

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

В. В. КОВАЛЬЧУК

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Конспект лекцій

Одеса
Одеський державний екологічний університет
2015

ББК 73
УДК 681.3
К - 56

Рекомендовано методичною радою Одеського державного екологічного університету Міністерства освіти і науки України як конспект лекцій (протокол № 9 від 25. 06. 2015 р.)

Ковальчук В. В.

Інтелектуальний аналіз даних: конспект лекцій. Одеса, ОДЕКУ, 2015. 206 с.

Пропонований конспект лекцій містить загальну інформацію про сутність, створення та практичне застосування технологій інтелектуального аналізу даних. Основні етапи розробки та застосування таких систем супроводжуються використанням різноманітного програмного забезпечення, яке досить широко і детально представлено в кожному розділі конспекту лекцій. Велика кількість ілюстративного матеріалу, структурованість розділів та підрозділів допомагають швидко оволодіти запропонованими технологіями.

Конспект лекцій призначено для студентів денної та заочної форми навчання, напрямку “Комп’ютерні науки”.

ISBN 978-966-186-002-4

ЗМІСТ

	ПЕРЕДМОВА.....	4
1	ОСНОВНІ ПОНЯТТЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	8
1.1	Сутність аналітичних технологій.....	8
1.2	Поняття інтелектуального аналізу даних.....	11
1.3	Етапи та методи пошуку нових знань.....	22
1.4	Основні моделі інтелектуальних обчислень.....	29
1.5	Засоби програмної підтримки інтелектуального аналізу даних.....	34
1.6	Новітні напрями застосування Data Mining.....	53
2	СХОВИЩА ДАНИХ ТА OLAP-ТЕХНОЛОГІЇ.....	60
2.1	Концепція сховищ даних.....	60
2.2	Технології побудови сховищ даних.....	73
2.3	Вітрини та кіоски даних.....	85
2.4	OLAP-технологія.....	94
2.5	Основні архітектури OLAP-систем.....	103
2.6	OLAP-системи та інтернет-технології.....	117
3	НЕЙРОКОМП'ЮТЕРНІ ТЕХНОЛОГІЇ ТА МЕРЕЖІ.....	127
3.1	Поняття та можливості нейрокомп'ютерних технологій.....	127
3.2	Архітектура нейронних мереж.....	146
3.3	Програмні засоби реалізації нейромережевих технологій.....	173
3.4	Сучасна практика та перспективні напрями застосування нейротехнологій.....	190
	ВИКОРИСТАНА ЛІТЕРАТУРА.....	206

ПЕРЕДМОВА

Сьогодні ми є свідками активного розвитку технологій інтелектуального аналізу даних (Data Mining), виникнення яких пов'язане, в першу чергу, з не обхідністю аналітичної обробки великих обсягів інформації, що накопичуються в сучасних базах даних. Більшість компаній зберігають під час своєї діяльності величезну кількість даних, але головна їх мета одержання корисної інформації. Як дізнатися за допомогою цих даних про те, що є вигіднішим для клієнтів компанії, як розмістити ресурси найефективніше або як мінімізувати втрати? Для вирішення цих проблем і призначені новітні технології інтелектуального аналізу, які використовуються для пошуку моделей і закономірностей, прихованих у середовищі даних, що не можуть бути отримані звичайними методами.

Суть і мету технологій Data Mining можна охарактеризувати так: це технології, призначені для пошуку у великих обсягах даних неочевидних, об'єктивних і корисних на практиці закономірностей. Сфера застосування Data Mining нічим не обмежена — вона поширюється на будь-які дані. Але насамперед методи Data Mining сьогодні заінтригували компанії, що розгортають проекти на основі сучасних інформаційних технологій. Досвід багатьох компаній показує, що віддача від використання Data Mining може досягати 1000 % .

Технології інтелектуального аналізу даних потрібні зокрема спеціалістам, які ухвалюють важливі рішення, керівникам, аналітикам, експертам, консультантам. Дохід компанії більшою мірою визначається якістю цих рішень — точністю прогнозів, оптимальністю вибраних стратегій. І від якості цих рішень залежить розвиток компанії. Зазначимо, що для реальних задач бізнесу і виробництва не існує чітких алгоритмів рішення. Тому керівники і експерти вирішують такі задачі тільки на основі власного досвіду. Часто класичні методика виявляються малоефективними для багатьох практичних завдань, оскільки неможливо точно описати реальність за допомогою невеликої кількості параметрів моделі, або розрахунок моделі займає дуже багато часу і обчислювальних ресурсів. Аналітичні технології дають можливість створювати моделі, що істотно підвищують ефективність рішень.

Пропонований конспект лекцій містить загальну інформацію про сутність, створення та практичне застосування технологій інтелектуального аналізу даних. Основні етапи розробки та застосування таких систем супроводжуються використанням різноманітного програмного забезпечення, яке досить широко і детально представлено в кожному розділі конспекту лекцій. Велика кількість ілюстративного матеріалу, структурованість розділів та підрозділів допомагають швидко

оволодіти запропонованими технологіями.

