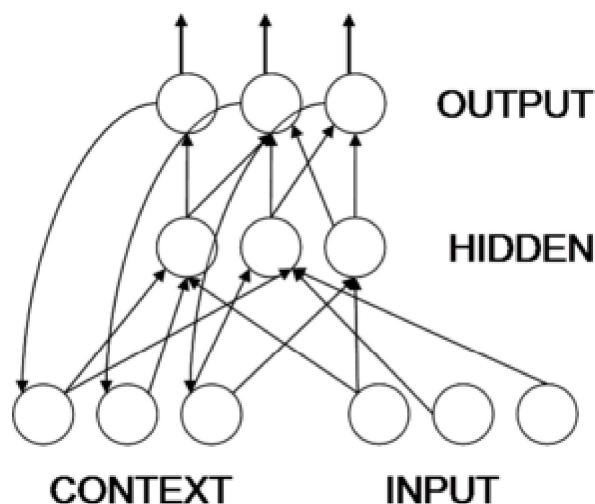


Рис.3.15. Архітектура рекурентної мережі Джордана.

Нейронна мережа Елмана. Нейронна мережа Елмана - один з видів рекурентної мережі, яка так само як і мережа Джордана виходить з багат шарового персептрона введенням зворотних зв'язків, тільки зв'язки йдуть не від виходу мережі, а від виходів внутрішніх нейронів. Це дозволяє врахувати передісторію спостережуваних процесів і накопичити інформацію для вироблення правильної стратегії управління.

Когнітрон та неокогнітрон. Когнітрон - штучна нейронна мережа на основі принципу самоорганізації. Своєю архітектурою когнітрон схожий на будову зорової кори, має ієрархічну багат шарову організацію, в якій нейрони між шарами зв'язані тільки локально. Навчається конкурентним навчанням (без вчителя). Кожен шар мозку реалізує різні рівні узагальнення; вхідний шар чутливий до простих образів, таких, як лінії, і їх орієнтації в певних областях візуальної області, тоді як реакція інших шарів є складнішою, абстрактнішою і незалежною від позиції образу.

Когнітрон конструюється у вигляді шарів нейронів, сполучених синапсами. Предсинаптічеській нейрон в одному шарі пов'язаний з постсинаптічним нейроном в наступному шарі. Є два типи нейронів: збудливі вузли, які прагнуть викликати збудження постсинаптічного вузла, і гальмуючі вузли, які гальмують це збудження. Збудження нейрона визначається зваженою сумою його збудливих і гальмуючих входів, проте насправді механізм є

складнішим, ніж просте підсумовування. Дана нейронна мережа одночасно є як моделлю процесів сприйняття на мікрорівні, так і обчислювальною системою, що застосовується для технічних завдань розпізнавання образів.

Неокогнітрон є подальшим розвитком ідеї когнітрона і точніше відображає будову зорової системи, дозволяє розпізнавати образи незалежно від їх перетворень, обертань, спотворень і змін масштабу. Неокогнітрон може як самонавчатися, так і навчатися з вчителем. Неокогнітрон отримує на вході двовимірні образи, аналогічні зображенням на сітчастій оболонці ока, і обробляє їх в подальших шарах аналогічно тому, як це було виявлено в зоровій корі людини. Звичайно, в неокогнітроне немає нічого, що обмежує його використання тільки для обробки візуальних даних, він достатньо універсальний і може знайти широке застосування як узагальнена система розпізнавання образів.

Хаотичні нейронні мережі. По структурі такі мережі є одношаровими рекурентними мережами, елементи яких зв'язані «кожен з кожним», без утворення зв'язку «сам на себе». Структуру мережі утворюють N нейронів, кожний з яких відповідає за конкретний об'єкт при цьому зв'язки симетричні. Навчання хаотичної мережі полягає у формуванні один раз вагових коефіцієнтів мережі, які визначають міру взаємного впливу пар нейронів один на одного. Основна відмінність від інших мереж полягає в утворенні групової поведінки нейронів і виділенні з первинного хаосу деякого унікального порядку, властивого тільки заданому вхідному набору об'єктів, представлених у вигляді окремих точок деякого умовного зображення. Прийнято вважати, що такі властивості нейронної мережі як великі обчислювальні можливості і стійкість до помилки в нейродинамічних системах обумовлюється колективною поведінкою всіх нейронів. Це стає можливим завдяки сильній взаємодії між нейронами, коли кожен пов'язаний з кожним. Має місце глобальний характер зв'язку: якщо не розрізняються миттєві стани елементів, то не відрізняється і діюче на них поле.

У хаотичній нейромережі на вхід в початковий момент часу подається відразу все зображення (представляються відразу всі дані, що підлягають кластеризації і лише один раз), на відміну від інших нейронних мереж, побудованих на основі шару змагання, де елементи представляються по черзі і неодноразово. Стан мережі задається станом нейронів. Щоб значення функції активації нейронів, коливалися в діапазоні $[-1, 1]$, використовується спеціальна логістична функція

$$f(x) = 1 - 2x^2$$

Мережа відпускається на деякий період часу у вільне функціонування для того, щоб можна було побудувати фазові портрети поведінки, що відображають поведінку кожного нейрона окремо. Подача на вхід мережі набору об'єктів може бути розглянута як вплив навколишнього середовища на спочатку неврегульовану поведінку нейронів і формування під її дією нейронних ансамблів. При цьому сумісне функціонування нейронів є проявом інтелекту на рівні малого колективу, вирішального завдання кластеризації.

Розглядаючи сумісну динаміку поведінки нейронів після проходження перехідного процесу, виділяються пари нейронів, які міняють свої стани синхронно. Формуються дані про сумісну інформацію, яку несе кожна синхронізована пара нейронів і на її основі робиться висновок про те, які нейрони слід вважати такими, що відносяться до одного кластера. Первинний хаос перехідного процесу, з часом зменшується, оскільки за рахунок виявлення взаємозв'язку значень виходів деяких пар нейронів починають змінюватися синфазно. Саме тому робота нейромережі розділяється на дві частини: перехідний період роботи і період спостереження, коли вважається, що хаос мінімальний, і можна починати витягувати інформацію.

Осциляторні нейронні мережі. Осциляторні нейронні мережі - нейронні мережі, основними структурними одиницями яких є осцилятори. Функціонують такі мережі за рахунок коливань окремих елементів або груп елементів і їх взаємодії. Осциляторні мережі представляють науковий інтерес, оскільки істотну роль в розумових процесах людини грають коливання.

3.3. Нейронні мережі Хопфілда та Кохонена.

Багато нейронних мереж не мають зворотних зв'язків, тобто зв'язків, що йдуть від виходів мереж до їх входів. Відсутність зворотного зв'язку гарантує безумовну стійкість мереж. Вони не можуть увійти до режиму, коли вихід безперервно блукає від стану до стану і не придатний до використання. Але це вельми бажана властивість досягається не безкоштовно, мережі без зворотних зв'язків володіють більш обмеженими можливостями в порівнянні з мережами із зворотними зв'язками.

Оскільки мережі із зворотними зв'язками мають шляхи, що передають сигнали від виходів до входів, то відгук таких мереж є динамічним, тобто після появи нового входу обчислюється вихід і, передаючись по мережі зворотного зв'язку, модифікує вхід. Потім вихід повторно обчислюється, і процес повторюється знову і знову. Для стійкої мережі послідовні ітерації приводять до все меншим змінам виходу, поки врешті-решт вихід не стає постійним. Для багатьох мереж процес ніколи не закінчується, такі мережі називають *нестійкими*. Нестійкі мережі володіють цікавими властивостями і вивчалися як приклад хаотичних систем.

Проблема стійкості ставила в безвихідь перших дослідників. Ніхто не був в змозі передбачити, які з мереж будуть стійкими, а які знаходитимуться в постійній зміні. Більш того, проблема представлялася настільки важкою, що багато дослідників було налагоджена песимістично відносно можливості застосування нейронних мереж. На щастя, була отримана теорема, що описала підмножину мереж із зворотними зв'язками, виходи яких врешті-решт досягають стійкого стану. Це чудове досягнення відкрило дорогу подальшим дослідженням і сьогодні багато вчених займаються дослідженням складної поведінки і можливостей цих систем. Дж. Хопфілд зробив важливий вклад як в теорію, так і у вживання систем із зворотними зв'язками. Тому деякі з конфігурацій відомі як мережі Хопфілда [22, 34, 42].

Розглянемо мережу із зворотними зв'язками, що складається з двох шарів (рис. 3.16). Нульовий шар не виконує обчислювальної функції, а лише розподіляє виходи мережі назад на входи. Кожен нейрон першого шару обчислює зважену суму своїх входів, даючи сигнал NET, який потім за допомогою нелінійної функції F перетвориться в сигнал OUT. Ці операції схожі з нейронами інших мереж.

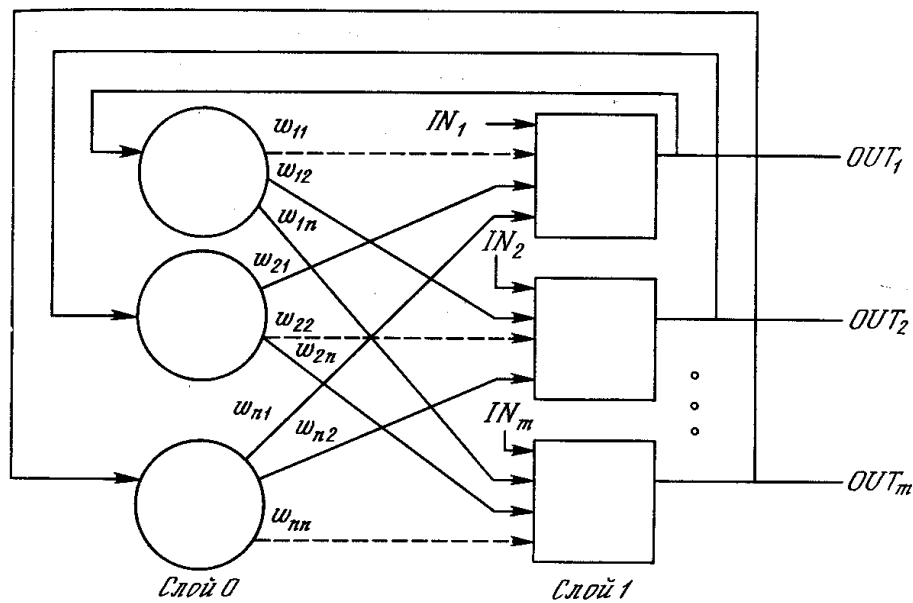


Рис. 3.16. Одношарова мережа з зворотними зв'язками.

В перших варіантах нейронних мереж Хопфілда функція F була просто пороговою функцією. Вихід такого нейрона дорівнює одиниці, якщо зважена сума виходів з інших нейронів більше порогу T_j , інакше вона дорівнює нулю. Він обчислюється таким чином:

$$NET_j = \sum_{i \neq j} w_{ij} OUT_i + IN_j,$$

$$OUT = 1, \text{ якщо } NET_j > T_j,$$

$$OUT = 0, \text{ якщо } NET_j < T_j$$

$$OUT \text{ не вимірюється, якщо } NET_j = T_j.$$

Стан мережі – це просто множина поточних значень сигналів OUT від всіх нейронів. Оскільки виходом бінарного нейрона може бути лише нуль або одиниця (проміжних рівнів немає), то поточний стан мережі є двоїчним числом, кожен біт якого є сигналом OUT деякого нейрона.

Функціонування мережі легко візуалізується геометрично. На рисунку 3.17 показаний випадок двох нейронів у вихідному шарі, причому кожній вершині квадрата відповідає один з чотирьох станів системи (00, 01, 10, 11). На рисунку 3.18 показана трьохнейронна система, представлена кубом, що має вісім вершин, кожна з яких помічена трьохбітовим бінарним числом. У загальному випадку система з n нейронами має 2^n різних станів і представляється n -мірним гіперкубом.

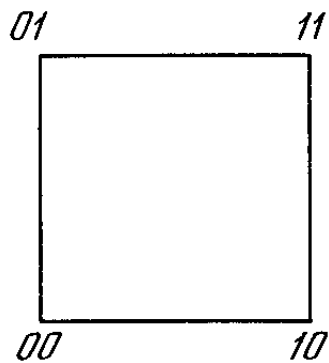


Рис. 3.17. Два нейрона породжують систему з чотирма станами.

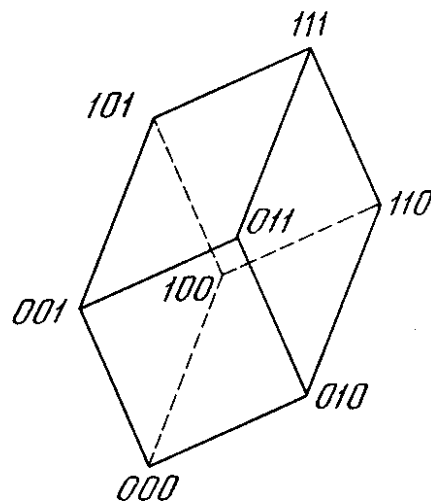


Рис. 3.18. Три нейрона породжують систему з восьма станами.

Коли подається новий вхідний вектор, мережа переходить з вершини у вершину, поки не стабілізується. Стійка вершина визначається мережевими вагами, поточними входами і величиною порогу. Якщо вхідний вектор частково неправильний або неповний, то мережа стабілізується у вершині, найближчій до бажаної.

Людська пам'ять асоціативна, тобто деякий спогад може породжувати велику пов'язану з ним область. Наприклад, декілька музичних тактів можуть викликати цілу гамму спогадів, включаючи пейзажі, звуки і запахи. Навпаки, звичайна комп'ютерна пам'ять є локально адресуємою, пред'являється адреса і витягується інформація за цією адресою. Мережа із зворотним зв'язком формує асоціативну пам'ять. Подібно до людської пам'яті по заданій частині потрібної інформації вся інформація витягується з «пам'яті». Аби організувати асоціативну пам'ять за допомогою мережі із зворотними зв'язками, ваги повинні вибиратися так, щоб утворювати енергетичні мінімуми в потрібних вершинах одиничного гіперкуба. Хопфілд розробив асоціативну пам'ять з безперервними виходами, що змінюються в межах від +1 до -1, відповідних двоїчним значенням 0 і 1. Інформація, що запам'ятовується, кодується двоїчними векторами і зберігається у вагах згідно наступній формулі:

$$w_{ij} = \sum_{d=1..m} (OUT_{i,d} OUT_{j,d}),$$

де m – число вихідних векторів, що запам'ятовуються; d – номер вихідного вектора, що запам'ятовується; $OUT_{i,j}$ – i -компонента вихідного вектора, що запам'ятовується.

Цей вираз може стати яснішим, якщо відмітити, що ваговий масив W може бути знайдений обчисленням зовнішнього множиння кожного вектора, що запам'ятовується, з самим собою і підсумовуванням матриць, отриманих таким чином. Це може бути записано у вигляді

$$W = \sum_i D_i^T D_i,$$

де D_i – i -й вектор-рядок, що запам'ятовується.

Як тільки ваги задані, мережа може бути використана для здобуття вихідного вектора, що запам'ятовується, по даному вхідному вектору, який може бути частково неправильним або неповним. Для цього виходам мережі спочатку надають значення цього вхідного вектора. Потім вхідний вектор забирається і мережі надається можливість «розслабитися», опустившись в найближчий глибокий мінімум. Мережа йде по локальному нахилу функції енергії і може бути захоплена локальним мінімумом, не досягнувши найкращого в глобальному сенсі рішення.

Недоліком мереж Хопфілда є їх тенденція стабілізуватися в локальному, а не глобальному мінімумі функції енергії. Ця трудність долається в основному за допомогою класу мереж, відомих під назвою машин Больцмана, в яких зміни станів нейронів обумовлені статистичними, а не детермінованими закономірностями. Існує тісна аналогія між цими методами і відпалом металу, тому і самі методи часто називають імітацією відпалу.

Розглянемо застосування нейронної мережі Хопфілда на прикладі розв'язання відомої задачі комівояжера.

Задача комівояжера є оптимізаційною задачею, що часто виникає на практиці. Вона може бути сформульована таким чином: для деякої групи міст із заданими відстанями між ними потрібно знайти найкоротший маршрут з відвідинами кожного міста один раз і з поверненням у вихідну точку. Було доведено, що ця задача належить великій множині задач, названих «NP-повними». Для NP-повних задач не відомо кращого методу розв'язання, ніж повний перебір усіх можливих варіантів. Оскільки такий повний пошук практично нездійснений для великого числа міст, то евристичні методи використовуються для знаходження прийнятних, хоча і неоптимальних рішень.

Запропоноване рішення, засноване на мережах із зворотними зв'язками, є типовим в цьому відношенні. Відповідь знаходиться так швидко, що в певних випадках метод може виявитися корисним. Припустимо, що міста, які необхідно відвідати, помічені буквами A, B, C і D, а відстані між парами міст є d_{ab} , d_{bc} і т. д. Рішенням є впорядкована множина з n міст. Завдання полягає у

відображенні його в обчислювальну мережу з використанням нейронів. Кожне місто представлене рядком з n нейронів. Вихід одного і лише одного нейрона з них дорівнює одиниці (всі інші дорівнюють нулю). Цей вихід нейрона показує порядковий номер, в якому дане місто відвідується при обході. На рисунку 3.19 показаний випадок, коли місто С відвідується першим, місто А – другим, місто D – третім і місто В – четвертим.

		Порядок дотримування			
Місто		1	2	3	4
А		0	1	0	0
В		0	0	0	1
С		1	0	0	0
D		0	0	1	0

Рис. 3.19. Варіант розв'язання задачі комівояжера

Для такого варіанта потрібно n^2 нейронів – число, яке швидко зростає із збільшенням числа міст. Довжина такого маршруту була б рівна $d_{ca} + d_{ad} + d_{db} + d_{bc}$. Оскільки кожне місто відвідується лише один раз і в кожен момент відвідується лише одне місто, то в кожному стовпці є по одній одиниці. Для задачі з n містами всього є $n!$ різних маршрутів обходу. Якщо $n = 60$, то існує $6934155 \cdot 10^{78}$ можливих маршрутів. Якщо прийняти до уваги, що в нашій галактиці є лише 10^{11} зірок, то стане зрозумілим, що повний перебір всіх можливих маршрутів для 1000 міст навіть на найшвидшому в світі комп'ютері займе час, порівнянний з геологічною епохою.

Продемонструємо тепер, як сконструювати мережу для вирішення цієї NP-повної проблеми. Кожен нейрон забезпечений двома індексами, які відповідають місту і порядковому номеру його відвідин в маршруті. Наприклад, $OUT_{x_j} = 1$ показує, що місто x було j -им по порядку містом маршруту.

Функція енергії повинна задовольняти двом вимогам: по-перше, повинна бути малою тільки для тих рішень, які мають по одній одиниці в кожному рядку і в кожному стовпці; по-друге, повинна надавати перевагу рішенням з короткою довжиною маршруту. Перша вимога задовольняється введенням наступної, складеної з трьох сум, функції енергії:

$$E = \frac{A}{2} \sum_x \sum_i \sum_{j \neq i} \text{OUT}_{xi} \text{OUT}_{xj} + \frac{B}{2} \sum_i \sum_x \sum_{y \neq x} \text{OUT}_{xi} \text{OUT}_{yi} + \frac{C}{2} \left[\left(\sum_x \sum_i \text{OUT}_{xi} \right) - n \right]^2,$$

де А, В і С - деякі константи. Цим досягається виконання наступних умов:

1. Перша потрібна сума дорівнює нулю в тому і лише в тому випадку, якщо кожен рядок (місто) містить не більш за одну одиницю.
2. Друга потрібна сума дорівнює нулю в тому і лише в тому випадку, якщо кожен стовпець (порядковий номер відвідин) містить не більш за одну одиницю.
3. Третя сума дорівнює нулю в тому і лише в тому випадку, якщо матриця містить рівно n одиниць.