Автор зробив все можливе, щоб ознайомлення з матеріалом можна було починати з будь-якої частини, тому досвідчений спеціаліст може використовувати конспект лекцій як довідник з системи технологій інтелектуального аналізу даних.

Початківцям радимо опановувати матеріал у запропонованому порядку, що даватиме змогу послідовно аналізувати сутність та основні проблеми, які виникають під час створення та застосування різних технологій інтелектуального аналізу даних.

Кожен розділ конспект лекцій містить тести з відповідної теми, виконання яких допоможе самостійно закріпити опанований теоретичний матеріал. Розділ 8 містить опис виконання лабораторних робіт, що сприятиме оволодінню на практиці різноманітними технологіями інтелектуального аналізу даних для широкого класу задач з реальної діяльності українських підприємств і організацій.

Перший розділ конспекту лекцій присвячений сутності, принципам і особливостям аналітичних технологій, які суттєво відрізняють їх від інших інформаційних технологій. Еволюція таких технологій зумовила появу нового класу систем — систем інтелектуального аналізу даних (Data Mining). У розділі розглядається їх сутність, основні моделі інтелектуальних обчислень, засоби комп'ютерної підтримки, сучасна практична діяльність та перспективні напрями розвитку.

Застосування технологій інтелектуального аналізу даних, як правило, базується на обробці великих обсягів інформації, що накопичуються в сучасних сховищах даних. Існують різні концепції, технології та практичні підходи до побудови таких сховищ, крім того, розроблена досить потужна технологія їх використання — OLAP-технологія, яка має різноманітні архітектури, особливості і значні практичні можливості. Детальне вивчення цих технологій розглядається у другому розділі конспекту лекцій.

У процесі діяльності багатьох компаній часто доводиться вирішувати задачі, постановка яких неформальна, вирішення — неоднозначне. Навіть якщо застосування чіткого алгоритмічного підходу неможливе, і отримати точне рішення принципово не можна, існують інші ефективні способи рішення. Важливе місце серед них займають нейрокомп'ютерні технології та нейронні мережі. У матеріалі третього розділу розглянуто сутність, архітектури, практика побудови та програмні засоби реалізації нейрокомп'ютерних технологій, а також нейронні мережі Хопфілда і Кохонена та сучасна практика і перспективні напрямки їх застосування.

У четвертому розділі подано матеріал, присвячений пошуку асоціативних правил та побудови дерев рішень. Зокрема розглядаються питання теорії асоціативних правил, програмних засобів для їхнього пошуку та практичний аспект застосування цієї технології. Розглянута

технологія дерев рішень, яка є однією з найбільш популярних методів вирішення задач класифікації і прогнозування, а саме — охарактеризовані загальні принципи технології, комп'ютерні системи та напрямки застосування дерев рішень.

Еволюційна теорія довела свою ефективність як під час вирішення складноформалізованих задач кластеризації, асоціативного пошуку, так і трудомістких задач оптимізації, апроксимації, інтелектуальної обробки даних. Концепції еволюційних обчислень включають генетичні алгоритми, генетичне програмування, еволюційні стратегії й еволюційне програмування. Еволюційні технології інтелектуального аналізу сьогодні успішно застосовуються для вирішення низки великих і економічно значущих задач у бізнесі та інших важливих проектах. У п'ятому розділі подані матеріали про концептуальні засади еволюційної теорії, основні положення теорії генетичних алгоритмів та їх моделі, а також інформація про програмне забезпечення та сфери застосування генетичних алгоритмів.

Мурашині алгоритми та генетичне програмування є сучасними напрямками еволюційної технології і майже не розглядалися раніше в підручниках про технології інтелектуального аналізу даних.

Важливим напрямком розвитку інтелектуального аналізу даних є широке застосування теорії нечітких обчислень, яка в сучасному світі розглядається як консорціум обчислювальних методологій, що колективно забезпечують основи для розуміння, конструювання і розвитку інтелектуальних систем, зокрема систем інтелектуального аналізу даних. Тому в шостому розділі конспекту лекцій розглянута концепція нечітких обчислень, а також її складові — теорія нечітких множин та нечітка логіка. Подані матеріали з програмного забезпечення нечітких методів та сучасна практика застосування таких методів.

Сьомий розділ конспекту лекцій присвячено класичним технологіям інтелектуального аналізу, зокрема технологіям класифікації та кластеризації, які є досить поширеним і потужним засобом аналізу даних. У наведеному матеріалі розглянуті як традиційні, так і сучасні методи та алгоритми класифікації, а також сучасні програмні засоби їх підтримки, стан та перспективні напрямки застосування. Задачі кластеризації схожі на задачі класифікації і є їх логічним продовженням. Сучасною наукою розроблено велику кількість методів та алгоритмів цього виду аналізу, створені потужні програмні засоби їх реалізації та методології практичного впровадження, про що також йдеться в цьому розділі.