Друга вимога - перевага коротким маршрутам - задовольняється за допомогою додавання наступного члена до функції енергії:

$$E = \frac{D}{2} \sum_x \sum_{y \neq x} \sum_i d_{xy} \text{OUT}_{xi} (\text{OUT}_{y,i+1} + \text{OUT}_{y,i-1}).$$

Відмітимо, що цей член є довжиною будь-якого допустимого маршруту. Для зручності індекси визначаються по модулю n , тобто $\text{OUT}_{n+j} = \text{OUT}_j$, а D - деяка константа. При достатньо великих значеннях А, В і С низькоенергетичні стани представлятимуть допустимі маршрути, а великі значення D гарантують, що буде знайдений короткий маршрут.

Тепер задамо значення вагів, тобто встановимо відповідність між членами у функції енергії і членами загальної форми. Отримуємо

$$\begin{aligned} w_{xi,yi} &= -A\delta_{xy}(1 - \delta_{ij}) && \text{(не допускає більш за одну одиницю в рядку)} \\ & -B\delta_{ij}(1 - \delta_{xy}) && \text{(не допускає більш за одну одиницю в стовпці)} \\ & -C && \text{(глобальне обмеження)} \end{aligned}$$

$-Dd_{xy}(\delta_{j,i+1} + \delta_{j,i-1})$ (член, що відповідає за довжину циклу),

де $\delta_{ij} = 1$, якщо $i = j$, інакше $\delta_{ij} = 0$.

Результати експерименту, в якому задача комівояжера була вирішена для 10 міст продемонстрували, що 16 з 20 прогонів зійшлися до допустимого маршруту і близько 50% рішень виявилися найкоротшими маршрутами, як це було встановлено за допомогою повного перебору. Цей результат стане більш вражаючим, якщо усвідомити, що є 181440 допустимих маршрутів.

Розглянемо практичне застосування нейромереж на прикладі розв'язання задачі «Видача кредиту клієнтові» в аналітичному пакеті Deductor. В якості навчального набору даних виступає база даних, що містить інформацію про клієнтів, зокрема: Сума кредиту, Термін кредиту, Мета кредитування, Вік, Рід, Освіта, Приватна власність, Квартира, Площа квартири. На основі цих даних необхідно побудувати модель, яка зможе дати відповідь, чи входить клієнт, охочий отримати кредит, в групу ризику неповернення кредиту, тобто користувач повинен отримати відповідь на питання чи «Видавати кредит?». Завдання відноситься до групи завдань класифікації, тобто навчання з вчителем.

Дані для аналізу знаходяться у задалегідь підготовленому файлі. Імпортуємо дані з файлу за допомогою майстра імпорту. Запускаємо майстер обробки і вибираємо метод обробки даних - нейронна мережа. Задаємо призначення початкових стовпців даних. Вихідний стовпець в нашому завданні – «Давати кредит», всі інші – вхідні (рис. 3.20).

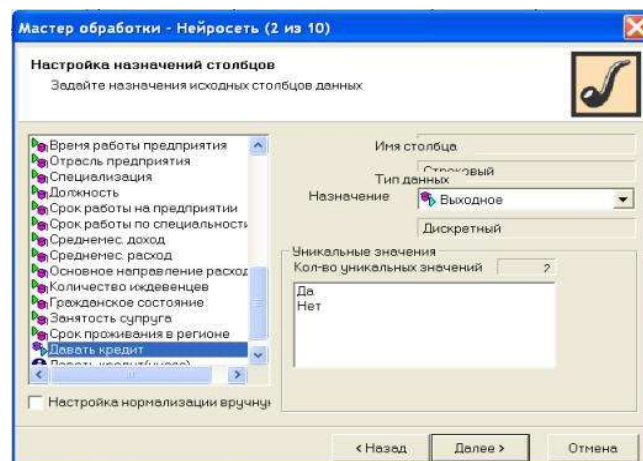


Рис. 3.20. Настройка назначений столбцов.

На наступному кроці розбиваємо початкову множину даних на навчальну і тестову та визначаємо структуру нейронної мережі, тобто вказуємо кількість нейронів у вхідному шарі - 33 (кількість вхідних змінних), в прихованому шарі - 1, у вихідному шарі - 1 (кількість вихідних змінних). Активаційна функція - Сигмоїда, і її крутизна рівна одиниці (рис. 3.21).

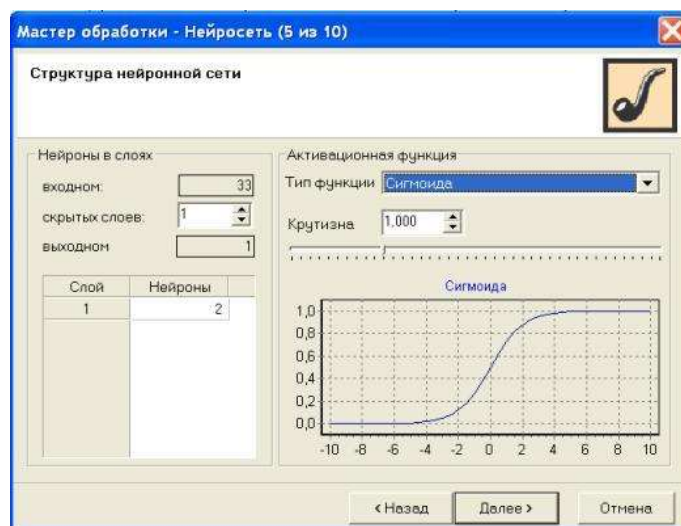


Рис. 3.21. Структура нейронної мережі.

Далі вибираємо алгоритм і параметри навчання нейронної мережі і налаштуємо умови зупинки навчання. Вважатимемо приклад розпізнаним, якщо помилка менше 0.005, і вкажемо умову зупинки навчання досягши епохи 10000. Запускаємо процес навчання і спостерігаємо за зміною величини помилки і відсотком розпізнаних прикладів в навчальній і тестовій множинах. У нашому випадку ми бачимо, що на епосі № 4536 в навчальній множині розпізнано 83,10% прикладів, а на тестовому - 85,71% прикладів (рис. 3.22).

Після закінчення процесу навчання для інтерпретації отриманих результатів ми маємо можливість вибрати візуалізатори із списку запропонованих, зокрема, таблицю зв'язаності, граф нейросеті, аналіз "що, якщо", і за допомогою їх проаналізувати отримані дані.

На рисунку 3.23 показана таблиця зв'язаності. По її діагоналі розташовані приклади, які були правильно розпізнані, тобто 55 клієнтів, яким можна видавати кредит, і 89 клієнтів, яким видавати кредит не варто. У решті

осередків розташовані ті клієнти, які були віднесені до іншого класу (1 і 4). Можна вважати, що правильно класифіковані практично всі приклади - 96,64%.

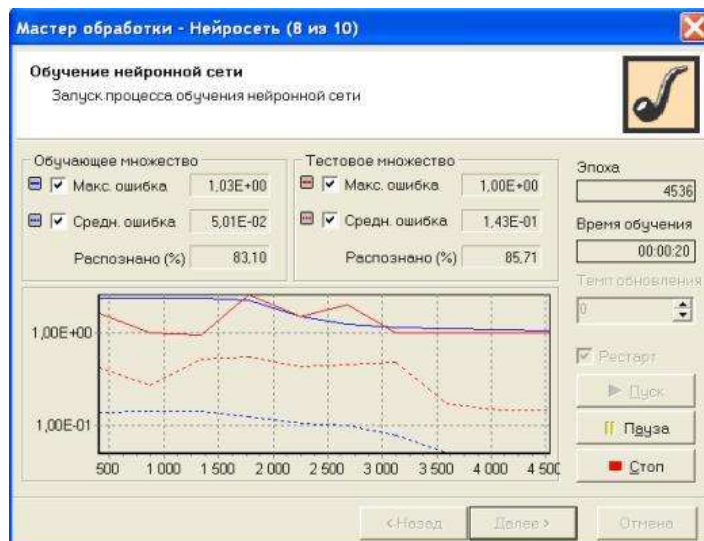


Рис. 3.22. Навчання нейтронної мережі.

Кросс таблица			
Фактически	Классифицировано		
	Да	Нет	Итого
Да	55	4	59
Нет	1	89	90
Итого	56	93	149

Рис. 3.23. Данні таблиці зв'язаності.

Тепер нейронна мережа підготовлена для практичного використання.

Важливим напрямком в розвитку нейронних мереж є самоорганізуючі карти (Self-Organizing Maps, SOM) або мережі (карти) Кохонена [21, 41, 53].

Мережі, названі картами Кохонена, - це один з різновидів нейронних мереж, проте вони принципово відрізняються від розглянутих вище, оскільки використовують неконтрольоване навчання. Нагадаємо, що при такому навчанні навчальна множина складається лише із значень вхідних змінних, в процесі навчання немає порівняння виходів нейронів з еталонними значеннями. Можна сказати, що така мережа вчиться розуміти структуру даних.

Ідея мережі Кохонена належить фінському ученому Тойво Кохонену (1982 рік). Основний принцип роботи мереж - введення в правило навчання

нейрона інформації щодо його розташування. У основі ідеї мережі Кохонена лежить аналогія з властивостями людського мозку. Кора головного мозку людини є плоским листом і згорнута складками. Таким чином, можна сказати, що вона володіє певними топологічними властивостями (ділянки, відповідальні за близькі частини тіла, примикають один до одного і все зображення людського тіла відображається на цю двовимірну поверхню).

Кarti, що самоорганізуються, можуть використовуватися для вирішення таких задач, як моделювання, прогнозування, пошук закономірностей у великих масивах даних, виявлення наборів незалежних ознак і стиснення інформації. Найбільш поширене застосування мереж Кохонена - рішення задачі класифікації без вчителя, тобто кластеризації. Нагадаємо, що при такій постановці задачі нам дан набір об'єктів, кожному з яких зіставлений рядок таблиці (вектор значень ознак). Потрібно розбити початкову множину на класи, тобто для кожного об'єкту знайти клас, до якого він належить. В результаті отримання нової інформації про класи можлива корекція існуючих правил класифікації об'єктів.

Ось два з найбільш поширених застосувань карт Кохонена: розвідувальний аналіз даних і виявлення нових явищ.

Розвідувальний аналіз даних. Мережа Кохонена здатна розпізнавати кластери даних, а також встановлювати близькість класів. Таким чином, користувач може поліпшити своє розуміння структури даних, щоб потім уточнити нейромережеву модель. Якщо в даних розпізнані класи, то їх можна позначити, після чого мережа зможе вирішувати задачі класифікації. Мережі Кохонена можна використовувати і в тих завданнях класифікації, де класи вже задані, - тоді перевага буде в тому, що мережа зможе виявити схожість між різними класами.

Виявлення нових явищ. Мережа Кохонена розпізнає кластери в навчальних даних і відносить всі дані до тих або інших кластерів. Якщо після цього мережа зустрінеться з набором даних, несхожим ні на один з відомих

зразків, то вона не зможе класифікувати такий набір і тим самим виявить його новизну.

Мережа Кохонена, на відміну від багат шарової нейронної мережі, дуже проста, вона представляє собою два шари: вхідний і вихідний. Її також називають самоорганізуючою картою. Елементи карти розташовуються в деякому просторі, як правило, двовимірному (рис. 3.24).

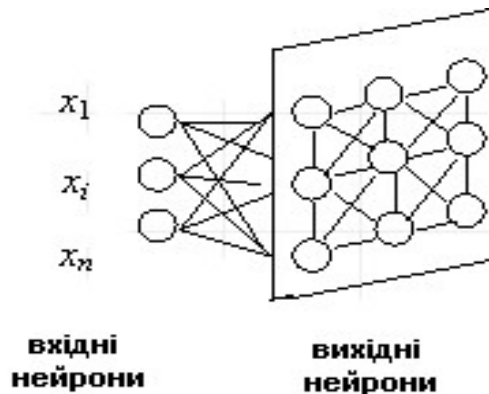


Рис. 3.24. Мережа Кохонена.

Мережа Кохонена навчається методом послідовних наближень. В процесі навчання таких мереж на входи подаються дані, але мережа при цьому підстроюється не під еталонне значення виходу, а під закономірності у вхідних даних. Починається навчання з вибраного випадковим чином вихідного розташування центрів. В процесі послідовної подачі на вхід мережі навчальних прикладів визначається найбільш схожий нейрон (той, у якого скалярне множення вагів і поданого на вхід вектора мінімально). Цей нейрон оголошується переможцем і є центром при підстроюванні вагів у сусідніх нейронів. Таке правило навчання припускає «змагальне» навчання з урахуванням відстані нейронів від «нейрона-переможця». Навчання при цьому полягає не в мінімізації помилки, а в підстроюванні вагів (внутрішніх параметрів нейронної мережі) для найбільшого збігу з вхідними даними.

Основний ітераційний алгоритм Кохонена послідовно проходить ряд епох, на кожній з яких обробляється один приклад з навчальної вибірки. Вхідні сигнали послідовно пред'являються мережі, при цьому бажані вихідні сигнали

не визначаються. Після пред'явлення достатнього числа вхідних векторів синаптичеські ваги мережі стають здатні визначити кластери. Ваги організовуються так, що топологічно близькі вузли чутливі до схожих вхідних сигналів. В результаті роботи алгоритму центр кластера встановлюється в певній позиції, задовільним чином кластеризующей приклади, для яких даний нейрон є «переможцем». В результаті навчання мережі необхідно визначити міру сусідства нейронів, тобто околицю нейрона-переможця. Околицею є декілька нейронів, які оточують нейрон-переможець. Спочатку до околиці належить велике число нейронів, далі її розмір поступово зменшується. Мережа формує топологічну структуру, в якій схожі приклади утворюють групи прикладів, що близько знаходяться на топологічній карті. Отриману карту можна використовувати як засіб візуалізації при аналізі даних. В результаті навчання карта Кохонена класифікує вхідні приклади на кластери (групи схожих прикладів) і візуально відображає багатовимірні вхідні дані на площині нейронів. Унікальність методу карт, що самоорганізуються, полягає в перетворенні n -мірного простору в двомірне. Застосування двомірних сіток пов'язане з тим, що існує проблема відображення просторових структур більшої розмірності. Маючи таке представлення даних, можна візуально визначити наявність або відсутність взаємозв'язку у вхідних даних. Нейрони карти Кохонена розташовують у вигляді двомірної матриці, розфарбовують цю матрицю залежно від аналізованих параметрів нейронів (рис. 3.25).

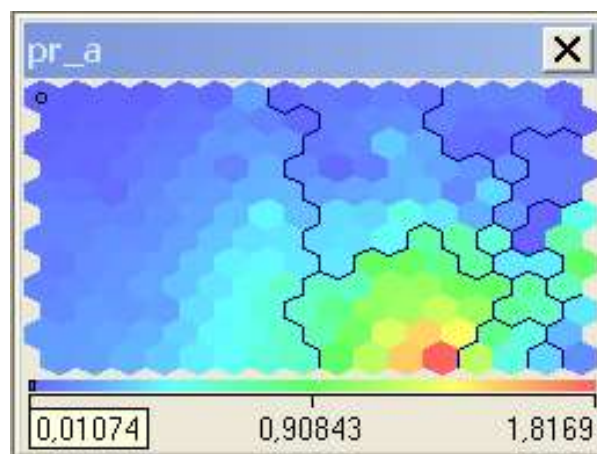


Рис. 3.25. Приклад карти Кохонена.

Що ж означає її розфарбовування? На рисунку 3.26 приведено розфарбовування карти, а точніше, її деякої ознаки, в тривимірному уявленні. Як ми бачимо, темно-сині ділянки на карті відповідають найменшим значенням показника, червоні - найвищим.

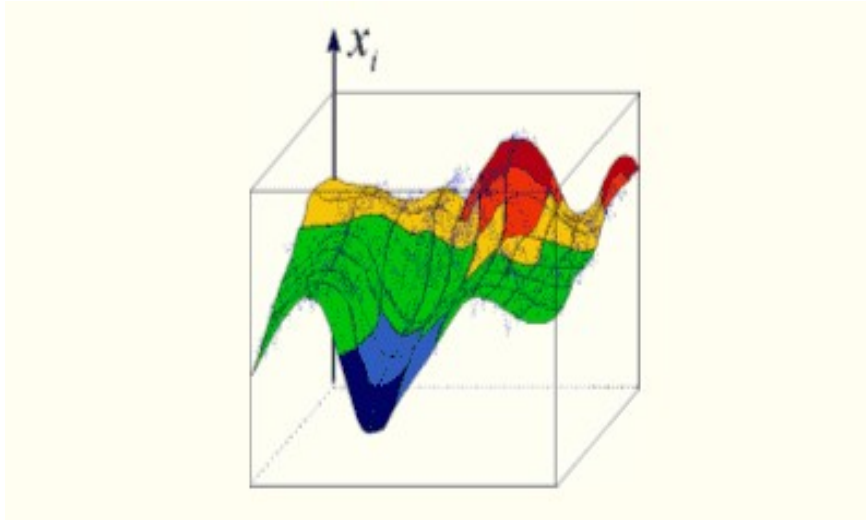


Рис. 3.26. Розфарбування ознаки в тривимірному просторі.

Тепер ми можемо сказати, які об'єкти мають найбільші значення даного показника (група об'єктів, позначена червоним кольором), а які - найменші значення (група об'єктів, позначена синім кольором). Таким чином, карти Кохонена (як і географічні карти) можна відображати:

- у двовірному вигляді, тоді карта розфарбовується відповідно до рівня виходу нейрона;
- у тривимірному вигляді.

В результаті роботи алгоритму отримуємо такі карти:

- карта входів нейронів;
- карта виходів нейронів;
- спеціальні карти.

Координати кожної карти визначають положення одного нейрона. Так, координати [15:30] визначають нейрон, який знаходиться на перетині 15-го стовпця з 30-м поряд в матриці нейронів. Розглянемо, що ж є ці карти.

Карта входів нейронів. Ваги нейронів підстроюються під значення вхідних змінних і відображають їх внутрішню структуру. Для кожного входу

малюється своя карта, розфарбована відповідно до значення конкретної ваги нейрона. При аналізі даних використовують декілька карт входів. На одній з карт виділяють область певного кольору - це означає, що відповідні вхідні приклади мають приблизно однакове значення відповідного входу. Колірний розподіл нейронів з цієї області аналізується на інших картах для визначення схожих або відмітних характеристик.

Карта виходів нейронів. На карту виходів нейронів проектується взаємне розташування досліджуваних вхідних даних. Нейрони з однаковими значеннями виходів утворюють кластери - замкнуті області на карті, які включають нейрони з однаковими значеннями виходів.

Спеціальні карти. Це карта кластерів, матриця відстаней, матриця щільності попадання і інші карти, які характеризують кластери, отримані в результаті навчання мережі Кохонена.

Важливо розуміти, що між всіма розглянутими картами існує взаємозв'язок - всі вони є різними розфарбовуваннями одних і тих же нейронів. Кожен приклад з навчальної вибірки має одне і те ж розташування на всіх картах.

Програмне забезпечення, що дозволяє працювати з картами Кохонена, зараз представлене безліччю інструментів. Це можуть бути як інструменти, що включають тільки реалізацію методу карт, що самоорганізуються, так і нейропакети з цілим набором структур нейронних мереж. Також даний метод реалізований в деяких універсальних інструментах аналізу даних.