Восьмий розділ має практичний напрям і зорієнтований на набуття навичок використання теоретичного матеріалу і сучасних комп'ютерних засобів для вирішення аналітичних задач з практики діяльності українських підприємств, фірм і організацій. У цьому розділі запропонований перелік лабораторних робіт, які, на думку авторів, допоможуть отримати перший практичний досвід і поєднати знання з

системи технологій інтелектуального аналізу з їх практичним втіленням.

Основним завданням конспекту лекцій є надання студентам економічних спеціальностей цілісного уявлення про процес інтелектуального аналізу даних, зміст його етапів, технологію залучення інструментальних засобів тощо.

Для полегшення подання та сприйняття інформації у пропонованому виданні прийняті такі позначення: найбільш важливі терміни, визначення та назви методів, алгоритмів, програмних засобів тощо виділені напівжирним шрифтом. Всі пункти меню, назви піктограм меню, назви кнопок управління виділені курсивом та лапками.

У конспекті лекцій можна знайти теоретичне підґрунтя та практичні приклади застосування різноманітних технологій інтелектуального аналізу даних, що дає змогу оптимізувати час впровадження та кошторис витрат.

1 ОСНОВНІ ПОНЯТТЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

1.1 Сутність аналітичних технологій

Аналітичні технології почали застосовуватися людством досить давно. Простим прикладом аналітичної технології є теорема Піфагора, яка дає змогу визначити довжину гіпотенузи, маючи відомі довжини катетів, за відомою формулою $c^2 = a^2 + b^2$. Іншим прикладом аналітичної технології можна назвати алгоритм обробки інформації людським мозком. Навіть мозок дитини може виконувати задачі, непідвладні сучасним комп'ютерам, наприклад розпізнавати знайомі обличчя в юрбі або ефективно керувати декількома десятками м'язів під час гри у футбол. Унікальною особливістю мозку є здібність до розв'язання нових задач — гри в шахи, водіння автомобіля і т. ін. Але при цьому мозок погано пристосований до обробки великих обсягів числової інформації — людина не може перемножити два багатозначні числа, не використовуючи калькулятора або алгоритму обчислення в стовпчик. Реальні задачі з числами набагато складніші, ніж множення, і людині для розв'язання задач необхідні додаткові методики та інструменти.

Під *аналітичною технологією* розуміємо методики, які на основі певних моделей, алгоритмів, математичних теорем дають можливість за відомими даними оцінити значення невідомих характеристик і параметрів.

Аналітичні технології потрібні в першу чергу людям, що ухвалюють важливі рішення, — керівникам, аналітикам, експертам, консультантам. Дохід компанії більшою мірою визначається якістю цих рішень — точністю прогнозів, оптимальністю обраних стратегій. І від якості цих рішень залежить розвиток компанії. За допомогою аналітичних технологій можна вирішувати проблеми прогнозування, наприклад курсів валют, цін на сировину, попиту, доходу компанії, рівня безробіття й оптимізації, наприклад плану закупівель, плану інвестицій, стратегії розвитку. Зазначимо, що для реальних задач бізнесу і виробництва не існує чітких алгоритмів їх розв'язання. Тому керівники і експерти вишукують розв'язки таких задач тільки на основі особистого досвіду. Часто класичні методики виявляються малоефективними для багатьох практичних завдань, оскільки неможливо точно описати реальність за допомогою невеликої кількості параметрів моделі, або розрахунок моделі займає багато часу і обчислювальних ресурсів. Аналітичні технологи допомагають створювати моделі, що істотно підвищують ефективність

рішень.

Серед класичних підходів до аналізу даних на практиці найбільш поширеними виявилися детерміновані та ймовірнісні технології.

Детерміновані аналітичні технології типу теореми Піфагора використовуються людиною багато сторіч. За цей час було створено величезну кількість формул, теорем і алгоритмів для розв'язання класичних задач - визначення об'ємів, знаходження розв'язків систем лінійних рівнянь, пошуку коренів багаточленів. Розроблені складні і ефективні методи аналізу задач оптимального управління, розв'язання диференціальних рівнянь та ін. Всі ці методи діють за однією схемою (рис. 1.1).

*Повний опис задачі---►-Відомий алгоритм -----► Відповідь(модель)
Відомі параметри*

Рисунок 1.1 - Схема детермінованої аналітичної технології

Для застосування алгоритму необхідно, щоб це завдання цілком описувалося певною детермінованою моделлю (деяким набором відомих функцій і параметрів). У такому разі алгоритм дає точну відповідь. Наприклад, для застосування теореми Піфагора потрібно перевірити, що трикутник є прямокутним.

На практиці часто трапляються завдання, пов'язані зі спостереженням випадкових величин, наприклад задача прогнозування курсу акцій. Для подібних проблем не можна будувати детерміновані моделі, тому застосовується принципово інший, ймовірнісний підхід. Параметри ймовірнісних моделей - це розподіли випадкових величин, їх середні значення, дисперсії та ін. Як правило, ці параметри наперед невідомі, а для їх оцінки використовуються статистичні методи, вживані до вибірок зафіксованих значень (історичних даних)

*Історичні----► Статистика----►Параметри-----► Прогноз
дані ймовірнісної моделі*

Рисунок 1.2 - Схема ймовірнісної аналітичної технології

Такі методи передбачають, що відома деяка ймовірнісна модель задачі. Наприклад, розв'язуючи задачу прогнозування курсу можна припустити, що завтрашній курс акцій залежить тільки від курсу за останні 2 дні (авторегресійна модель). Якщо це правильно, то спостереження курсу протягом декількох місяців дають змогу досить точно оцінити коефіцієнти цієї залежності і прогнозувати курс у майбутньому).