До інструментарію, що включає реалізацію методу карт Кохонена, відносяться SoMine, Statistica, NeuroShell, NeuroScalp, Deductor і багато інших. Для вирішення задач використовуватимемо аналітичний пакет Deductor. Хай є база даних комерційних банків з показниками діяльності за поточний період. Необхідно провести їх кластеризацію, тобто виділити однорідні групи банків на основі показників з бази даних, всього показників - 21. Початкова таблиця знаходиться у файлі «banks.xls». Вона містить показники діяльності

комерційних банків за звітний період. Спочатку імпортуємо дані з xls-файлу в середу аналітичного пакету.

На першому кроці запускаємо майстер обробки і вибираємо із списку метод обробки «Карта Кохонена». Далі слід настроїти призначення стовпців, тобто для кожного стовпця вибрати одне з призначень: вхідне, вихідне, не використовується і інформаційне. Вкажемо всім стовпцям, відповідно до показників діяльності банків, призначення «Вхідні». «Вихідний» не призначаємо. Наступний крок пропонує розбити початкову множину на навчальну, тестову і валідаційну. За умовчанням, програма пропонує розбити множину на навчальне - 95% і тестове - 5%. Ці кроки аналогічні крокам в майстрові обробки для нейронних мереж. Далі задаємо параметри карти: кількість ячеек по X і по Y, їх форму (шестикутну або чотирикутну), встановлюємо параметри зупинки навчання і встановлюємо епоху, по досягненню якої навчання буде припинено.

На сьомому кроці настроюються інші параметри навчання: спосіб початкової ініціалізації, тип функції сусідства. Можливі два варіанти кластеризації: автоматичне визначення числа кластерів з відповідним рівнем значущості і фіксована кількість кластерів (визначається користувачем). Оскільки нам невідома кількість кластерів, виберемо автоматичне визначення їх кількості (рис. 3.27).

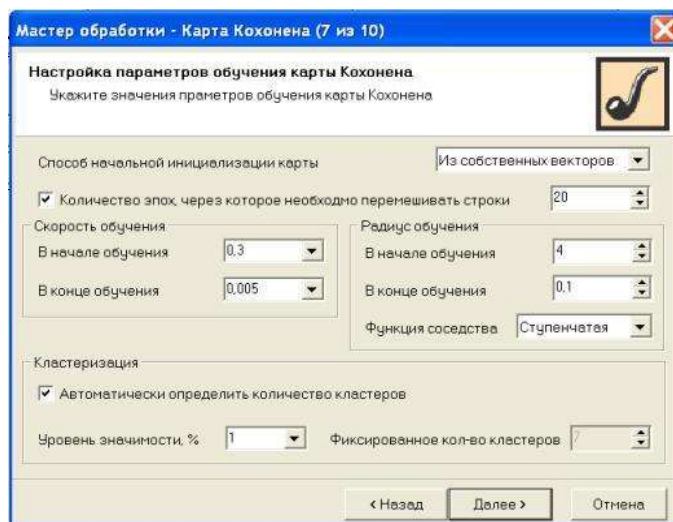


Рис. 3.27. Настройка параметров навчання.

На восьмому кроці запускаємо процес навчання мережі. Під час навчання можемо спостерігати зміну кількості розпізнаних прикладів і поточні значення помилок. Цей процес аналогічний тому, що ми розглядали при навчанні нейронних мереж. Після закінчення навчання в списку візуалізаторов виберемо «Карту Кохонена» і візуалізатор «Що-коли». На останньому кроці налаштуємо відображення карти Кохонена.

Вкажемо відображення всіх вхідних, вихідних стовпців, кластерів. При аналізі карт входів рекомендують використовувати відразу декілька карт. Досліджуємо фрагмент карти, що складається з карт трьох входів, який приведений на рис. 3.28.

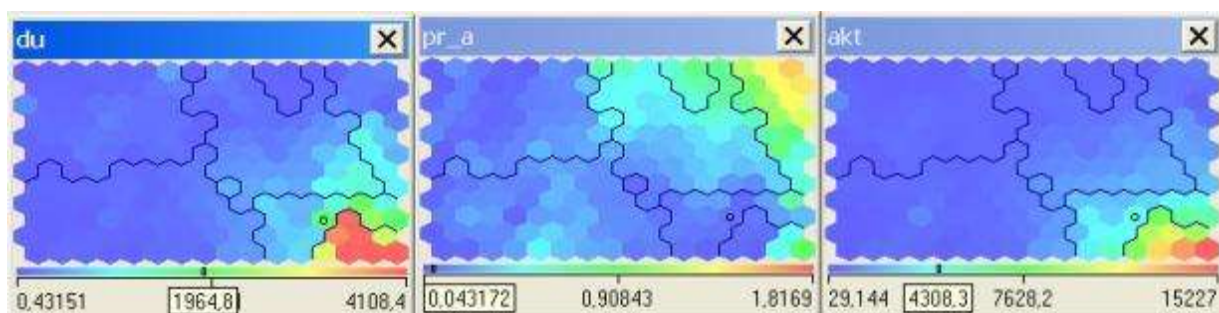


Рис. 3.28. Карты трёх входів.

На одній з карт виділяємо область з найбільшими значеннями показника. Далі має сенс вивчити ці ж нейрони на інших картах. На першій карті найбільші значення мають об'єкти, розташовані в правому нижньому кутку. Розглядаючи одночасно три карти, ми можемо сказати, що ці ж об'єкти мають найбільші значення показника, зображеного на третій карті. Також по розфарбовуванню першої і третьої карти можна зробити висновок, що існує взаємозв'язок між цими показниками. Також ми можемо визначити, наприклад, таку характеристику: кластер, розташований в правому верхньому кутку, характеризується низькими значеннями показників «депозити юридичних осіб» і «активи банку» і високими значеннями показників «прибутковість активів». Ця інформація дозволяє так охарактеризувати кластер, що знаходиться в правому верхньому кутку: це банки з невеликими активами, невеликими повернутими депозитними засобами від юридичних осіб, але з найбільш

прибутковими активами, тобто це група невеликих, але найбільш прибуткових банків. Це лише фрагмент висновку, який можна зробити, досліджуючи карту.

На рисунку 3.29 приведена ілюстрація карт входів і виходів, остання - це карта кластерів. Ми бачимо декілька карт входів (показників діяльності банків) і сформовані кластери, кожен з яких виділений окремим кольором.

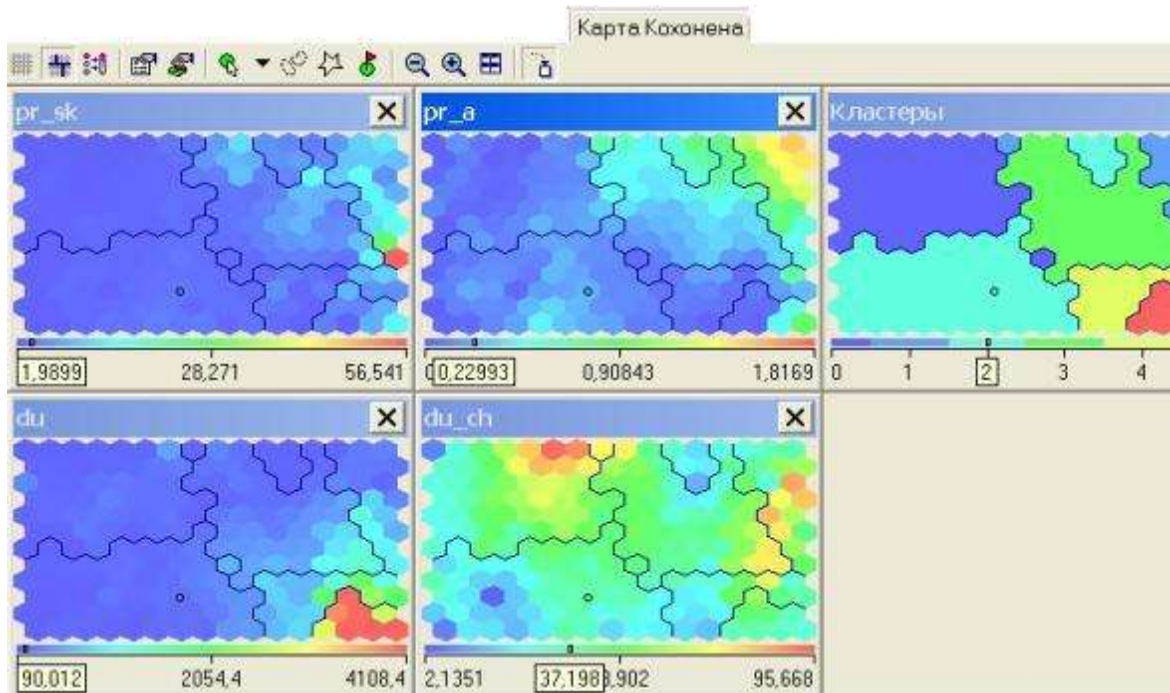


Рис. 3.29. Карти входів та виходів.

Для знаходження конкретного об'єкту на карті необхідно натискувати правою кнопкою миші на досліджуваному об'єкті і вибрати пункт «Знайти ячейку на карті» (рис. 3.30). В результаті ми можемо бачити як сам об'єкт, так і значення того виміру, який ми переглядаємо. Таким чином, ми можемо оцінити положення аналізованого об'єкту, а також порівняти його з іншими об'єктами.

В результаті застосування карт, що самоорганізуються, багатовимірний простір вхідних чинників був представлений в двомірному вигляді, в якому його досить зручний аналізувати. Зокрема, банки були класифіковані на 7 груп, для кожної з яких можливе визначення конкретних характеристик, виходячи з розфарбовування відповідних показників.

Таким чином, мережі Кохонена дозволяють спростити багатовимірну структуру, їх можна розглядати як одним з методів проектування багатовимірного простору в простір з нижчою розмірністю. Інтенсивність кольору в певній точці карти визначається даними, які туди попали: ячейка з мінімальними значеннями зображаються темно-синім кольором, ячейка з максимальними значеннями - червоним.

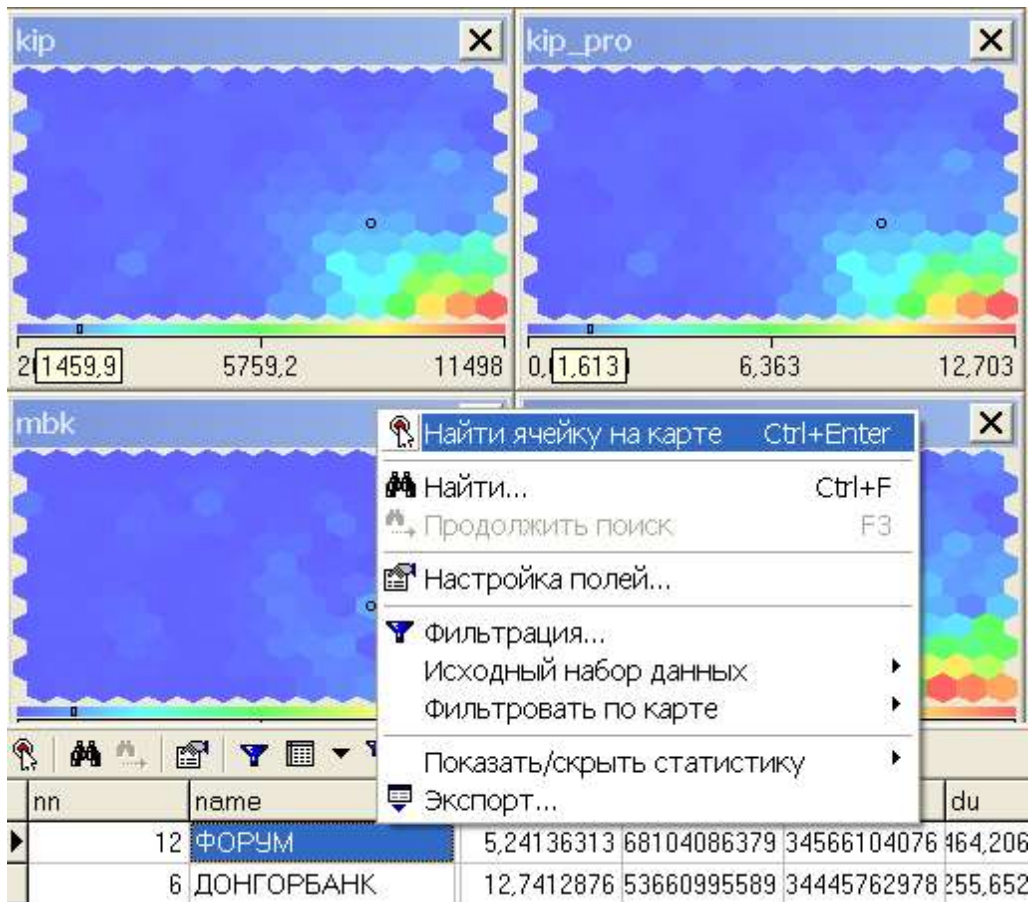


Рис. 3.30. Ячейка на карті Кохонена.

Інша принципова відмінність карт Кохонена від інших моделей нейронних мереж - інший підхід до навчання, а саме - некероване або неконтрольоване навчання. Цей тип навчання дозволяє даним навчальної вибірки містити значення лише вхідних змінних. Мережа Кохонена вчиться розуміти саму структуру даних і вирішує завдання кластеризації.

3.4. Програмні засоби реалізації нейромережевих технологій.

Нейронні мережі активно застосовуються в задачах аналізу вже більше 10 років. На фоні постійно змінної динаміки ринку, інтерес до нейромережевих технологій не лише не слабшає, а зростає з кожним роком. Зараз вже накопичений багатий досвід успішного використання нейронних мереж в практичних застосуваннях. По кількості реальних впроваджень лідирують системи інтелектуального аналізу даних в бізнесі і в управлінні процесами.

Більшість програмних комплексів, заснованих на нейромережевих технологіях, включають наступну послідовність дій:

- створення мережі (вибір користувачем параметрів або схвалення встановлених за умовчанням);
- навчання мережі;
- видача користувачеві рішення.

Існує величезна різноманітність нейропакетів, можливість використання нейромереж включена також практично у всі відомі статистичні пакети. Критерії порівняння нейропакетів: простота вживання, наочність інформації, що представляється, можливість використовувати різні структури, швидкість роботи, наявність документації. Вибір визначається кваліфікацією і вимогами користувача [2, 17, 34, 41, 53,70].

Для застосування методів нейронних мереж в процесі інтелектуального аналізу даних в бізнес-додатках розроблений ряд інструментальних засобів високого рівня. До них відносяться в першу чергу системи MathLab, STATISTICA Neural Networks, NeuroSolutions, BrainMakerPro, NeuroShell 2, 4Thought, SENN Sales та інші.

З метою більшої наочності розглянемо особливості застосування різних програмних комплексів підтримки нейромережевих технологій на конкретних практичних прикладах.

Пакет MathLab. Пакет MathLab надає користувачам можливість роботи з нейронними мережами. Стандартне постачання MathLab «Neural

Network Toolbox» надає широкі можливості для роботи з нейронними мережами всіх типів. Перевага цього пакету полягає в тому, що при його використанні користувач не обмежений моделями нейронних мереж і їх параметрами, жорстко закладеними в нейросимуляторі, а має можливість самостійно сконструювати ту мережу, яку рахує оптимальною для вирішення поставленого завдання. Пакет містить функції командного рядка і графічний інтерактивний майстер для швидкого покрокового створення нейромереж. Окрім цього Neural Network Toolbox забезпечує підтримку Simulink, що дозволяє моделювати нейромережі і створювати блоки на основі розроблених нейромережевих структур. Ключовими можливостями Neural Network Toolbox є:

- графічний інтерфейс користувача і майстер покрокового створення нейронних мереж;
- підтримка найбільш поширених мережевих парадигм;
- повний набір засобів для тренування нейромереж з вчителем і без;
- динамічно навчаємі нейромережі, включаючи нейромережі з запізненням, нелінійні і авторегресійні (NARX);
- підтримка Simulink для моделювання нейромережі, створення блоків на основі розроблених нейромережевих структур для адаптивних систем управління;
- модульне представлення мережі, що дозволяє створювати необмежене число шарів і міжмережевих зв'язків;
- візуалізація топології нейронної мережі.

Розглянемо приклад конструювання нейронної мережі в пакеті Matlab. Хай є 15 незалежних змінних - показників діяльності фірми і одна залежна змінна - об'єм продажів. Маємо базу даних за минулий рік. Необхідно побудувати потижневий прогноз об'ємів продажів на місяць. Для вирішення завдання пропонується використовувати тришарову мережу зворотного розповсюдження.

Сформуємо таку мережу, яка включає 15 нейронів у вхідному шарі (по кількості вхідних змінних), 8 нейронів в другому шарі і 1 нейрон у вихідному шарі (по кількості вихідних змінних). Для кожного шару виберемо передавальну функцію: перший шар - `logsig`, другий - `logsig`, третій - `purelin`. У середовищі MatLab синтаксис такої нейронної мережі виглядає таким чином:

$$\text{Net}=\text{netff}(\text{PR}, [\text{S1}, \text{S2}, \dots, \text{Sn}], \{\text{TF1}, \text{TF2} \dots, \text{TFn}\}, \text{btf}, \text{blf}, \text{pf}),$$

де `PR` - масив мінімальних і максимальних значень для `R` векторів входу; `Si` - кількість нейронів в *i*-му шарі; `TFi` - функція активації шару *i*; `btf` - навчальна функція, що реалізує метод зворотного розповсюдження; `blf` - функція настройки, що реалізує метод зворотного розповсюдження; `pf` - критерій якості навчання.

Активаційною функцією може виступати будь-яка функція, що диференціюється, наприклад, `tansig`, `logsig`, `purelin`.

$$\text{Net}=\text{netff}(\text{minmax}(\text{P}), [\text{n}, \text{m}, \text{l}], \{\text{logsig}, \text{logsig}, \text{purelin}\}, \text{trainpr}),$$

де `P` - множина вхідних векторів; `n` - кількість входів нейромережі; `m` - кількість нейронів в прихованому шарі; `l` - кількість виходів нейромережі.

Необхідно також встановити метод розрахунку значення помилки. Наприклад, якщо вибраний метод найменших квадратів, то ця функція виглядатиме так: `Net.performFcn='SSE'`. Для встановлення максимальної кількості епох рівної 10000 скористаємося функцією: `net.trainParam.epochs=10000`. Запустити процес навчання можна таким чином:

$$[\text{net}, \text{tr}]=\text{train}(\text{net}, \text{P}, \text{T}).$$

Після закінчення навчання мережі її можна зберегти у файлі, наприклад, з ім'ям `rvz.mat`. Для цього необхідно виконати команду: `save rvz net`.

Таким чином, в пакеті можливе конструювання мережі будь-якої складності і немає необхідності прив'язуватися до обмежень, що накладаються нейросимуляторами. Проте для роботи з нейронними мережами в пакеті MatLab необхідно вивчити як саме середовище, так і більшість функцій Neural Network Toolbox.

Розглянемо інший більш складніший приклад застосування нейронних мереж в пакеті MathLab.

Прогноз фінансових результатів - необхідний елемент будь-якої інвестиційної діяльності. Сама ідея інвестицій - вкладення грошей зараз з метою отримання доходу в майбутньому - ґрунтується на ідеї прогнозування майбутнього. Відповідно, прогноз фінансових результатів лежить в основі діяльності всієї індустрії інвестицій - всіх бірж і небіржових систем торгівлі цінними паперами. У останнє десятиліття спостерігалось стійке зростання популярності технічного аналізу - набору емпіричних правил, заснованих на різного роду індикаторах поведінки ринку. Технічний аналіз зосереджується на індивідуальній поведінці даного фінансового інструменту, поза його зв'язком з рештою цінних паперів. Але технічний аналіз дуже суб'єктивний і погано працює на правому краю графіка - саме там, де потрібно прогнозувати напрям ціни. Тому все більшої популярності набуває нейромережевий аналіз, оскільки на відміну від технічного, не припускає ніяких обмежень на характер вхідної інформації. Це можуть бути як індикатори даного тимчасового ряду, так і відомості про поведінку інших ринкових інструментів. Недаремно нейромережі активно використовують саме інституційні інвестори (наприклад, крупні пенсійні фонди), що працюють з великими портфелями, для яких особливо важливі кореляції між різними ринками.

Як відомо, для хорошого прогнозу потрібно користуватися по-перше, дуже якісно підготовленими даними, а по-друге, нейропакетамі з великою функціональністю. Дані можна отримати стандартними засобами MetaTrader (рис. 3.31). Для отримання прогнозу на основі нейронної мережі застосуємо пакет MathLab. З командного рядка по команді *anfisedit* запускаємо пакет ANFIS. Редактор складається з чотирьох панелей - для даних (Load data), для генерації мережі (Generate FIS), для тренування (Train FIS) і для її тестування (Test FIS). Верхня панель призначена для проглядання структури отриманої нейромережі (ANFIS Info).

Для початку роботи завантажуюмо дані, підготовлені на попередніх етапах. Для цього натискаємо кнопку Load Data і вказуємо файл з даними вибірки. Після цього створюємо неймережу натисненням кнопки Generate EIS (рис. 3.32).

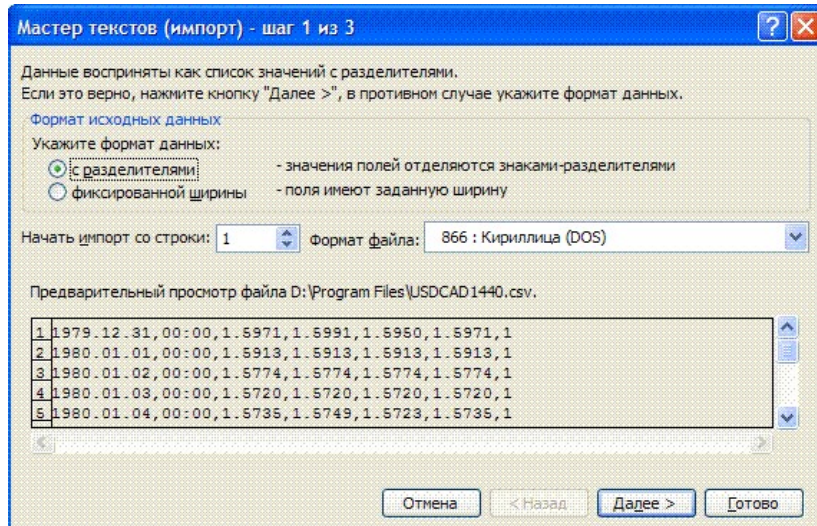


Рис. 3.31. Підготовка даних засобами пакету MetaTrader.

Для кожної з вхідних змінних задамо по 3 лінгвістичних змінних з трикутною функцією приналежності. В якості функції приналежності вихідної функції задамо лінійну функцію.

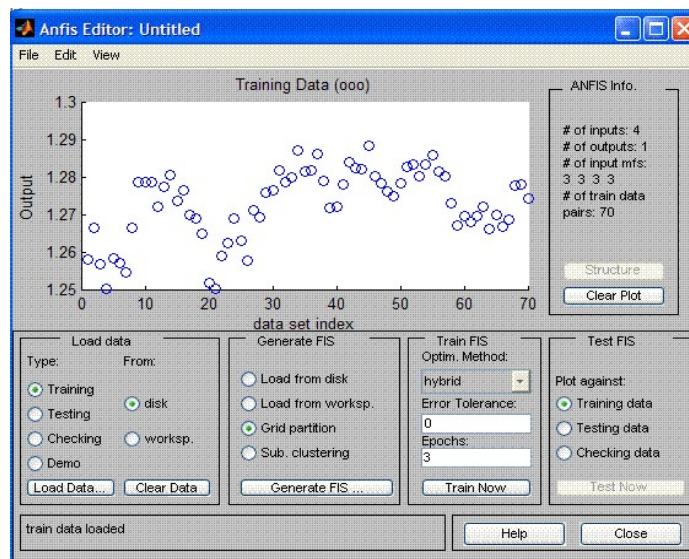


Рис. 3.32. Застосування пакету MathLab.

Для навчання нейромереж в пакеті передбачено 2 алгоритми навчання - зворотного розповсюдження і гібридний. При гібридному способі навчання мережа навчається буквально за пару-трійку проходів. На тренувальній вибірці (60 значень) після навчання прогноз мережі відрізняється від реального на декілька пунктів (рис. 3.33).

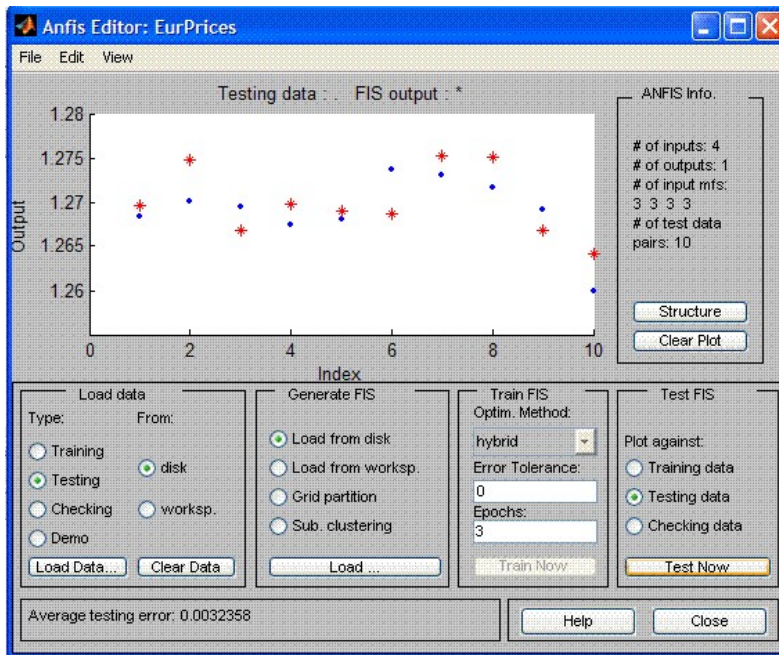


Рис.3.33. Прогнозування фінансових результатів.

Результатом роботи є багатoshарова гібридна нейронна мережа, яка здатна прогнозувати абсолютні значення цін на невелике майбутнє. Отриману нейромережу можна побачити при натисненні кнопки Structure (рис. 3.34).

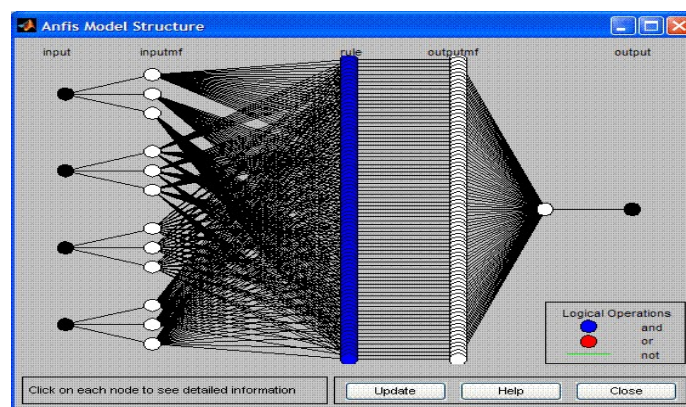


Рис. 3.34. Архітектура нейронної мережі.

Пакет STATISTICA Neural Networks. Ця система є могутнім і надзвичайно швидким середовищем аналізу нейросетевих моделей, що надає наступні можливості:

пре- і пост-процесування, включаючи вибір даних, кодування номінальних значень, шкалювання, нормалізація, видалення пропущених даних з інтерпретацією для класифікації, регресії і завдання тимчасових рядів;

могутні методи розвідувальних і аналітичних технологій, зокрема *Аналіз головних компонент* і *Пониження розмірності* для вибору потрібних вхідних змінних в розвідувальному (нейромережевому) аналізі даних;

найсучасніші, оптимізовані і могутні алгоритми навчання мережі, повний контроль над всіма параметрами, що впливають на якість мережі, такими як функції активації і помилок, складність мережі;

підтримка комбінацій нейромереж практично необмеженого розміру, створених в *Наборах мереж - Network Sets*, вибіркоче навчання нейромережевих сегментів; об'єднання, і збереження наборів мереж в окремих файлах;

повна інтеграція з системою STATISTICA, всі результати, графіки, звіти можуть бути надалі модифіковані за допомогою могутніх графічних і аналітичних інструментів STATISTICA (рис. 3.35).

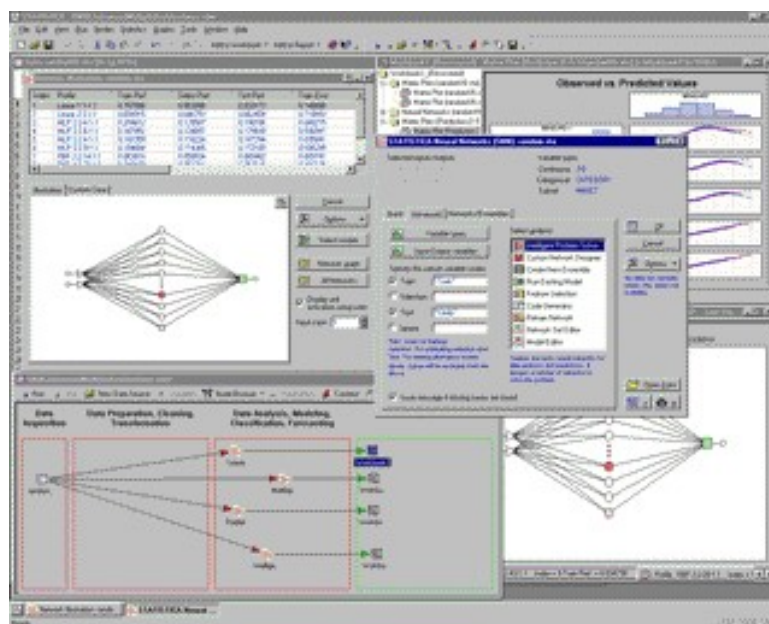


Рис. 3.35. Вигляд меню пакету STATISTICA Neural Networks.

В пакеті STATISTICA Neural Networks реалізовані зворотний і прямий алгоритми покрокового вибору. Крім того, нейро-генетичний алгоритм відбору вхідних даних поєднує в собі можливості генетичних і PNN/GRNN алгоритмів (PNN - імовірнісні нейронні мережі, GRNN - узагальнено-регресійні нейронні мережі) для автоматичного пошуку оптимальних комбінацій вхідних змінних, у тому числі і в тих випадках, коли між ними є кореляції і нелінійні залежності.

Перед тим, як дані будуть введені в мережу, вони повинні бути певним чином підготовлені. Так же важливо, щоб вихідні дані можна було правильно інтерпретувати. У пакеті є можливість автоматичного масштабування вхідних і вихідних даних, а також можуть бути автоматично перекодовані змінні з номінальними значеннями (наприклад, Стать = { Чоловіча, Жіноча}). STATISTICA Neural Networks містить також засоби роботи з пропущеними даними. Реалізовані такі функції нормування, як «єдинична сума», «переможець отримує все» і «вектор єдиничної довжини». Є засоби підготовки і інтерпретації даних, спеціально призначені для аналізу часових рядів. У задачах класифікації є можливість встановити довірчі інтервали, які система використовує потім для віднесення спостережень до того або іншого класу. У поєднанні із спеціальною реалізованою в STATISTICA Neural Networks функцією активації Софтмакс і кросс-ентропійними функціями помилок це дає принциповий теоретико-імовірнісний підхід до задач класифікації (рис. 3.36).

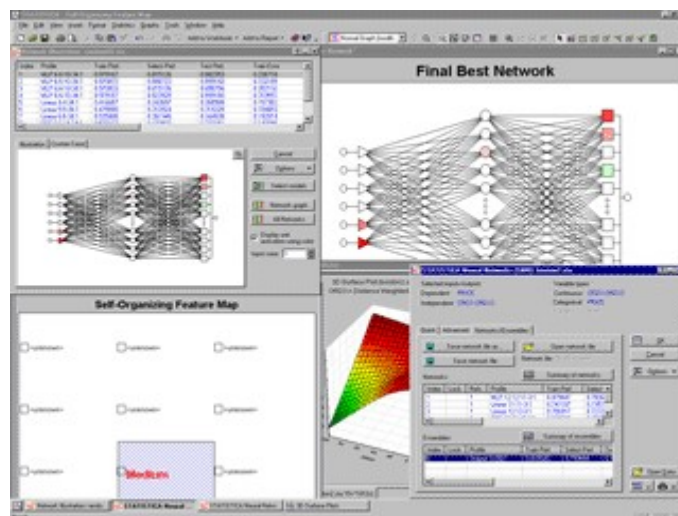


Рис.3.36. STATISTICA Neural Networks в задачах класифікації.

Різноманіття моделей нейронних мереж і безліч параметрів, які необхідно встановити (розміри мережі, параметри алгоритму навчання і т.д.), може поставити іншого користувача в безвихідь. Саме для цього в пакеті існує *Майстер рішень*, який може автоматично провести пошук відповідної архітектури мережі будь-якої складності. У системі STATISTICA Neural Networks реалізовані всі основні типи нейронних мереж, що використовуються при рішенні практичних задач, зокрема: багат шарові перцептрони, мережі на радіальних базисних функціях, карти Кохонена, імовірнісні нейронні мережі; мережі для кластеризації, лінійні мережі і т. д. (рис. 3.37).

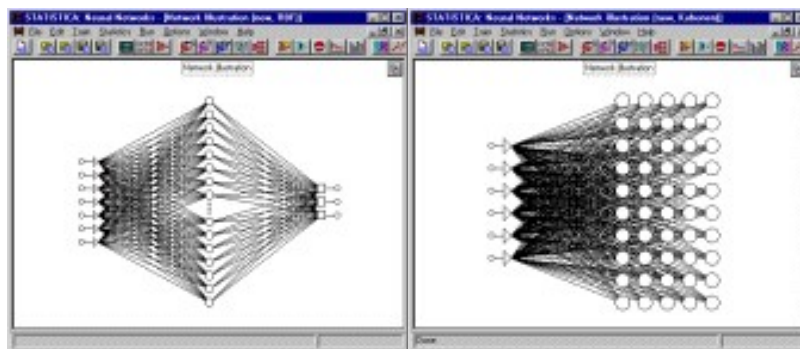


Рис. 3.37. Типи нейромереж в пакеті STATISTICA Neural Networks.

Також, в системі реалізовані мережеві ансамблі, що формуються з випадкових (але значущих) комбінацій вищеперелічених мереж. Існує ще один зручний засіб: ви можете зв'язати мережі, щоб вони запускалися послідовно. Це корисно при препроцесуванні для знаходження рішень з мінімальною вартістю (рис. 3.38).

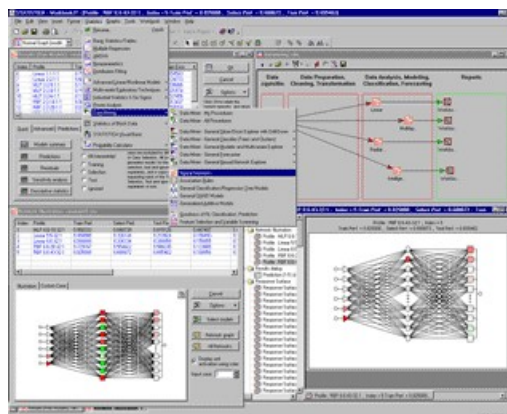


Рис. 3.38. Мережеві ансамблі.

У пакеті STATISTICA Neural Networks є численні засоби, що полегшують користувачеві вибір відповідної архітектури мережі. Статистичний і графічний інструментарій системи включає гистограми, матриці і графіки помилок для всієї сукупності і за окремими спостереженнями, підсумкові дані про вірну або невірну класифікацію, а всі важливі статистики - наприклад, пояснена частка дисперсії - обчислюються автоматично. Для візуалізації даних в пакеті реалізовані діаграми розсіяння і тривимірні поверхні відгуку, що допомагають користувачеві зрозуміти «поведінку» мережі (рис. 3.39).

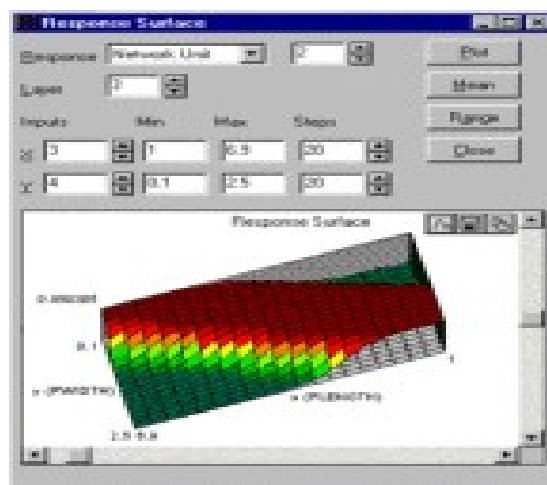


Рис.3.39. Візуалізація даних в пакеті STATISTICA Neural Networks

Для навчання багат шарових перцептронів в системі STATISTICA Neural Networks реалізований, перш за все, метод зворотного розповсюдження - із змінною в часі швидкістю навчання і коефіцієнтом інерції, перемішуванням спостережень перед черговим кроком алгоритму і додаванням аддитивного шуму для робастного узагальнення. Крім цього, в системі реалізовано два швидкі алгоритми другого порядку - методи зв'язаних градієнтів і Левенберга-Маркара. Останній є незвичайно могутнім сучасним алгоритмом нелінійної оптимізації. Ітеративний процес навчання супроводжується автоматичним відображенням поточної помилки навчання і обчислюваної незалежно від неї помилки на перевіірочній множині, при цьому показується і графік сумарної помилки. Якщо перенавчання має місце, це не повинно турбувати користувача:

пакет автоматично запам'ятовує екземпляр якнайкращої мережі, отриманої в процесі навчання, і до цього варіанту мережі завжди можна звернутися, натиснувши відповідну кнопку. Після того, як навчання мережі завершено, Ви можете перевірити якість її роботи на окремому тестовому множині (рис. 3.40).

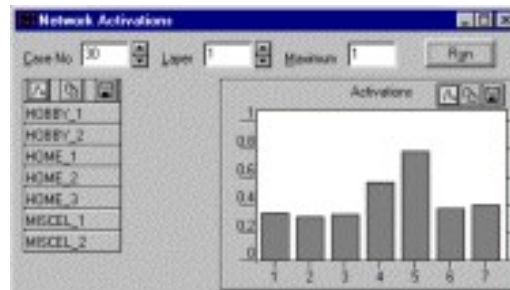


Рис. 3.40. Ітеративний процес навчання.