На жаль, класичні методики виявляються малоефективними в

багатьох практичних задачах. Це пов'язано з тим, що неможливо достатньо повно описати реальність за допомогою невеликої кількості параметрів моделі, або розрахунок моделі потребує дуже багато часу і обчислювальних ресурсів. Розглянемо проблеми, що виникають при пошуку розв'язку задачі оптимального розподілу інвестицій:

- у реальній задачі жодна з функцій не відома точно, а відомі лише приблизні або очікувані значення прибутку. Для того, щоб позбавитися від невизначеності, ми вимушені зафіксувати функції, втрачаючи при цьому в точності опису задачі;

- детермінований алгоритм для пошуку оптимального розв'язку може бути застосований тільки в тому випадку, якщо всі дані функції лінійні. У реальних задачах бізнесу ця умова не виконується. Хоча дані функції можна апроксимувати лінійними, рішення в цьому випадку буде далеким від оптимального;

- якщо одна з функцій нелінійна, то симплекс-метод непридатний, і залишається два традиційні шляхи пошуку розв'язку цієї задачі. Перший шлях використовувати метод градієнтного спуску для пошуку максимуму прибутку. У випадку, коли область визначення функції прибутку має складну форму, а сама функція — декілька локальних максимумів, градієнтний метод може призвести до неоптимального рішення. Другий шлях — провести повний перебір варіантів інвестування. Якщо кожна з 10 функцій задана в 100 крапках, то доведеться перевірити близько 1020 варіантів, що потребуватиме не менше декількох місяців роботи сучасного комп'ютера.

Ймовірнісні технології також мають істотні недоліки при розв'язанні практичних задач. Вище було проілюстровано дію імовірнісного підходу на прикладі простої лінійної авторегресійної моделі, проте залежності, що трапляються на практиці, часто нелінійні. Навіть якщо й існує проста залежність, то її вигляд наперед невідомий. Якщо ми хочемо враховувати для прогнозування курсу акцій декілька взаємопов'язаних чинників (наприклад кількість операцій, курс долара і т. ін.), то доведеться звернутися до побудови багатовимірної статистичної моделі. Проте такі моделі або припускають гауссівський розподіл спостережень (що не виконується на практиці), або не обґрунтовані теоретично. У багатовимірній статистиці за відсутності кращого нерідко застосовують малообґрунтовані евристичні методи, які за своєю суттю дуже близькі до технології нейронних мереж.

Останніми роками відбувається бурхливий розвиток аналітичних систем нового типу. У їх основі — технології штучного інтелекту, що імітують природні процеси, наприклад діяльність нейронів мозку або процес природного відбору.

При розробці сучасних аналітичних технологій враховується їх здатність:

- до розуміння задачі, загального процесу і знання можливостей інших систем і людей, що беруть участь у взаємодії;
- зв'язку з користувачами за допомогою розуміння природної мови, малюнків, зображень і знаків;
- знань, заснованих на здоровому глузді;
- координування ухвалення рішень, планування і дії;
- навчання на попередньому досвіді і адаптації поведінки.

Розуміння цих здібностей у людях і втілення їх при розробці програм є центральним у створенні новітніх аналітичних технологій, здатних накопичувати і використовувати знання. Від зростання потужностей для проведення інформаційного аналізу, ухвалення рішення, гнучкого проектування і виробництва залежить національна конкурентоспроможність. При додаванні інтелекту до комп'ютерних систем усуваються багато обмежень у розв'язанні реальних задач.

1.2 Поняття інтелектуального аналізу даних

Більшість організацій накопичують під час своєї діяльності величезні масиви даних з метою одержання корисної інформації. Як можна дізнатися за допомогою даних про те, що вигідно для клієнтів організації, як розмістити ресурси ефективно або як мінімізувати втрати? Для вирішення цих проблем

призначені новітні технології інтелектуального аналізу, які використовують для пошуку моделей і відносин, прихованих у середовищі даних, — моделей, які не можуть бути знайдені звичайними методами.