У системі STATISTICA Neural Networks є інтелектуальні засоби, що дозволяють відрізувати шматки від вже наявних мереж і сполучати декілька мереж воедино. Так, можна видаляти або додавати окремі нейрони, видаляти з мережі цілком деякий шар, а мережі, узгоджені по числу входів/виходів, послідовно сполучати одна з одною. Завдяки цим можливостям пакет дозволяє використовувати такі засоби, як пониження розмірності за допомогою асоціативних мереж і матриця втрат (для ухвалення рішень з найменшими втратами). Матриця втрат автоматично використовується при роботі з імовірнісними нейронними мережами (рис. 3.41).

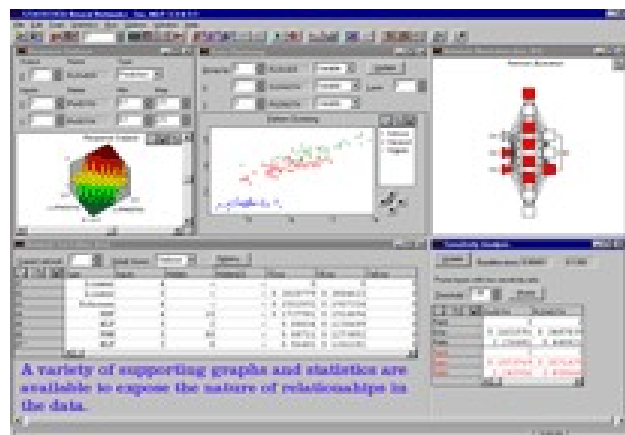


Рис. 3.41. Переформатування нейронних мереж.

Нейропакет NeuroSolutions. Програмний комплекс призначений для моделювання великого набору нейронних мереж. Основне його достоїнство полягає в гнучкості: крім традиційних нейросетевих парадигм (повнозв'язних і багат шарових нейронних мереж, карт Кохонена, що самоорганізуються) нейропакет включає могутній редактор візуального проектування нейронних мереж, що дозволяє створювати будь-які нейронні структури і алгоритми їх навчання, а також вводити власні критерії навчання. Пакет має хороші засоби візуалізації структур, процесів і результатів навчання і функціонування нейронних мереж. Це ставить даний нейропакет на рівень САД-систем (систем автоматизованого проектування) і моделювання нейронних мереж.

Крім засобів взаємодії з операційною системою (OLE), нейропакет забезпечений генератором початкового коду і дозволяє використовувати зовнішні модулі при створенні і навчанні нейронної мережі. Пакет підтримує програми, написані на мові C++ для компіляторів Microsoft Visual C++ і Borland C++, а також у вигляді DLL-коду. Таким чином, NeuroSolutions є гнучкою відкритою системою, яку можна при необхідності доповнювати і модифікувати. Пакет містить вбудовану макромову, що дозволяє проводити практично будь-яку настройку під конкретне завдання.

У пакеті реалізується великий перелік нейронів, включаючи зважений суматор (нейрон першого порядку), нейрони вищих порядків (з перемножуванням входів), а також безперервний інтегруючий нейрон. Функція активації нейрона може бути вибрана з п'яти стандартних функцій, а також задана користувачем. Зв'язки між нейронами задаються довільно на етапі проектування і можуть бути змінені в процесі роботи. Підтримуються всі типи зв'язків: прямі, перехресні і зворотні. При цьому добре реалізована схема організації зв'язків: можна задати один векторний зв'язок із заданою ваговою матрицею, а не набір скалярних зв'язків з ваговими коефіцієнтами.

Нейропакет містить могутні засоби для організації навчальних вибірок. Вбудовані конвертори даних підтримують графічні зображення, текстові файли з числовими або символічними даними, а також функції безперервного

аргументу (наприклад, часу), задані в аналітичному вигляді або у вигляді вибірки значень. Нейропакет дозволяє використовувати будь-які зовнішні конвертори даних.

На етапі навчання може бути використаний широкий круг критеріїв навчання, як дискретних, так і безперервних. Крім цього можна вводити власні критерії. Також можливо використовувати як вбудований алгоритм навчання типу back-propagation або дельта-правила, так і використовувати власний. Система візуалізації процесу навчання дозволяє проводити аналіз зміни вагів безпосередньо в процесі навчання і вносити корективи. Пакет містить генератор (майстер) стандартних нейромережевих архітектор (Neural Wizard), за допомогою якого швидко задається архітектура, підбирається навчальна вибірка, критерії і методи навчання нейронної мережі.

Нейропакет NeuralWorks Professional II/Plus. NeuralWorks Professional є могутнім засобом для моделювання нейронних мереж. У ньому реалізовано 28 нейронних парадигм, а також велику кількість алгоритмів навчання. Додатковий модуль UDND (User Define Neural Dynamics) дозволяє створювати власні нейронні структури. Як і NeuroSolutions, NeuralWorks Professional має хорошу систему візуалізації даних: структури нейронної мережі, зміни помилки навчання, зміни вагів і їх кореляції в процесі навчання. Останнє є унікальною властивістю пакету і корисна при аналізі поведінки мережі. У NeuralWorks Professional можна інтегрувати зовнішні програмні модулі. Він має вбудований генератор коду, що підтримує компілятор Microsoft Visual C++.

Нейропакет Brain Maker Pro. Одним з поширених зарубіжних нейросистем є пакет Brain Maker. Хай необхідно вирішити задачу прогнозу ціни закриття на сьогоднішніх торгах по валютних тисячедоларових тримісячних ф'ючерсних контрактах. Хай нас влаштовує точність прогнозу, при якій правильно вказується ціновий тренд і зміна ціни з точністю не нижче 90% від останнього стрибка. Застосування нейронної мережі починається з підготовки вхідних даних: курс долара, індекс інфляції, ставка межбанка,

біржові індекси, об'єм торгів, кількість операцій, максимальні і мінімальні ціни і ін. Після попередньої настройки мережі починається ітераційний процес навчання, в результаті якої нейромережа настраює свою логічну структуру для точної реакції ринку на ті або інші дії. Для цього в пакеті Brain Maker передбачений могутній аналітичний блок, який дозволяє побачити, які параметри роблять позитивний вплив на ситуацію, а які - негативний. Потім мережа знову навчається і далі тестується на якість і адекватність, і після вдалого тестування використовується для прогнозів. За десять біржових днів мережа жодного разу не помилилася в знаку відхилення ф'ючерських котирувань, а дев'ять днів з десяти відхилення прогнозу від реальної ціни склало менше 10 доларів. Brain Maker - це програма, з якою почалася історія застосування нейронних мереж в Росії та Україні. У цьому пакеті на професійному рівні реалізована класична багат шарова нейронна мережа. Це єдина програма, в якій є можливість настройки всіх параметрів нейронних мереж і алгоритмів навчання. Останнім часом Brain Maker найчастіше використовується не як самостійна програма, а як надбудова до програми TradeStation для аналізу в режимі реального часу.

Нейропакет NeuroShell 2. Пакет є однією з трьох програм, що входять до складу пакету The AI Trilogy і є універсальним нейропакетом для моделювання декількох найбільш відомих нейронних парадигм: багат шарових мереж, мереж Кохонена і т.д. З його допомогою можна вирішувати широкий спектр завдань, починаючи з широко поширених задач прогнозування курсів акцій і закінчуючи менш поширеними зворотніми задачами геофізиці. Він включає систему для початківця, систему для професіонала, засоби автономного використання. Крім того, існують пакети доповнень до NeuroShell 2, які дозволяють істотно спростувати рішення окремих задач. Зокрема:

Пакет ринкових індикаторів з оптимізацією містить близько 150 найбільш відомих і могутніх технічних індикаторів, які можна застосовувати для фінансових даних або інших часових рядів.

Прогноз результатів скачок. Виконує передобробку статистичних даних про результати скачок для того, щоб зробити прогнози точнішими. Програма бере дані про результати всіх коней в одному забігу і перетворює їх у файл з великою кількістю рядів, в кожному з яких приведені результати тільки для 2 коней. При такій формі представлення даних точність прогнозів нейромережі значно підвищується. Це доповнення застосовне не тільки для прогнозу результатів скачок, але і для вирішення інших задач, пов'язаних з ранжируванням даних.

Прогнозування результатів виборів (ранжирування). Дозволяє успішно застосовувати нейронні мережі для прогнозування результатів виборів за наслідками ряду передуючих їм регулярних соціологічних опитів. Вхідні дані можуть включати дані про результати опитів, однакові для всіх кандидатів, а також дані для кожного кандидата. Кожен тренувальний приклад може включати дані одного опиту для декількох кандидатів, що беруть участь у виборах. Кількість кандидатів в різних опитах може бути різним. Це доповнення здійснює передобробку даних так, щоб мережа могла порівнювати за один раз тільки по двох кандидатів. Це спрощує тренування мережі і збільшує точність прогнозів.

Складовою частиною нейропакету є програмний комплекс **NeuroShell Trader** - система, призначена для прогнозування і пошуку ефективних торгових стратегій на фінансових ринках. NeuroShell Trader - це сімейство продуктів, розроблене спеціально для трейдерів і покликане допомогти їм в ухваленні рішень при торгівлі. В них реалізовані технології штучного інтелекту, що дозволяють прогнозувати фінансові часові ряди, будувати і оптимізувати торгові стратегії. Пакет спочатку розроблявся як інструмент для нейромережевого аналізу біржових даних, тому побудова в ньому прогнозів і торгових стратегій за допомогою нейронних мереж і генетичних алгоритмів проста і зрозуміло навіть для користувача, що не є професіоналом в цій області. Крім того, будучи спеціалізованим інструментом для трейдерів, NeuroShell Trader має дружній графічний інтерфейс, багаті можливості для імпорту даних і

могутню бібліотеку індикаторів. Він дозволяє відображати біржові дані на робочих листах у вигляді графіків, діаграм, японських свічок, дозволяє використовувати більше 800 вбудованих індикаторів і створювати свої індикатори, а також одночасно працювати з декількома фінансовими інструментами. Нейронна мережа або індикатор, побудовані для одного фінансового інструменту, автоматично застосовуються до всіх інструментів, які вказуються. NeuroShell Trader дозволяє користувачеві отримати важливу статистику та іншу інформацію, що стосується застосування торгових стратегій і нейронних мереж до даних (рис. 3.42).



Рис. 3.42. Нейромережевий аналіз біржових даних.

NeuroShell Series. Нейромережева архітектура, яка лежить в основі програм даної серії, є найостаннішими досягненнями наукового пошуку, результатом якого з'явилося створення алгоритму «самопобудови» нейронної мережі, що володіє рекордними швидкостями навчання. Ці програми надзвичайно прості у використанні. Тепер користувач повинен зосередитися тільки на формулюванні завдання, все інше програми даної серії зроблять самі. До складу серії входять:

- **NeuroShell Predictor** – Провісник.
- **NeuroShell Classifier** – Класифікатор.

- **NeuroShell Run-time Server** - Засоби автономного використання мереж, отриманих в NeuroShell Predictor і NeuroShell Classifier.

Пакет NeuroShell Predictor дає можливість створювати системи для вирішення задач прогнозування на основі наявної бази даних. Це можуть бути прогнози наступних значень параметрів часового ряду, наприклад, прогноз курсу акцій, або оцінка якої-небудь величини, визначуваної набором незалежних чинників, наприклад, оцінка вартості квартир або вживаних автомобілів (рис. 3.43).

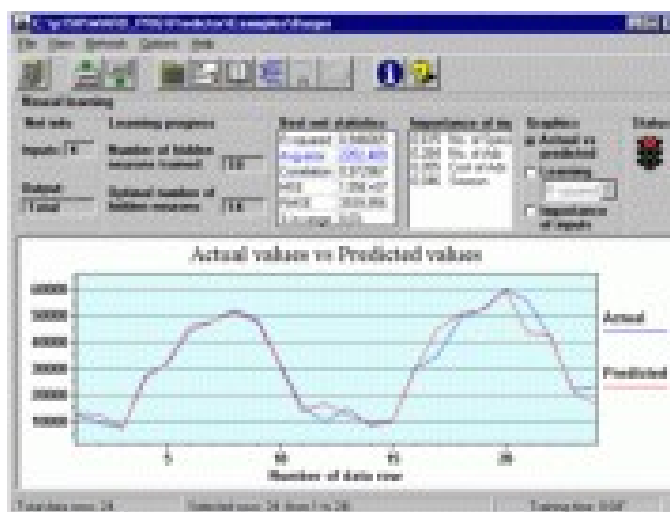


Рис. 3.43. Застосування пакету NeuroShell Predictor.

NeuroShell Predictor настільки простий у використанні, що відпадає необхідність в керівництві користувача. Замість нього в програмі є «Інструктор», який проведе через всі етапи створення моделі для прогнозів. Інструктор дає можливість вибрати стовпці у файлі даних, які використовуватимуться як входи, і вказати один стовпець як вихід.

Пакет NeuroShell Classifier призначений для вирішення задач розпізнавання образів, пов'язаних з визначенням приналежності образу (ситуації) до тієї або іншої категорії. Наприклад, по набору біржових показників виробляти сигнал для покупки або продажу акцій тієї або іншої компанії. Пакет був розроблений, подібно NeuroShell Predictor, для тих, хто не мав попереднього досвіду роботи з нейронними мережами. На відміну від

програми NeuroShell Predictor, яка на виході нейронної мережі дає безперервне значення величини, що передбачається, нейронна мережа NeuroShell Classifier має декілька виходів, які визначають вірогідність приналежності пред'явленого образу до кожної з декількох категорій. Як приклади категорій можна привести такі, як { купити, продати, утриматися від операцій} або { ракова пухлина, доброякісна пухлина}. У NeuroShell Classifier реалізовано два ефективні алгоритми класифікації : один є спеціальною нейронною мережею, а інший - статистичний класифікатор, керований генетичним алгоритмом. Дані алгоритми були розроблені спеціально для того, щоб зробити моделі класифікації точнішими, а також щоб «не сушити голову» кожного разу при створенні моделі, підбираючи її параметри. (рис. 3.44).

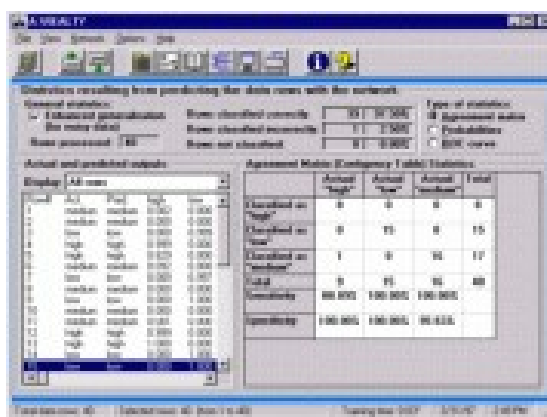


Рис. 3.44. Задача класифікації в пакеті NeuroShell Classifier.

Пакет NeuroShell Run-time Server містить ряд програм, які дозволяють використовувати мережі, створені за допомогою NeuroShell Predictor і NeuroShell Classifier, або з робочих листів Microsoft Excel, або у власних програмах.

Програмний комплекс DEDUCTOR. Deducator Studio є аналітичним ядром платформи Deducator. Він містить повний набір механізмів імпорту, обробки, візуалізації і експорту даних для швидкого і ефективного аналізу інформації. В пакеті використовуються наймогутніші технології, такі як

багатовимірний аналіз, нейронні мережі, дерева рішень, карти Кохонена, спектральний аналіз і множина інших. При цьому акцент зроблений на самонавчальні методи і машинне навчання, що дозволяє будувати адаптивні системи, тобто здатні реагувати на зміну ситуації. Використання самонавчальних методів і майстрів для настройки, дозволяє понизити вимогу до підготовки персоналу, роблячи сучасні технології доступними широкому колу користувачів. Всі механізми уніфіковані і виконуються за допомогою майстрів (рис. 3.45).

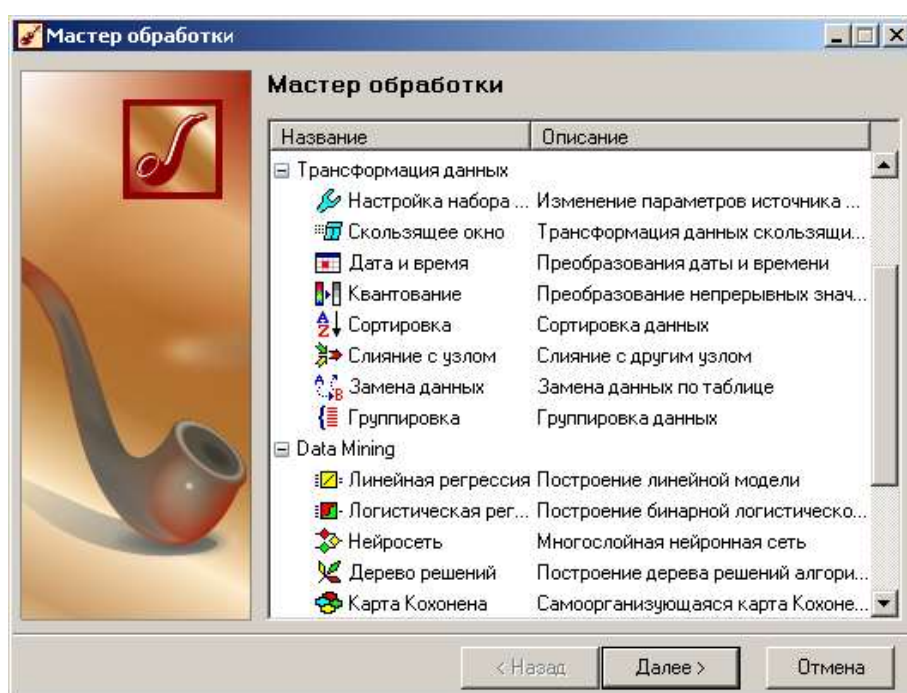


Рис. 3.45. Аналітичні технології платформи Deductor.

Нейронні мережі в пакеті підтримуються за допомогою бібліотеки компонентів NeuralBase. Як приклад, створені компоненти, що реалізують дві нейромережеві парадигми: рекурентну нейронну мережу (мережу Хопфілда) і багат шарову нейронну мережу навчену по алгоритму зворотного розповсюдження помилки. Основним призначенням бібліотеки є інтеграція нейронних мереж в інформаційні системи, для розширення аналітичних можливостей систем. Реалізація нейронних мереж у вигляді компонентів, наявність відкритого коду дозволяє легко вбудовувати їх в інші програми.

Існує три базові класи TNeuron, TLayer, TNeuralNet. Всі інші є похідними від них. TNeuron є базовим класом для нейронів, несе всю основну функціональність, має індексовану властивість Weights, що є ваговими коефіцієнтами (синапсами), властивістю Output, яка є виходом нейрона (результатом обчислень) і суматор, роль якого, виконує метод ComputeOut. TNeuronHopf, нащадок TNeuron, реалізує нейрон використовуваний в нейронній мережі Хопфілда, єдиною відмінністю від базового класу, є використання активаційної функції в перекритому методі ComputeOut. Наступним породженим класом, є TneuronBP, яка служить для програмної реалізації багат шарових нейронних мереж. Основним призначенням базового класу TLayer і його нащадків TLayerHopf і TLayerBP є об'єднання нейронів в шар, для спрощення роботи з нейронами. Компонент TNeuralNetHopf реалізує нейронну мережу Хопфілда.

Система 4Thought. Даний продукт входить в сімейство засобів Business Intelligence, пропонує одним з провідних розробників інструментальних засобів для створення аналітичних додатків в бізнесі, канадсько-американською компанією Cognos. Засоби аналізу результатів, передбачені в даній системі роблять її зручним інструментом як для кінцевого користувача-аналітика, так і для фахівця-математика. Інтеграція із засобами OLAP (система PowerPlay) дозволяє застосовувати нейромереві методи аналізу і прогнозування безпосередньо до гіперкубів PowerPlay.

Використовуючи технологію нейронних мереж, 4Thought дозволяє аналізувати ефективність роботи підприємств і їх підрозділів, віддачу від інвестицій або різних видів реклами і т.д. Параметр, що характеризує ефективність, - це цільовий показник, а всі інші, що впливають на нього ми називатимемо чинниками (причому значення показника і чинників можуть бути як числовими, так і символічними). Початкові дані аналізу представляються у вигляді таблиці, один із стовпців якої - цільовий показник (наприклад, прибуток або об'єм продажів), а інші - впливаючі чинники (витрати на рекламу, пора року, регіон і ін.). Пакет будує кількісну модель залежності значень

показника від значень чинників, після чого дозволяє проводити перехресний аналіз (проглядати в графічній і аналітичній формі залежність модельованого показника від будь-якого з вибраних чинників при фіксованих або усереднених значеннях інших чинників), перевіряти гіпотези «що якщо», оцінювати значущість чинників за мірою їх впливу на цільовий показник, а також використовувати отриману модель для прогнозування значення показника виходячи з відомих значень чинників. Розглянемо декілька прикладів, в основі яких лежать реальні додатки.

Аналіз роботи філіалів. В даному прикладі досліджувався вплив різних чинників (регіон, масштаб і види діяльності) на ефективність роботи філіалу компанії. До регіональних чинників можна віднести категорію регіону і тип населеного пункту, де розташований торговий філіал: міський або сільський. Масштаб діяльності може характеризуватися величиною обороту і потоками наявних засобів, а види діяльності - видами товарів. Як цільовий показник можна вибрати величину невиробничих витрат філіалу. Для ефективного управління компанією необхідно розуміти, від чого залежать витрати в кожному філіалі та що є вирішальним чинником. На підставі таблиці, що характеризують дані по філіалах, 4Thought дозволяє будувати модель впливу вибраних стовпців-чинників на величину витрат філіалу. Це досить складне завдання, яке неможливо вирішити традиційними статистичними методами. Система використовує дані про роботу філіалів для навчання нейронної мережі, яка моделює залежність витрат філіалів від різних чинників. Готову модель можна використовувати для кількісного аналізу впливу чинників на витрати філіалів. Зокрема, побудувавши залежність модельних значень витрат від обороту (перехресний аналіз), можна оцінити результат економії на масштабах: чим більше оборот, тим менше витрачання. Модель дозволяє досліджувати навіть ті чинники, які не були в неї введені. Наприклад, якість управління філіалом не можна пов'язувати безпосередньо з його прибутковістю - на прибутковість впливають і інші чинники. Проте можна порівняти фактичні показники ефективності (наприклад, витрати) з очікуваними, які обчислюються

за допомогою побудованої моделі, а отже, в них враховується вплив регіональних і інших чинників. Якщо очікувані витрати вище фактичних, значить, ефективність роботи філіалу вище середнього, а це, швидше за все, викликано високою якістю управління (і навпаки).

Окупність капіталовкладень. Сьогодні багато компаній вкладають величезні засоби в розвиток і використання інформаційних технологій. Але як оцінити, наскільки ефективно використовуватиметься обчислювальна техніка і складне програмне забезпечення? Чи окупляться капіталовкладення в цю сферу? Яким чином вплине впровадження нової технології на доходи підприємства і на продуктивність співробітників? Відповіді на ці питання складно - термін окупності часто дуже великий, а на ефективність роботи компанії впливають і інші чинники. Як приклад візьмемо мережу магазинів роздрібною торгівлі, де встановлювалася інформаційна система для касових терміналів в 2000 точок збуту. Об'єм інвестицій склав 75 млн. дол. Передбачалося, що завдяки впровадженню нової системи вдасться підвищити рівень обслуговування клієнтів. Проте після вкладення чверті грошових коштів виникли сумніви в окупності проекту. Задача оцінки окупності інвестицій виявилася складною. Вплив саме цих капіталовкладень на зростання валового прибутку в кожному окремому підрозділі необхідно було відокремити від впливу інших чинників: кваліфікації менеджерів, розширення філіалу, недавніх витрат на ремонт і т.д. По всіх торгових точках були зібрані дані про перераховані чинники, а також інформація про зростання доходів і про те, скільки місяців пройшло з моменту впровадження інформаційної системи. Щоб оцінити ефективність впроваджуваної системи потрібно було побудувати модель зростання доходів залежно від комбінації даних чинників. За допомогою системи 4Thought був проведений аналіз взаємозалежності з метою отримання кількісної оцінки дії кожного чинника окремо (рис. 3.46).

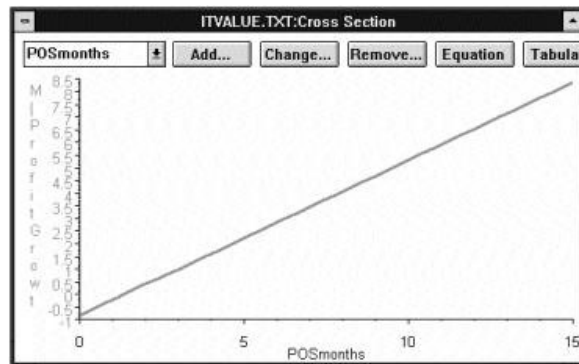


Рис. 3.46. Залежність прибутку компанії від терміну впровадження.

Побудований системою графік показує, що чим довше працює інформаційна система (горизонтальна вісь), тим більше зростання прибутку (вертикальна вісь). Зокрема, видно, що інвестиції повністю окупляться протягом 12 місяців. Отримавши таку інформацію, менеджери без коливань вирішили продовжити вкладення тих, що залишилися 75% коштів.

Оптимізація цін. Один з найкорисніших додатків нейроаналітики - оптимізація цін. Очевидно, чим більше ціна на товар, тим вище прибуток від його продажу. Проте у міру підвищення ціни об'єм продажів починає знижуватися, і, може так трапитися, що товар перестануть купувати зовсім. Але якщо ціна дуже низька, то отримати прибуток від збуту не вдасться. Необхідно встановити деяке проміжне, оптимальне значення. У системі 4Thought для вирішення цього завдання шляхом аналізу даних про продажі за минулий період з'ясовується, як міняються витрати і об'єми продажів залежно від ціни на товар. Величина оптимальної ціни залежить від реакції клієнтів на зміну цін, від собівартості об'єму продукції, а також від зовнішніх чинників.

Компанія, що займається ремонтом автомобілів, використовувала 4Thought для виставляння оптимальних цін на свої послуги. Щоб визначити чутливість клієнтів до зміни цін, система повинна встановити, яким чином у минулому мінявся попит на продукт залежно від різних чинників. Що стосується компанії по ремонту автомобілів, то тут необхідно було промодельовати зміна попиту залежно від цін на послуги самої фірми, середнього рівня цін конкурентів, пори року, витрат на рекламу, а також певних

змін в законодавстві, які вплинули на структуру ринку. Були зібрані дані за три роки. Інформацію про ціни конкурентів вдалося отримати від менеджерів філіалів, які добре знайомі з діяльністю інших фірм. Щоб визначити, які чинники впливають на попит, була побудована модель зміни об'єму продажів відповідно до цін на послуги, рівнем цін конкурентів, порою року і витратами на рекламу. Ця інформація використовувалася для оцінки співвідношення ціна/об'єм продажів.

На рисунку 3.47 показано, як росте попит на послуги (вертикальна вісь) при зниженні цін (горизонтальна вісь). Щоб визначити оптимальну ціну, необхідно побудувати модель витрат, що показує, яким чином міняються витрати залежно від об'єму продажів. Витрати - розрахункова величина, яку можна ввести в модель за допомогою обчислюваного стовпця. Комбінуючи моделі продажів і витрат, можна отримати повну картину впливу різних чинників на прибутковість бізнесу. Зокрема, можна встановити залежність між ціною і прибутком.

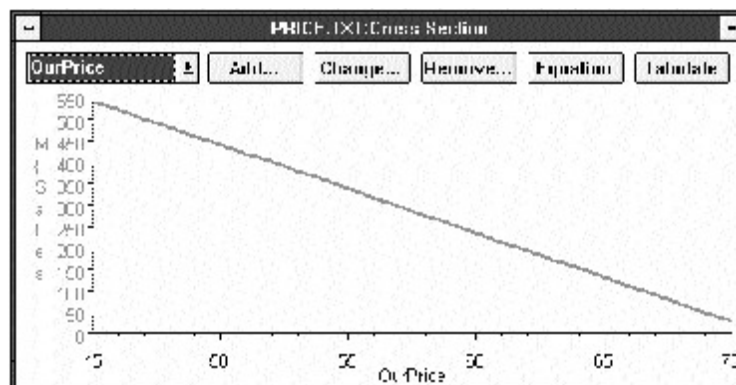


Рис. 3.47. Залежність попиту від цін на послуги.

На рисунку 3.48 показана модель зміни прибутку (вертикальна вісь) залежно від зміни цін на послуги (горизонтальна вісь). У такий спосіб вдається визначити оптимальну ціну для будь-якої іншої галузі бізнесу. Компанія по ремонту автомобілів використовувала отриману інформацію для встановлення цін на свої послуги, при цьому тільки за рахунок правильного ціноутворення їй вдалося підвищити прибуток на декілька млн. дол.

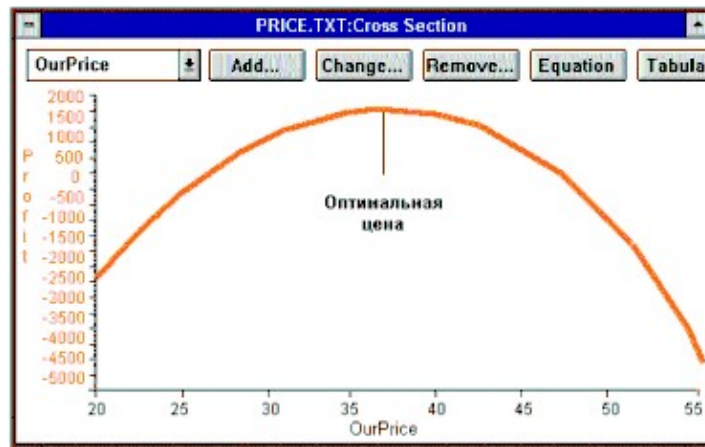


Рис. 3.48. Залежність прибутку компанії від цін на послуги.

Система SENN Sales. Даний продукт є спеціалізованим засібом для інтелектуального аналізу даних у фінансовій і комерційній сферах. Як і 4Thought, ця система також може завантажувати інформацію з корпоративних баз даних без обмежень на кількість стовпців і рядків в таблицях. SENN Sales використовується для вирішення задач профілізації клієнтів, маркетингового аналізу, прогнозування попиту і т.д. Розглянемо досвід її застосування для прогнозування фінансових рядів.

Надійне і точне прогнозування ф'ючерсних курсів обміну валют, кредитних ставок, цін і динаміки продажів дозволяє укласти вигідні міжнародні валютні угоди, формувати фінансові і інвестиційні пакети, давати точну оцінку поточної ситуації на ринку і т.д. Дані за минулі періоди (індекси, курси обміну, кредитні ставки і криві продажів) містять структурні залежності, виявивши які, можна визначити поведінку системи в майбутньому. Використовуючи метод моделювання, що забезпечує точне відтворення динаміки поведінки системи, можна описати залежності в наявних даних і побудувати прогноз.

Оптимальне управління портфелем цінних паперів. При розв'язанні даної задачі за допомогою системи SENN результати прогнозування доходів від капіталовкладень оброблялися методами ухвалення оптимальних рішень. Спочатку для визначення якнайкращого способу розміщення капіталу розглядався індекс фондової біржі, а також довгострокові і короткострокові

кредитні ставки для кожної з країн (Англія, Франція, Німеччина, Японія і США). При цьому були прийняті до уваги зміни в системі рахунків в Європі і країнах великої сімки, викликані появою на фінансових ринках нової європейської валюти. При традиційному підході до управління портфелем прогнозується дохід на кожен із статей капіталовкладень: за допомогою методів нейроаналітики була отримана ставка доходу 21%. На основі побудованих моделей була розроблена стратегія розміщення фондів, що оптимізує дохід від капіталовкладень для портфеля активів і що оцінює можливі ризики.

3.5. Сучасна практика та перспективні напрямки застосування нейротехнологій.

Нейронні мережі все частіше застосовуються в реальних бізнес-додатках. У деяких областях, таких як виявлення фальсифікацій і оцінка ризику, вони стали безперечними лідерами серед використовуваних методів. Їх використання в системах прогнозування і системах маркетингових досліджень постійно росте. Варто відзначити, що оскільки, економічні, фінансові і соціальні системи дуже складні і є результатом дій і протидій різних процесів, то є дуже складним створити повну математичну модель з урахуванням всіх можливих дій і протидій. Практично неможливо детально апроксимувати модель, засновану на таких традиційних параметрах, як максимізація корисності або максимізація прибутку.

У системах подібної складності є природним і найбільш ефективним використовувати моделі, які безпосередньо імітують поведінку суспільства і економіки. А це якраз те, що здатна запропонувати методологія нейронних мереж. Розглянемо деякі важливі області, в яких ефективність застосування нейронних мереж доведена на практиці [9, 17, 22, 45, 70, 77].

Застосування нейронних мереж в менеджменті. Протягом останнього десятиліття в журналах і газетах, таких як «Management Science», «Man and Cybernetics», «Decision Sciences», «Computers & Operations Research», «European Journal of Operational Research» були надруковані незліченні пропозиції по застосуванню нейронних мереж в бізнесі і дослідженні операцій. Більшість варіантів застосування нейронних мереж в менеджменті стосуються задач, що потрапляють в наступні чотири категорії: класифікація, побудова емпіричної кривої і аналіз часових рядів, кластеризація і оптимізація. Нижче приведені приклади кожної категорії.

1. Класифікація. Належним чином розроблена нейронна мережа може використовуватися як класифікатор. Після навчання історичним даним нейронна мережа може визначати клас приналежності деякої характерної межі. Нейронні мережі можна використовувати при аналізі кредитоспроможності, щоб передбачити банкрутство фірми. Нейронні мережі можуть також оцінити активи і зобов'язання. У багатьох банках нейронні мережі можна використовувати для виявлення підробки кредитної картки.

2. Побудова емпіричної кривої і аналіз часових рядів. Процес навчання в багатьох типах нейронних мереж може розглядатися як побудова емпіричної кривої. Крім того, нейронні мережі можуть використовуватися для визначення моделі коливань часового ряду. Аналітики сфери маркетингу, використовуючи нейронні методи мережі, можуть визначати ринкові функції відгуку, засновані на часових даних. Керівники виробництва можуть передбачати продуктивність фірми, ґрунтуючись на кривих, представлених навченими нейронними мережами. Багато фінансових установ використовують нейронні мережі для фінансового прогнозу і управління інвестиціями.

3. Кластеризація. Неконтрольовані нейронні мережі, що навчаються, зазвичай використовуються в кластерному аналізі для групування об'єктів без апіорного знання класів. Ідентифікація споживчих сегментів і групування технологічних деталей можуть служити прикладом в цій прикладній категорії.

4. Оптимізація. Оскільки процес навчання в нейронних мережах повинен мінімізувати наперед певну помилку або енергію, то нейронні мережі можуть використовуватися для вирішення задач оптимізації. Задачі на зразок оптимального планування робіт, оптимального планування роботи магазину і мінімізації втрат можуть бути вирішені з використанням нейронних мереж.

Прогноз ризиків і рейтингування. Існують дві базові інвестиційні стратегії: активна, заснована на прогнозах прибутковості тих або інших активів, і пасивна, в якій ринок вважають непередбачуваними, і головною метою ставлять мінімізацію ризиків. Оцінка інвестиційного ризику, таким чином, є одним з наріжних каменів фінансового аналізу. Ці методики використовують два основні підходи: навчання з вчителем - на прикладах експертних оцінок або збанкрутілих фірм, і навчання без вчителя - шляхом категоризації наявних даних.

Спочатку розглянемо перший, прямолінійніший підхід. З цією метою систематизуємо поняття рейтингу цінних паперів.

Рейтинг корпоративних облігацій. Істотну частину ринку цінних паперів складають корпоративні облігації - позики корпорацій під фіксований відсоток. Тільки на Нью-Йоркській Фондовій біржі в 2005 році оберталися облігації близько 15000 компаній із загальною номінальною вартістю понад \$260 млрд. Для оцінки ризику невиконання відсотків або неповернення грошей по облігації практично для всіх таких корпорацій існують і періодично оновлюються рейтинги, що складаються незалежними рейтинговими агентствами.

У рейтинговому бізнесі домінують декілька компаній, наприклад, Standard & Poor's і Moody's. Понад 4000 боргових емітентів поставляють свої фінансові звіти цим двом організаціям. Рейтинги цих агентств надзвичайно авторитетні, від них безпосередньо залежать процентні ставки по облігаціях: чим нижче рейтинг емітента - тим дорожче обходиться емітенту обслуговування свого боргу, оскільки інвестори бажають отримати плату за додатковий ризик. Більш того, в США деяким категоріям інвесторів, таким, як

банки і страхові компанії, законодавчо заборонено купувати облігації з рейтингом Standard & Poor's і Moody's нижче певного рівня.

Алгоритм складання описаних вище рейтингів невідомий, більш того, агентства стверджують, що він не заснований в чистому вигляді на статистичному аналізі фінансової інформації, а містить ще оцінки експертів, наприклад для таких параметрів, що важко формалізуються, як «якість менеджменту». Така ситуація цілком влаштовує самі рейтингові агентства, перетворюючи їх продукцію на унікальний товар. Проте багато інвесторів зацікавлено у володінні своїми власними алгоритмами рейтингування, що «емюлюють» рейтинги великої двійки, - принаймні по трьом причинам.

- По-перше, не для кожної облігації є офіційний рейтинг. Багато паперів, обійдених увагою крупних рейтингових агентств, можуть у результаті виявитися вельми привабливими для інвестицій, якщо зуміти грамотно оцінити ступінь їх ризикованої.

- По-друге, оновлення офіційних рейтингів відбувається не так часто, як хотілося б. Уміння заздалегідь, до того як це стане загальнодоступною інформацією, передбачити зміну рейтингів, очевидно, дає інвесторам додаткові конкурентні переваги.

- Нарешті, розгадавши стратегію «офіційного» рейтингування, інвестори можуть сподіватися поліпшити якість оцінки фінансового стану емітентів шляхом інтенсивнішого статистичного аналізу, отримавши, таким чином, перевагу над тими, хто користується офіційними рейтингами.

Приведені вище доводи обґрунтовують наступну постановку задачі для нейроаналіза: на основі загальнодоступної фінансової звітності компаній-емітентів постаратися відтворити рейтинги Standard & Poor's або Moody's. Не дивлячись на наявність неформальної компоненти, представляється вірогідним, що алгоритмічна складова цих рейтингів досить велика. Врешті-решт, загальна чисельність аналітиків в обох провідних агентствах разом узятих не перевищує 100 чоловік. Отже справитися з обробкою постійно оновлюваних даних про

4000 емітентів вони можуть лише використовуючи в основному автоматизовані процедури.