Модель, як і карта, — це абстрактне подання реальності. Карта може вказувати на шлях від аеропорту до будинку, але вона не може показати аварію, яка створила пробку, або ремонтні роботи, які ведуться в даний момент і потребують об'їзду. Доти поки модель не відповідає наявним реальним відносинам, неможливо отримати сприятливий результат. Є два види моделей: прогнозувальні й описові. Перші використовують один набір даних із відомими результатами для побудови моделей, які явно прогнозують результати для інших наборів, а другі описують залежності в наявних даних, які у свою чергу використовуються для ухвалення рішень або дій. Звичайно, компанія, що давно перебуває на ринку і знає своїх клієнтів, користується безліччю моделей. Технології інтелектуального аналізу можуть не тільки підтвердити ці емпіричні спостереження, а й знайти нові, невідомі раніше моделі. Спочатку це може дати користувачеві лише невелику перевагу, але якщо цю інформацію об'єднати за кожним товаром і кожним клієнтом, це допоможе зробити великий крок вперед від тих, хто не використовує таких технологій. З іншого боку, за допомогою методів інтелектуального аналізу можна знайти таку модель, яка приведе до радикального поліпшення у фінансовому і ринковому становищі

компанії.

Термін Data Mining отримав свою назву від двох понять: пошуку цінної інформації у великій базі даних (Data) і видобутку гірської руди (Mining). Обидва процеси потребують або просіювання величезної кількості “сірого” матеріалу, або розумного дослідження і пошуку корисних цінностей. Найчастіше Data Mining перекладається як “видобуток даних”, “витягання інформації”, “розкопка даних”, інтелектуальний аналіз даних, засоби пошуку закономірностей, витягання знань, аналіз шаблонів, “витягання зерен знань із надр даних”, розкопка знань у базах даних, інформаційна проходка даних, “промивання” даних. Поняття “виявлення знань у базах даних” (Knowledge Discovery in Databases, KDD) можна вважати синонімом Data Mining.

Поняття Data Mining вперше з’явилося в 1978 році та стало популярним у сучасному трактуванні приблизно з першої половини 1990-х років. Донині обробка і аналіз даних здійснювалися у рамках прикладної статистики, при цьому в основному вирішувалися завдання обробки невеликих баз даних.

В основу сучасної технології Data Mining покладена концепція шаблонів (партнерів), що відображають фрагменти багатоаспектних відношень у даних. Ці шаблони є закономірностями, властивими підвибіркам даних, які можуть бути компактно виражені в зрозумілій людині формі. Пошук шаблонів проводиться методами, не обмеженими рамками апріорних припущень про структуру вибірки і виду розподілів значень аналізованих показників.

Важливий аспект Data Mining — нетривіальність розшукуваних шаблонів. Це означає, що отримані шаблони повинні відображати неочевидні, несподівані (unexpected) регулярності в даних, такі, що становлять так звані приховані знання (hidden Knowledge). Суспільство зрозуміло, що неякісні дані (raw Data) містять глибинний пласт знань, при технології знизу до верху грамотній розробці якого можуть бути виявлені справжні самородки. У цілому технологію Data Mining досить точно визначає Григорій Піатецький-Шапіро (Gregory Piatetsky-Shapiro) — один із засновників цього напрямку: “*Data Mining*— це процес виявлення в “сірих” даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для ухвалення рішень у різних сферах людської діяльності”.

Суть і мету технології Data Mining можна охарактеризувати так: це технологія, яка призначена для пошуку у великих обсягах даних неочевидних, об’єктивних і корисних на практиці закономірностей. *Неочевидних* — означає, що виявлені закономірності не виявляються стандартними методами обробки інформації або експертним шляхом. *Об’єктивних* — це означає, що виявлені закономірності повністю відповідатимуть дійсності, на відміну від експертної думки, яка завжди є

суб'єктивною. *Практично корисних* — що висновки мають конкретне значення, якому можна знайти практичне застосування.

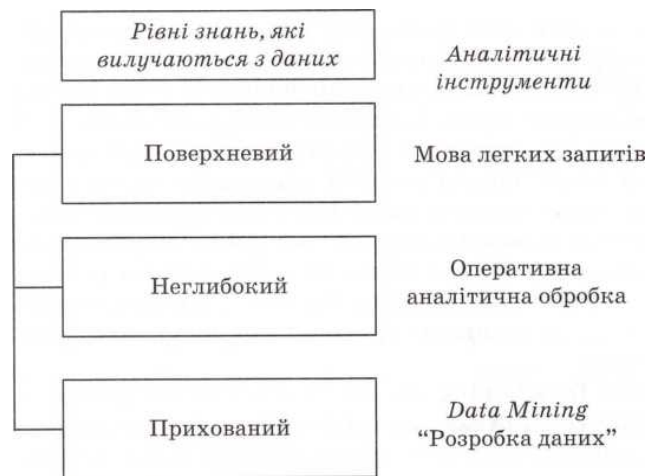


Рисунок 1.1 - Рівні знань, які вилучаються із даних

Наведемо ще декілька визначень поняття DataMining.

Data Mining це процес виділення, дослідження і моделювання великих обсягів даних для виявлення невідомих до цього структур (patterns) з метою досягнення переваг у бізнесі (визначення SASInstitute).

Data Mining — процес, мета якого — виявити нові значущі кореляції, зразки і тенденції в результаті просіювання великого обсягу даних, що зберігаються, з використанням методик розпізнавання зразків плюс застосування статистичних і математичних методів (визначення GartnerGroup).

Data Mining — мультидисциплінарна область, що виникла і розвивалася на базі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних тощо (рис. 1.2).