Спроби змоделювати алгоритм рейтингування облігацій робилися з 60-х років і базувалися на методі лінійної регресії. Типовий відсоток вгадування рейтингу в цих моделях складає приблизно 60%. Оскільки можливості нелінійного нейромережевого моделювання ширші, не дивно, що перші ж спроби застосувати нейромережі показали істотно кращі результати - на рівні 88% для відтворення окремої градації рейтингу. Складніші нейромережеві моделі здатні з прийнятною точністю відтворювати широкий діапазон рейтингів облігацій по набору ключових фінансових індикаторів фірм-імітентів. Відмітимо, що не дивлячись на непогані, загалом, результати, подібні нейромережеві моделі вельми компактні. В якості вхідних змінних зазвичай використовується від 6 до 10 фінансових індикаторів, що є відношенням найбільш значущих статей балансів і звітів про прибутки і збитки корпорацій. Якість відтворення «тонких» градацій рейтингу агентства Standard & Poor's, досягнуте цією моделлю, ілюструється на малюнку 3.49.

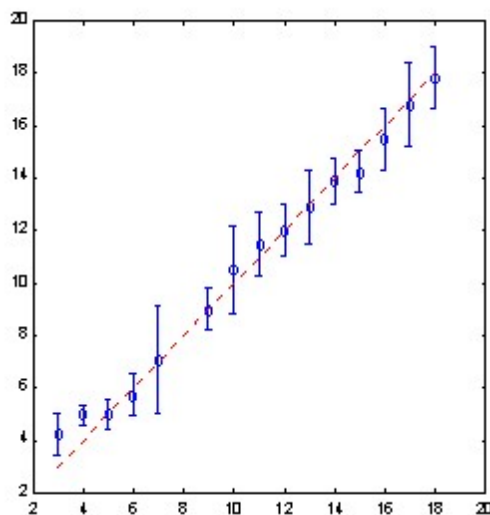


Рис. 3.49. Відтворення 15-ти градацій рейтингу Standard & Poor's.

Оцінка акцій. На відмінність від облігацій акції корпорацій не гарантують повернення відсотків і основної суми боргу. Проте, оцінка перспективності різних активів в пакетах акцій є одним з головних завдань

будь-якого інвестора. Одним з підходів до цього є рейтингування акцій, засноване на ширшому крузі фінансових показників компаній, доступних з їх фінансових звітів. Результативність подібного підходу ілюструють рейтинги провідного консультативного агентства США по інвестиціях в акції - Value Line. Раз в тиждень це агентство розбиває акції близько 1700 компаній по 5 рейтинговим категоріям. Статистичні дослідження підтверджують значущість рейтингу Value Line. А саме, пакети, складені з акцій вищої рейтингової категорії, систематично дають великий прибуток в перебігу найближчого кварталу. Є підстави припускати, що квартальні звіти корпорацій впливають на курс акцій. Зокрема, несподівано високі прибутки (збитки) статистично значущо корелюють з підвищенням (пониженням) курсів акцій. Причому, ця кореляція існує достатньо довго - в перебігу принаймні двох місяців з дня публікації звіту. Отже, інвестор має можливість отримати певну вигоду з фінансової звітності корпорацій.

У більш загальній постановці йдеться про прогнозування фінансового «здоров'я» корпорації на підставі її фінансової звітності. Нетривіальним моментом тут є кількісне визначення фінансового благополуччя. Можна, як і у випадку з облігаціями, скористатися для навчання мережі рейтингами, наприклад, згаданого вище агентства Value Line для відтворення цієї, загалом суб'єктивної оцінки компанії. Можна спробувати використовувати як індикатор благополуччя об'єктивніший критерій - ринковий курс акцій в найближчому або віддаленішому майбутньому. Проте, ринковий курс може бути схильний до сильних флуктуацій чисто спекулятивного характеру. Нарешті, можна скористатися вказівками найсுவорішого вчителя, досліджуючи крайню форму прояву фінансового «нездужання» - банкрутство. Аналіз банкрутства, таким чином, може служити джерелом об'єктивних оцінок стійкості фінансового положення фірм.

Прогноз ризиків банкрутства. Спочатку приведемо декілька цифр, що ілюструють «ціну питання». Світовий ринок тільки міжбанківських кредитів оцінюється в \$58 трлн. Це майже в два рази перевищує світовий об'єм цінних

паперів. Природно, що оцінка ризику неповернення кредитів має для банків першорядне значення. Кількість банкрутств в США впродовж 2000 - 2005 років зростала щорічно приблизно на 14%. Таким чином, прогноз банкрутств, особливо в кризових економічних умовах, є насущним завданням економічного аналізу.

Якщо в проблемі рейтингування завданням нейромережі було відтворити думки експертів про надійність корпорації, то нейромережевий прогноз банкрутств заснований на статистичній обробці конкретних прикладів банкрутств. У такій постановці задачі нейромережі важливо самій стати експертом, що визначає фінансову стабільність корпорації, ґрунтуючись виключно на об'єктивній інформації - показниках фінансової звітності. Зазвичай від нейромережі потрібно оцінити вірогідність банкрутства через певний проміжок часу по доступній на даний момент фінансовій звітності. Як входи використовують фінансові індикатори - відносини балансових статей, що найбільш повно відображають певні сторони фінансового положення фірми. Найбільш загальний підхід полягає у використанні в якості входів логарифмів укрупнених статей балансів і звітів про прибутки (збитки). Нейромережа в цьому випадку сама вибере найбільш значущі лінійні комбінації входів, яким відповідатимуть найбільш значущі відносини різних статей в потрібних пропорціях. Використання індикаторів, з іншого боку, допомагає в інтерпретації результатів нейромоделювання якщо скористатися, наприклад, технікою проріджування зв'язків і витягання правил. Відмітимо, що використання описаних вище індикаторів лежить також в основі загальноприйнятої методики рейтингування банків CAMEL.

Узагальнюючи досвід порівняльного аналізу прогнозів банкрутств різними методиками, відзначимо:

- Нейромережеве моделювання забезпечує якнайкращу точність прогнозу банкрутств: близько 90%, в порівнянні з 80%-85% точністю для інших статистичних методик.

- При бажанні можна підвищити «підозрілість» нейромережі, забезпечивши точність виявлення банкрутів аж до 99% - за рахунок зниження вимог до помилок.

- Банкрутства можна упевнено передбачати за декілька років до їх фактичного настання, причому точність прогнозу за два роки практично не відрізняється від точності прогнозу за рік. Таким чином, неявні сигнали неблагополуччя присутні у фінансовій звітності фірми задовго до її краху.

Корисність навчання мережі на прикладах збанкрутілих фірм полягає також в тому, що така мережа виробляє функцію - чисельний показник фінансового здоров'я фірми, міру її стійкості. Проте, стійкість не є єдиним можливим критерієм оцінки діяльності фірми. Акціонери, наприклад, зацікавлені не тільки в нескінченно довгому існуванні фірми, але і в отриманні достатньо вагомого прибутку. Важливо, крім того, не тільки стан фірми на справжній момент, але і характеристики існуючих тенденцій. Тут значущим може опинитися інший набір чинників, що дає іншу оціночну функцію. Так, висока прибутковість може забезпечити підвищення надійності в майбутньому. Тим часом, неясно яким чином можна навчати нейромережу на «майбутній успіх» за відсутності такого ж чіткого критерію успіху, яким є банкрутство для невдачі. Ці об'єктивні труднощі можна подолати, якщо пригадати, що фірма існує не сама по собі, а в співтоваристві подібних нею фірм-конкурентів. І саме в зіставленні з цим співтовариством можна говорити про сильні і слабкі сторони її діяльності. Ці міркування підводять нас до іншої постановки задачі: комплексної оцінки фінансового стану фірми шляхом систематичного порівняння її показників з показниками решти учасників даного ринку. Такий підхід не вимагає знання готових відповідей, оскільки заснований на навчанні без вчителя.

Порівняльний аналіз фінансового стану фірм. Порівняльний аналіз, на відмінність від рейтингування, припускає введення не однієї, а декількох оціночних координат. Це дозволяє краще використовувати наявну інформацію, точніше позиціонувати фірму серед інших. З іншого боку, для осяжності

результатів порівняльного аналізу, кількість параметрів порівняння повинна бути по можливості мінімальним. У вузькому сенсі «осяжність» вимагає введення не більше двох координат - щоб відносна позиція фірми могла бути представлена точкою на двовимірній карті, а різні фінансові показники могли бути візуалізовані у вигляді двовимірних поверхонь.

З математичної точки зору ця задача зводиться до оптимального стиснення інформації про фінансовий стан фірми, тобто відображенні інформації мінімальним числом параметрів при заданому рівні огрублення або мінімізації втрат інформації при заданому числі узагальнених координат. Для цілей візуалізації, вигідно обмежитися двоухпараметричним уявленням. Це вже істотний крок вперед в порівнянні з однопараметричним рейтингом.

Для ілюстрації описуваного підходу будемо використовувати дані Центрального Банку України про річні баланси і звіти про прибутки (збитки) приблизно 100 українських банків за 2004 - 2005 роки. Кожен банк при цьому описується 30 фінансовими показниками - відношенням балансових статей до загальної суми активів банку. Подібна нормалізація приводить всі статті до єдиного масштабу, згладжуючи відмінності між крупними і дрібними банками. Згідно такому підходу надійність банку будемо характеризувати одним фінансовим показником - відношенням власного капіталу до повернутого. Більш загальний підхід - використовувати не дві окремі компоненти, а дві лінійні комбінації всіх 30 початкових параметрів, що найкращим чином представляють наявні дані (рис. 3.50).

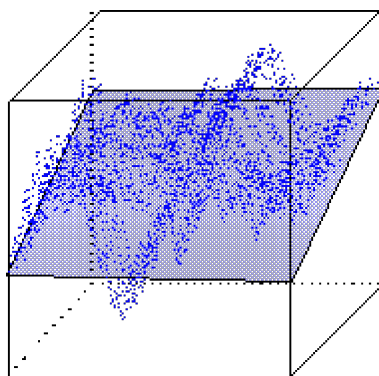


Рис. 3.50. Лінійна апроксимація багатовимірних даних.

Кожен банк представлений точкою в 30-мірному просторі і задача полягає в проведенні двовимірної площини в цьому просторі, що забезпечує мінімальне середньоквадратичне відхилення наявних точок від цієї площини. Подібне лінійне наближення створюється методом головних компонент. Якщо дійсне розташування точок не сильно відхиляється від площини, цей метод може дати непогане початкове наближення. Проте, виявляється, що в даному випадку це не так. Середньоквадратичне відхилення для випадку двох головних компонент виявилось рівним майже половині від загальної дисперсії. Таким чином, навіть оптимальний варіант лінійного стиснення не дає можливості візуалізувати фінансове положення банків.

У цій ситуації природно звернутися до нелінійного статистичного аналізу, тобто до нейромережевого моделювання. Нагадаємо, що методом, що дає оптимальне представлення інформації у вигляді координат двовимірної сітки, є побудова топографічних карт (карт Кохонена). Таким чином, можна з прийнятною точністю описати фінансовий стан українських банків використовуючи всього лише два узагальнені фінансові індикатори, а саме - дві координати на двовимірній карті Кохонена. Кожен банк за станом свого балансового звіту відображається конкретним осередком на карті. Осередки з однаковими координатами містять банки з схожим фінансовим станом. Чим далі на карті координати банків, тим більше відрізняється один від одного їх фінансовий портрет.

Розташування на карті банків з відкликаною ліцензією. Достоїнства карти Кохонена починають виявлятися після нанесення на неї якої-небудь графічної інформації. Малюнок 3.51 показує як виглядає карта Кохонена, на якій відмічені ячейки, що містять банки з відкликаними за наслідками 2004 року ліцензіями. Видно, що банки з відкликаними ліцензіями групуються в правому верхньому кутку карти – «зоні ризику». Ми побачимо, що ця зона має і інші ознаки неблагополуччя.

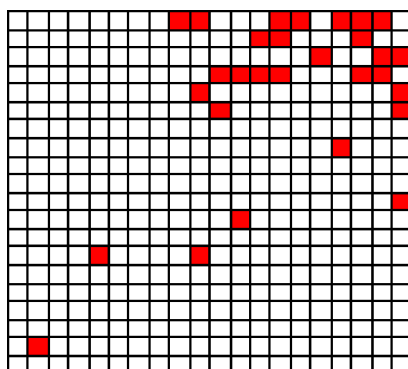


Рис. 3.51. Ячейки, що містять хоч би один банк з відкликаною в 2004 році ліцензією.

Відзначимо, що на відміну від аналізу банкрутств тут інформація про банкрутства не брала участь в навчанні мережі. Вона зображена на вже готовій карті, являється лише індикатором області параметрів з підвищеним ризиком банкрутства. Ця особливість описуваної методики дозволяє виявити область ризику по відносно невеликому числу прикладів.

Карта розмірів банків. Розглянемо, наприклад як розташовуються на побудованій карті банки різних розмірів (рис. 3.52). Розміри банків беруться в логарифмічній шкалі, причому клітки, що відрізняються на одну колірну градацію, містять банки з п'ятикратним відношенням активів. Нагадаємо, що величина активів банків була спочатку виведена з набору параметрів, оскільки вона використовувалася для нормування решти статей. Не дивлячись на це, банки різних розмірів розташовуються не хаотично, а регулярним чином, що свідчить про значущість розміру банку при виборі їм своїй фінансовій стратегії.

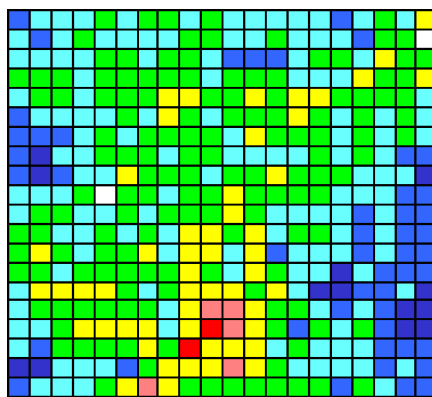


Рис. 3.52. Розмір активів українських банків.

Візуально на карті можна виділити наступні великі групи банків: великі банки (низ карти), малі банки (група зліва і група справа) і середні банки. Розфарбовування, що відображає відносний розмір статутного фонду показує, що між двома групами малих банків є істотні відмінності: банки в нижньому правому куті карти практично не ростуть (рис. 3.53).

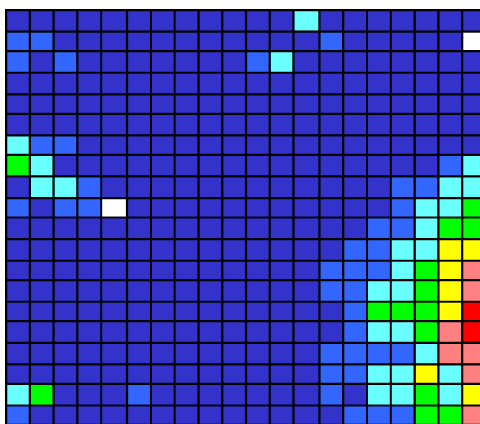


Рис. 3.53. Статутний фонд банків.

Порівняння з розташуванням банків-банкротів, показує, що вірогідність банкрутства як великих так і малих банків в 2004 - 2005 роках була невелика.

Приведені вище розфарбовування в сукупності утворюють атлас, що відображає фінансовий стан банків або інших фірм, що займаються схожими видами бізнесу. Цей атлас дає графічне відображення положення будь-якої конкретної фірми серед конкурентів і може використовуватися як зручний засіб фінансового аналізу. Зокрема, можна розглядати еволюцію фінансового положення окремої фірми в часі, виявляти існуючі тенденції і цикли. З погляду макроекономіки зручність такого роду карт полягає в тому, що площі на цій карті приблизно пропорційні долі фірм через більш-менш рівномірне заповнення ячеек. Таким чином можна зримо уявляти собі, наприклад, частку крупних банків або банків, що зазнають труднощі з поверненням кредитів.

Передбачення фінансових часових рядів. Прогноз фінансових часових рядів - необхідний елемент будь-якої інвестиційної діяльності. Сама ідея інвестицій - вкладення грошей зараз з метою отримання доходу в

майбутньому - ґрунтується на ідеї прогнозування майбутнього. Відповідно, прогноз фінансових часових рядів лежить в основі діяльності всієї індустрії інвестицій - всіх бірж і небіржових систем торгівлі цінними паперами. Приведемо декілька цифр, що ілюструють масштаб цієї індустрії прогнозів. Денний оборот ринку акцій тільки в США перевищує \$40 млрд. Депозитарій DTC (Depository Trust Company) в США, де зареєстровано цінних паперів на суму \$21 трлн, реєструє в день операцій приблизно на \$750 млрд. Ще активніше йде торгівля на світовому валютному ринку FOREX. Його денний оборот перевищує \$1000 млрд. Це приблизно 1/50 всього сукупного капіталу людства. Відомо, що 99% всіх операцій - спекулятивні, тобто направлені не на обслуговування реального товарообігу, а поміщені з метою витягання прибутку по схемі «купив дешевше - продав дорожче». Всі вони засновані на прогнозах зміни курсу учасниками операції. Причому, що важливо, прогнози учасників кожної операції протилежні один одному. Отже об'єм спекулятивних операцій характеризує ступінь відмінностей в прогнозах учасників ринку, тобто реально - ступінь непередбачуваності фінансових часових рядів.

Методика передбачення часових рядів. Спершу змалюємо загальну схему нейромережевого передбачення часових рядів (рис. 3.54). Почнемо з етапу занурення. Як ми зараз переконаємося, не дивлячись на те, що передбачення, здавалося б, є екстраполяцією даних, нейромережі, насправді, вирішують задачу інтерполяції, що істотно підвищує надійність рішення. Передбачення часового ряду зводиться до типової задачі нейроаналіза - апроксимації функції багатьох змінних по заданому набору прикладів - за допомогою процедури занурення ряду в багатовимірний простір. Наприклад, d -мірний лаговий простір ряду X_t складається із d значень ряду в послідовні моменти часу: $X_{t-d} = (X_{t-1}, \dots, X_{t-d})$.

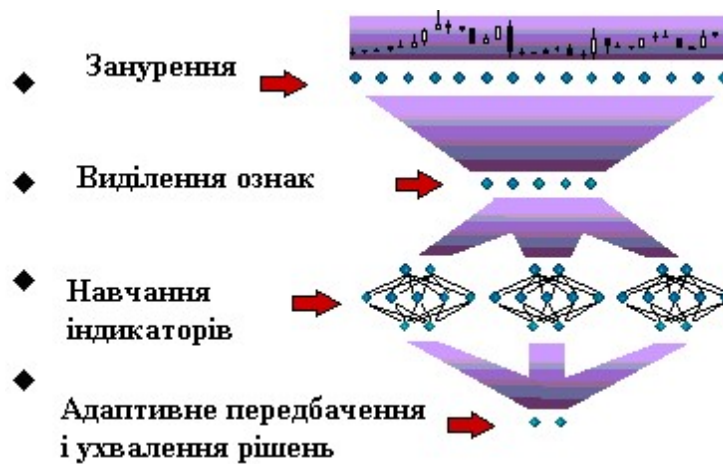


Рис. 3.54. Схема технологічного циклу передбачення ринкових часових рядів.

Для динамічних систем доведена наступна теорема Такенса. Якщо часовий ряд породжується динамічною системою, тобто значення X_t є довільна функція стану такої системи, існує така глибина занурення d (приблизно рівна ефективному числу мір свободи даної динамічної системи), яка забезпечує однозначний прогноз наступного значення часового ряду. Таким чином, вибравши чимале d можна гарантувати однозначну залежність майбутнього значення ряду від його d попередніх значень: $X_t = f(X_{t-d})$, тобто передбачення часового ряду зводиться до завдання інтерполяції функції багатьох змінних. Нейромережу далі можна використовувати для відновлення цієї невідомої функції по набору прикладів, заданих історією даного часового ряду. Навпаки, для випадкового ряду знання минулого нічого не дає для передбачення майбутнього. Тому, згідно теорії ефективного ринку, розкид значень ряду, що передбачається, на наступному кроці при зануренні в лаговий простір не зміниться.