Рисунок 1.2 - DataMining як мультидисциплінарна область

Розглянемо сутність деяких дисциплін, на межі яких з'явилася технологія DataMining:

- Статистика – це наука про методи збирання даних, їх обробки і аналізу для виявлення закономірностей, властивих явищу, що вивчається. Статистика є сукупністю методів планування експерименту, збирання даних, їх представлення і узагальнення, а також аналізу і одержання висновків на підставі цих даних. Вона оперує даними, отриманими в результаті спостережень або експериментів;

- Машинне навчання можна охарактеризувати як процес одержання програмою нових знань. М. Мітчелл у 1996 році дав таке визначення: "Машинне навчання — це наука, яка вивчає комп'ютерні алгоритми, що автоматично поліпшуються під час роботи". Одним з найбільш популярних прикладів алгоритму машинного навчання є нейронні мережі;

- *штучний інтелект*— науковий напрям, у рамках якого ставляться і вирішуються завдання апаратного або програмного моделювання видів людської діяльності, що традиційно вважаються інтелектуальними. Термін "інтелект" (intelligence) походить від латинського intellectus, що означає розум, розумові здібності людини. Відповідно штучний інтелект (Artificial Intelligence, AI) тлумачиться як властивість автоматичних систем виконувати окремі функції інтелекту людини. Штучним інтелектом називають властивість інтелектуальних систем виконувати творчі функції, які традиційно вважаються прерогативою людини.

Поняття Data Mining також тісно пов'язане з *технологіями баз даних* і поняттям "дані". Дослідження історичного аспекту цієї проблеми дозволяє виявити основні моменти, пов'язані з появою систем інтелектуального аналізу даних. Так, у 1968 році була введена в експлуатацію перша промислова СУБД система IMS фірми "IBM". Вже у 1975 році з'явився перший стандарт асоціації мов систем обробки даних — Conference on Data System Languages (CODASYL), що визначив ряд фундаментальних понять в теорії систем баз даних, які досі є основоположними для мережевої моделі даних. Подальший розвиток теорії баз даних пов'язаний з ім'ям американського математика Е.Ф. Кодда, який є творцем реляційної моделі даних. Протягом 80-х років багато дослідників експериментували з новим підходом у напрямках структуризації баз даних і забезпечення до них доступу. Метою цих пошуків було одержання реляційних прототипів для простішого моделювання даних. У результаті, в 1985 році була створена мова, названа SQL. На сьогодні практично всі СУБД забезпечують даний інтерфейс. Приблизно у цей час з'явилися специфічні типи даних "графічний образ", "документ", "звук", "карта", типи даних для часу, інтервалів часу, символічних рядків із двобайтовим представленням символів були додані в мову SQL. Результати цих всіх зусиль зробили можливим появу технології

Data Mining, сховищ даних, мультимедійних баз даних і web-баз даних.

Для успішного проведення процесу знаходження нового знання необхідною умовою є наявність сховища даних. *Сховище даних* - це предметно орієнтований, інтегрований, прив'язаний до часу, незмінний збір даних для підтримки процесу ухвалення рішень. *Предметна орієнтація* означає, які дані об'єднані в категорії і зберігаються відповідно до областей, що вони описують, а не до застосувань, що їх використовують. *Інтегрованість* означає, що дані задовольняють вимоги всього підприємства, а не однієї функції бізнесу. Цим сховище даних гарантує, що однакові звіти, генеровані для різних аналітиків, міститимуть однакові результати. *Прив'язка до часу* означає, що сховище можна розглядати як сукупність "історичних" даних, тобто можна відновити картину на будь-який момент часу. Атрибут часу завжди явно присутній у структурах сховища даних. *Незмінність* означає, що, потрапивши один раз у сховище, дані там зберігаються і не змінюються, лише додаються. Для організації і експлуатації інформаційного сховища створюється спеціалізоване програмне забезпечення, яке гарантує ефективну взаємодію з користувачем.

Ключовою можливістю застосування новітніх технологій стало величезне падіння ціни за останні декілька років на пристрої зберігання інформації з десятків доларів за зберігання мегабайта інформації, до десятків центів. Це істотно здешевило і збільшило можливості збору та зберігання великих обсягів інформації. Падіння цін на процесори з одночасним збільшенням їх швидкодії сприяло розвитку технологій, пов'язаних з обробкою величезних масивів інформації. У результаті було подолано безліч бар'єрів, що трапляються під час пошуку нового знання в сховищах інформації. Клієнт-серверна архітектура також є необхідним атрибутом технології інтелектуального аналізу даних. Такий підхід надає можливість виконувати найбільш трудомісткі процедури обробки даних на високопродуктивному сервері як розробникам проектів, так і користувачам. На цьому ж сервері можуть зберігатися, і за запитом клієнтів, виконуватися корпоративні проекти.