Для навчання нейромережі недостатньо сформулювати навчальні набори входів-виходів. Необхідно також визначити помилку прогнозів мережі. Середньоквадратична помилка, використовувана за умовчанням в більшості нейромережових додатків, не має великого «фінансового сенсу» для ринкових рядів. Тому окремо слід розглянути специфічні для фінансових часових рядів функції помилки і показати їх зв'язок з можливою нормою прибутку.

Наприклад, для вибору ринкової позиції надійне визначення знаку зміни курсу важливіше, ніж пониження середньоквадратичного відхилення. Хоча ці показники і зв'язані між собою, мережі оптимізовані по одному з них даватимуть гірші прогнози іншого.

Основна специфіка передбачення часових рядів лежить в області передобробки даних. Процедура навчання окремих нейромереж стандартна. Як завжди, наявні приклади розбиваються на три вибірки: навчальну, валідаційну і тестову. Перша використовується для навчання, друга - для вибору оптимальної архітектури мережі або для вибору моменту зупинки навчання. Нарешті, третя, яка взагалі не використовувалася в навчанні, служить для контролю якості прогнозу навченої нейросеті. Проте, для сильно зашумлених фінансових рядів істотний вигравш в надійності прогнозів здатне дати використання комітетів мереж. Поліпшення якості прогнозів можливе також за рахунок використання нейромереж із зворотними зв'язками. Такі мережі можуть володіти локальною пам'яттю, що зберігає інформацію про далеке минуле, ніж те, що в явному вигляді присутнє на входах.

В якості вхідних змінних логічно вибирати найбільш статистично незалежні величини, наприклад, зміни котирувань δC_t , або логарифм відносного приросту $\log(C_t / C_{t-1}) \approx \Delta C_t / C_{t-1}$. Останній вибір хороший для тривалих часових рядів, коли вже помітно вплив інфляції. В цьому випадку прості різниці в різних частинах ряду матимуть різну амплітуду, оскільки фактично вимірюються в різних одиницях.

Одним з самих «слабких місць» у фінансових прогнозах є дефіцит прикладів для навчання нейросеті. Фінансові ринки, взагалі кажучи, не стаціонарні. З'являються нові фінансові інструменти, для яких ще не накопичена історія, змінюється характер торгівлі на колишніх ринках. У цих умовах довжина доступних для навчання нейросеті часових рядів вельми обмежена. Проте, можна підвищити число прикладів, використовуючи для цього ті або інші апріорні міркування про інваріанти динаміки часового ряду. Це ще одне фізико-математичне поняття, здатне значно поліпшити якість

фінансових прогнозів. Йдеться про генерацію штучних прикладів, що отримуються з вже отриманих шляхом застосуванням до них різного роду перетворень.

Пояснимо основну думку на прикладі. Психологічно виправдано наступне припущення: гравці звертають увагу, в основному, на форму кривої цін, а не на конкретні значення по осях. Тому якщо небагато розтягнути по осі котирувань весь часовий ряд, то отриманий в результаті такого перетворення ряд також можна використовувати для навчання разом з результатним. Ми, таким чином, подвоїли число прикладів за рахунок використання апріорної інформації, витікаючої з психологічних особливостей сприйняття часових рядів учасниками ринку. Більш того, ми не просто збільшили число прикладів, але і обмежили клас функцій, серед яких шукається рішення, що також підвищує якість передбачення.

Особливістю передбачення фінансових часових рядів є прагнення до отримання максимального прибутку, а не мінімізації середньоквадратичного відхилення, як це прийнято у разі апроксимації функцій. У простому випадку щоденної торгівлі прибуток залежить від вірно вгаданого знаку зміни котирування. Тому нейромережі потрібно орієнтувати саме на точність вгадування знаку, а не самого значення. Знайдемо як зв'язана норма прибутку з точністю визначення знаку в простій постановці щоденного входження в ринок. Позначимо на момент t : повний капітал гравця K_t , відносна зміна котирування $x_t = \Delta C_t / C_{t-1}$, а в якості вихіду мережі візьмемо ступінь її упевненості в знаку цієї зміни $y_t \in [-1, 1]$. Така мережа з вихідною нелінійністю вигляду $y = \tanh(\alpha)$ навчається передбачати знак зміни і видає прогноз знаку з амплітудою пропорційної його вірогідності. Тоді зростання капіталу на кроці t запишеться у вигляді:

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

де δ - частка капіталу «в грі». Виграш за весь час гри:

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

нам і належить максимізувати, вибравши оптимальний розмір ставок σ . Хай в середньому гравець вгадує частку $p = \frac{1}{2} + \varepsilon$ знаків і, відповідно, помиляється з вірогідністю $q = \frac{1}{2} - \varepsilon$. Тоді логарифм норми прибутку

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

а отже і сам прибуток, буде максимальним при значенні $\delta = (p - q)(\frac{|x|}{x^2})$ і складе в середньому:

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

Тут ми ввели коефіцієнт $\alpha = |x|^2 / x^2 \leq 1$. Наприклад, для Гауссова розподілу $\alpha \approx 0.8$. Ступінь передбаченості знаку безпосередньо пов'язаний з кросс-ентропією, яку можна оцінити а ргіогу методом box-counting (рис. 3.55).

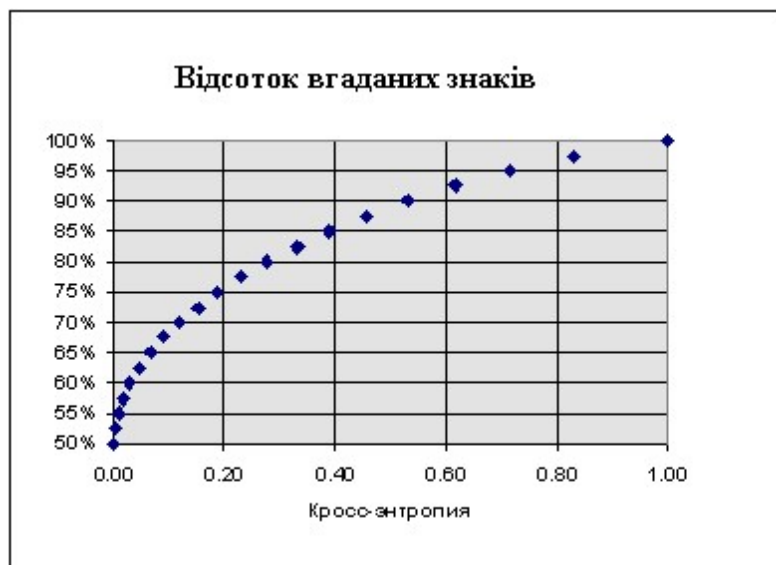


Рис. 3.55. Доля правильно вгаданих напрямків змін ряду.

У результаті отримуємо наступну оцінку норми прибутку при заданій величині передбаченості знаку I , вираженої в бітах $K_t = K_0 2^{\alpha I t}$. Тобто, для ряду з передбаченістю I в принципі можливо подвоїти капітал за $t = 1/\alpha I$ входжень в ринок. Таким чином, навіть невелика передбаченість знаку зміни котирувань здатна забезпечити вельми помітну норму прибутку. Підкреслимо, що

оптимальна норма прибутку вимагає достатньо акуратної гри, коли при кожному входженні в ринок гравець ризикує строго певною часткою капіталу:

$$\Delta K / K = \delta |x| = (p - q)(|x|^2 / x^2) = 2\alpha\varepsilon \approx 1.6\varepsilon,$$

де ΔK - типова при даній волатильності ринку $|x|$ величина виграшу або програшу. Як менші, так і більші значення ставок зменшують прибуток. Причому, занадто ризикована гра може привести до програшу при будь-якій передбачаючій здатності. Цей факт ілюструє рисунок 3.56.

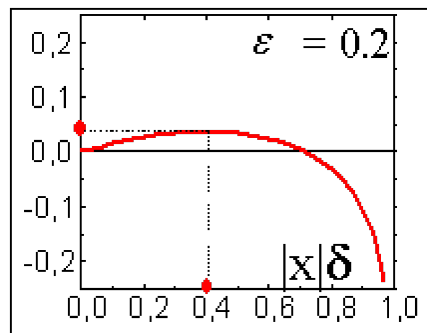


Рис. 3.56. Залежність середньої норми прибутку від вибору частки капіталу.

Використання комітетів мереж. Із-за випадковості у виборі початкових значень синаптичних вагів, прогнози мереж, навчених на одній і тій же вибірці, будуть різнитися. Цей недолік можна перетворити на перевагу, організувавши комітет нейро-експертів, що складається з різних нейромереж. Розкид в прогнозах експертів дасть уявлення про ступінь упевненості цих передбачень, що можна використовувати для правильного вибору стратегії гри. Легко показати, що середнє значень комітету повинно давати кращі прогнози, чим середній експерт з цього ж комітету. Хай помилка i -го експерта для значення входу дорівнює $\varepsilon_i(x)$. Середня помилка комітету завжди менше середньоквадратичної помилки окремих експертів в силу нерівності Коші:

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

Причому, зниження помилки може бути доволі помітним. Так, помилка комітету з L експертів в \sqrt{L} раз менше, ніж середня індивідуальна помилка одного експерта.

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

Тому, в прогнозах завжди краще спиратися на середні значення всього комітету. Ілюстрацією цього факту служить рисунок 3.57. Виграш комітету (кружки) вищий, ніж виграш середнього експерта. Рахунок вгаданих знаків для комітету 59:41.

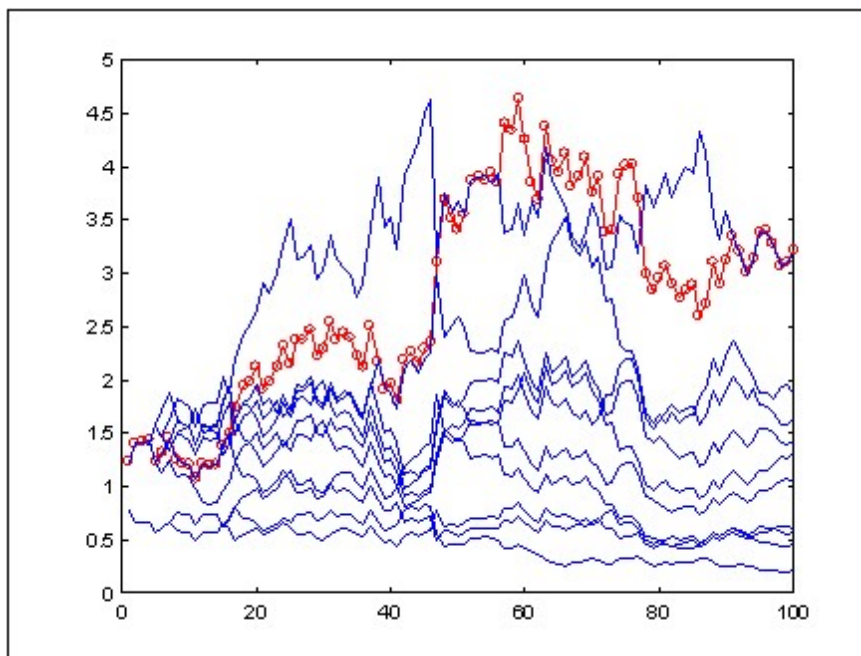


Рис. 3.57. Норма прибутку на останніх 100 значеннях ряду при передбаченні комітетом з 10 мереж.

Як видно з приведенного вище рисунка, в даному випадку виграш комітету навіть вище, ніж виграш кожного з експертів. Таким чином, метод комітетів може істотно підвищити якість прогнозування. Зверніть увагу на абсолютне значення норми прибутку: капітал комітету зріс в 3.25 разу при 100 входженнях в ринок.

Завершуючи огляд нейрокомп'ютерних технологій зупинимося на деяких перспективних напрямках застосування нейронних мереж.

Штучні нейронні мережі і експертні системи. Останніми роками над штучними нейронними мережами домінували логічні і символно-операційні дисципліни. Наприклад, широко пропагувалися експертні системи, у яких є багато помітних успіхів, так само, як і невдач. Дехто говорить, що штучні нейронні мережі замінять собою сучасний штучний інтелект, але багато що свідчить про те, що вони існуватимуть, об'єднуючись в системах, де кожен підхід використовується для вирішення тих завдань, з якими він краще справляється. Ця точка зору підкріплюється тим, як люди функціонують в нашому світі. Розпізнавання образів відповідає за активність, що вимагає швидкої реакції. Оскільки дії здійснюються швидко і несвідомо, то цей спосіб функціонування важливий для виживання у ворожому оточенні. Уявіть тільки, що було б, якби наші предки вимушені були обдумувати свою реакцію на хижака, що стрибнув? Коли наша система розпізнавання образів не в змозі дати адекватну інтерпретацію, питання передається у вищі відділи мозку. Вони можуть запитати додаткову інформацію і займуть більше часу, але якість отриманих в результаті рішень може бути вище. Можна уявити собі штучну систему, що наслідує такому розподілу праці.

Штучна нейронна мережа реагувала б в більшості випадків відповідним чином на зовнішнє середовище. Оскільки такі мережі здатні указувати довірчий рівень кожного рішення, то мережа «знає, що вона не знає» і передає даний випадок для дозволу експертній системі. Рішення, що приймаються на цьому вищому рівні, були б конкретними і логічними, але вони можуть потребувати збору додаткових фактів для отримання остаточного висновку. Комбінація двох систем була б могутнішою, ніж кожна з систем окремо, слідуючи при цьому високоефективній моделі, що дається біологічною еволюцією.

Надійність нейронних мереж. Перш ніж штучні нейронні мережі можна буде використовувати там, де поставлено на карту людське життя або цінне майно, повинні бути вирішені питання, що відносяться до їх надійності. Подібно до людей, структуру мозку яких вони копіюють, штучні нейронні мережі зберігають до певної міри непередбачуваність. Єдиний спосіб точно

знати вихід полягає у випробуванні всіх можливих вхідних сигналів. У великій мережі така повна перевірка практично нездійснена і повинні використовуватися статистичні методи для оцінки функціонування. В деяких випадках це неприпустимо. Наприклад, що є допустимим рівнем помилок для мережі, що управляє системою космічної оборони? Більшість людей скажуть, будь-яка помилка недопустима, оскільки веде до величезного числа жертв і руйнувань. Це відношення не міняється від тієї обставини, що людина в подібній ситуації також може припускатися помилок. Проблема виникає із-за допущення повної безпомилковості комп'ютерів. Оскільки штучні нейронні мережі іноді здійснюватимуть помилки навіть при правильному функціонуванні, то, як відчувається багатьма, це веде до ненадійності - якості, яку ми вважаємо неприпустимою для наших машин.

Схожа трудність полягає в нездатності традиційних штучних нейронних мереж «пояснити», як вони вирішують задачу. Внутрішнє уявлення, що виходить в результаті навчання, часто настільки складно, що його неможливо проаналізувати, за винятком найпростіших випадків. Це нагадує нашу нездатність пояснити, як ми пізнаємося людину, не дивлячись на відмінність у відстані, куті, освітленні і минулих роках. Експертна система може прослідкувати процес своїх міркувань в зворотному порядку, так що людина може перевірити її на розумність. Можливе вбудовування цієї здатності в штучні нейронні мережі може істотно вплинути на прийнятність цих систем.

Логічно прозорі нейронні мережі і виробництво явних знань з даних. Одним з основних недоліків нейронних мереж, з точки зору багатьох користувачів, є те, що нейронна мережа вирішує задачу, але не може розповісти як. Іншими словами з навченої нейронної мережі не можна витягувати алгоритм рішення задачі. Проте спеціальним чином побудована процедура дозволяє вирішити це завдання.

Задасмося класом мереж, які вважатимемо логічно прозорими (тобто такими, які вирішують задачу зрозумілим для нас способом, для якого легко сформулювати словесний опис у вигляді явного алгоритму). Наприклад

зажадаємо, аби всі нейрони мали не більше трьох вхідних сигналів. Задамося нейронною мережею в якій всі вхідні сигнали подаються на всі нейрони вхідного шару, а всі нейрони кожного наступного шару приймають вихідні сигнали всіх нейронів попереднього шару. Навчимо мережу безпомилковому рішенню задачі. Після цього вироблятимемо контрастування у декілька етапів. Процедура контрастування заснована на оцінці значущості вагів зв'язків в мережі. Основною її метою є спрощення технічної реалізацію мережі і створення навичку мережі, який був би зрозумілішим.

На першому етапі контрастуватимемо лише ваги зв'язків нейронів вхідного шару. Якщо після контрастування в деяких нейронів залишилося більше трьох вхідних сигналів, то збільшимо число вхідних нейронів. Потім аналогічну процедуру виробимо по черзі для всіх останніх шарів. Після завершення описаної процедури буде отримана логічно прозора мережа (рис. 3.58).

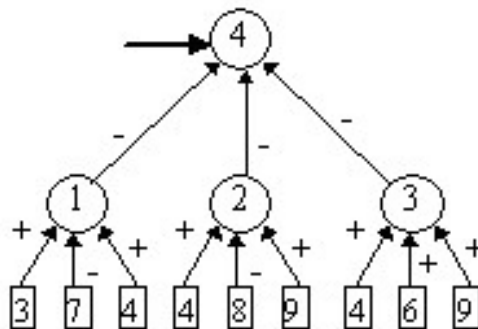


Рис. 3.58. Логічно прозора нейронна мережа.

Як приклад приведемо інтерпретацію алгоритму міркувань, отриманого по логічно прозорій мережі. Постановка задачі слідує: по відповідях на 12 питань необхідно передбачити перемогу правлячої або опозиційної партії. Відповіді на питання описують ситуацію на момент перед виборами. Негативний сигнал на виході мережі інтерпретується як передбачення перемоги правлячій партії. Інакше відповіддю вважається перемога опозиційної партії. Всі нейрони реалізовували порогову функцію, рівну 1, якщо сума алгебри

вхідних сигналів нейрона більше або рівна 0, і -1 при сумі меншою 0. Відповідь мережі базується на проявах двох синдромів: синдрому політичної нестабільності і синдрому поганої політики. Таким чином, для перемоги правлячої партії необхідна відсутність (-1) обох синдромів. На малюнку 3.59 приведена структура логічно прозорої нейронної мережі, яка вирішувала задачу про передбачення результатів виборів президента США.

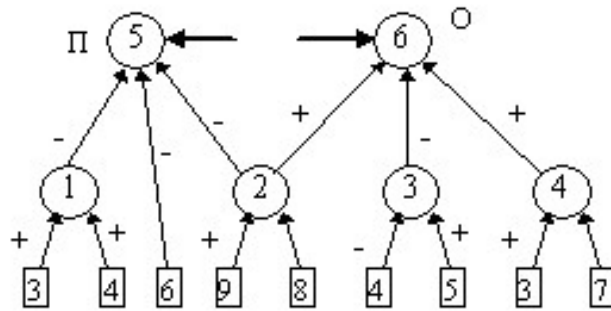


Рис.3.59. Логічно прозора мережа для задачі передбачення.

Технологія здобуття явних знань з даних за допомогою навчених нейронних мереж виглядає досить просто. Перший етап: навчаємо нейронну мережу вирішувати базову задачу. Зазвичай базовою є задача розпізнавання, передбачення і тому подібне. В більшості випадків її можна трактувати як задачу про заповнення пропусків в даних. Другий етап: за допомогою аналізу показників значущості, контрастування і донавчання приводимо нейронну мережу до логічно прозорого вигляду - так, щоб отриманий навик можна було «прочитати».