Технології інтелектуального аналізу даних є важливою частиною ринку сучасних інформаційних технологій. Агентство Gartner Group, що займається аналізом ринків інформаційних технологій, у 1980-х роках ввело термін "Business Intelligence" (BI), діловий інтелект, або бізнес-інтелект. Цей термін запропонований для опису різних концепцій і методів, які покращують бізнес-рішення шляхом використання систем підтримки ухвалення рішень. У 1996 р. агентство уточнило визначення цього терміна: "Business Intelligence — програмні засоби, що функціонують у рамках підприємства і забезпечують функції доступу й аналізу інформації, наявної у сховищі даних, а також сприяють ухваленню правильних і обґрунтованих управлінських рішень". Поняття BI об'єднує

різні засоби і технології аналізу й обробки даних масштабу підприємства. На основі цих засобів створюються ВІ-системи, мета яких — підвищення якості інформації для ухвалення управлінських рішень. ВІ-системи також відомі під назвою Систем Підтримки Прийняття Рішень (СППР, DSS, Decision Support System). Ці системи перетворюють дані в інформацію, на основі якої можна ухвалювати рішення. Gartner Group визначає склад ринку систем Business Intelligence як набір програмних продуктів таких класів:

- засоби побудови сховищ даних (Data Warehousing, ХД);
- системи оперативної аналітичної обробки (OLAP);
- інформаційно-аналітичні системи (Enterprise Information Systems, EIS);
- засоби інтелектуального аналізу даних (Data Mining);
- інструменти для виконання запитів і побудови звітів (Query and Reporting tools).

Класифікація Gartner базується на методі функціональних задач, де програмні продукти кожного класу виконують певний набір функцій або операцій з використанням спеціальних технологій.

Наведемо декілька коротких цитат найбільш впливових членів бізнес-організацій, які є експертами в цій відносно новій технології. Керівництво з придбання продуктів Data Mining (Enterprise Data Mining Buying Guide) компанії Aberdeen Group: “Data Mining — технологія одержання корисної інформації з баз даних. Проте у зв’язку з істотними відмінностями між інструментами, досвідом і фінансовим станом постачальників продуктів, підприємствам необхідно ретельно оцінювати передбачуваних розробників Data Mining і партнерів. Щоб максимально використовувати потужність інструментів Data Mining комерційного рівня, підприємству необхідно вибрати, очистити і перетворити дані, іноді інтегрувати інформацію, отриману із зовнішніх джерел, і встановити спеціальне середовище для роботи Data Mining-алгоритмів. Результати Data Mining значною мірою залежать від рівня підготовки даних, а не від “чудових можливостей” якогось алгоритму або набору алгоритмів. Близько 75 % роботи над Data Mining полягає в збиранні даних, яке здійснюється ще до того, як запускаються самі інструменти. Безграмотно застосувавши деякі інструменти, підприємство може безглуздо розпорошити свій потенціал, а іноді і мільйони доларів”.

Як зазначає Херб Едельштайн, відомий у світі експерт у галузі Data Mining, сховищ даних і CRM “Нещодавнє дослідження компанії Two Crows показало, що Data Mining перебуває ще на ранній стадії розвитку. Багато організацій цікавляться цією технологією, але лише деякі активно упроваджують такі проекти. Вдалося з’ясувати ще один важливий момент: процес реалізації Data Mining на практиці виявляється складнішим, ніж очікується. ІТ-команди захопилися міфом про те, що засоби Data Mining

прості у використанні. Вважається, що досить запустити такий інструмент на терабайтній базі даних, і вмість з'явиться корисна інформація. Насправді успішний Data Mining- проект потребує розуміння суті діяльності, знання даних та інструментів, а також процесу аналізу даних”.

Перш ніж використовувати технологію Data Mining, необхідно ретельно проаналізувати її проблеми, обмеження і критичні питання, що з нею пов'язані, а також зрозуміти, чого ця технологія не може. Зокрема, Data Mining не може замінити аналітика, а тільки дає йому потужний інструмент для полегшення і поліпшення роботи. Також технологія не може дати відповіді на ті питання, які не були поставлені. Оскільки ця технологія є мультидисциплінарною галуззю, для розробки додатка, що включає Data Mining, необхідно залучати фахівців із різних галузей, а також забезпечувати їх якісну взаємодію.

Різні інструменти Data Mining мають різний ступінь “доброзичливості” інтерфейсу і потребують певної кваліфікації користувача. Тому програмне забезпечення повинне відповідати рівню підготовки користувача. Використання Data Mining повинне бути нерозривно пов'язане з підвищенням кваліфікації користувача, проте фахівців із Data Mining, які б добре розбиралися в бізнесі, поки що мало.

Вдалиий аналіз потребує якісної переробки даних. За твердженням аналітиків і користувачів баз даних, процес переробки може зайняти до **80** % з усього Data Mining-процесу. Тому, щоб технологія працювала на себе, потрібно буде багато зусиль і часу, які підуть на попередній аналіз даних, вибір моделі та її коригування.

За допомогою Data Mining можна відшукувати дійсно цінну інформацію, яка незабаром дасть великі дивіденди у вигляді фінансової і конкурентної вигоди. Проте Data Mining достатньо часто робить безліч помилкових відкриттів, які не мають сенсу. Багато фахівців стверджують, що засоби Data Mining можуть видавати величезну кількість статистично недостовірних результатів. Щоб цього уникнути, необхідна перевірка адекватності отриманих моделей на тестових даних.

Зазначимо, що якісна Data Mining-програма може коштувати достатньо дорого для компанії. Варіантом є придбання вже готового рішення з попередньою перевіркою його використання, наприклад на демо-версії з невеликою вибіркою даних. Засоби Data Mining, на відміну від статистичних, теоретично не потребують наявності строго певної кількості ретроспективних даних. Ця особливість може стати причиною виявлення недостовірних, помилкових моделей і, як результат, ухвалення на їх основі неправильних рішень. Необхідно здійснювати контроль статистичної значущості виявлених знань.

Data Mining має суттєві відмінності від інших методів аналізу даних.

- традиційні методи аналізу даних (статистичні методи) і OLAP в основному орієнтовані на перевірку наперед сформульованих гіпотез

(verification-driven Data Mining) і на “грубий” розвідувальний аналіз, що становить основу оперативної аналітичної обробки даних (Online Analytical Processing, OLAP), тоді як одне з основних положень Data Mining — пошук неочевидних закономірностей. Інструменти Data Mining можуть виявляти такі закономірності самостійно і самостійно будувати гіпотези про взаємозв’язки. Оскільки саме формулювання гіпотези щодо залежностей є найскладнішим завданням, перевага Data Mining порівняно з іншими методами аналізу є очевидною. Більшість статистичних методів для виявлення взаємозв’язків у даних використовують концепцію усереднювання за вибіркою, що приводить до операцій над неіснуючими величинами, тоді як Data Mining оперує реальними значеннями. OLAP — лише придатна для розуміння ретроспективних даних, Data Mining спирається на ретроспективні дані для одержання відповідей на питання про майбутнє.

Потенціал Data Mining дає “зелене світло” для розширення меж застосування технології. Щодо перспектив Data Mining можливі такі напрями розвитку:

- виділення типів предметних областей із відповідними евристичними, формалізація яких полегшить розв’язання відповідних задач Data Mining, що належать до цих областей;
- створення формальних мов і логічних засобів, за допомогою яких будуть формалізовані міркування і автоматизація яких стане інструментом розв’язання задач Data Mining у конкретних наочних областях;
- створення методів Data Mining, здатних не тільки витягувати з даних закономірності, а і формувати певні теорії, що спираються на емпіричні дані;
- подолання істотного відставання можливостей інструментальних засобів Data Mining від теоретичних досягнень у цій галузі.

Якщо розглядати майбутнє Data Mining у короткостроковій перспективі, то очевидно, що розвиток цієї технології найбільш спрямований на галузі, пов’язані з бізнесом. Продукти Data Mining можуть стати такими ж звичайними і необхідними, як електронна пошта, і, наприклад, використовуватися користувачами для пошуку найнижчих цін на певний товар або найбільш дешевих квитків.

У довгостроковій перспективі майбутнє Data Mining є дійсно захоплюючим — це може бути пошук інтелектуальними агентами як нового способу лікування різних захворювань, так і нового розуміння природи всесвіту.

Проте Data Mining містить і потенційну небезпеку — адже все більша кількість інформації стає доступною через Всесвітню мережу, у

тому числі і відомості приватного характеру, і все більше знань можна здобути з неї. Не так давно найбільший онлайн-магазин “Amazon” опинився в центрі скандалу з приводу отриманого ним патенту “Методи і системи допомоги користувачам при купівлі товарів”, який є ні чим іншим, як черговим продуктом Data Mining, призначеним для збирання персональних даних про відвідувачів магазину. Нова методика дає змогу прогнозувати майбутні запити на підставі фактів покупок, а також робити висновки про їх призначення. Мета цієї методики - одержання якомога більшої кількості інформації про клієнтів, у тому числі і приватного характеру (стан, вік, переваги тощо). Таким чином, збираються дані про приватне життя покупців магазину, а також членів їх сімей, включаючи дітей. Останнє заборонено законодавством багатьох країн — збирання інформації про неповнолітніх можливе тільки з дозволу батьків.

Дослідження довели, що є як успішні рішення, які використовують Data Mining, так і невдалий досвід застосування цієї технології. Напрями, де застосування технологій Data Mining, швидше за все, будуть успішними, мають такі особливості:

- потребують рішень, що ґрунтуються на знаннях;
- мають навколишнє середовище, що змінюється;
- мають доступні, достатні і значущі дані;
- забезпечують високі дивіденди від правильних рішень.

Сфера застосування Data Mining нічим не обмежена — вона скрізь, де є будь-які дані. Але в першу чергу методи Data Mining зацікавляли комерційні підприємства, що розгортають проекти на основі інформаційних сховищ даних (Data Warehousing). Досвід багатьох таких підприємств показує, що віддача від використання Data Mining може досягати 1000 % . Наприклад, відомо про економічний ефект, що в 10—70 разів перевищив первинні витрати від 350 до 750 тис. дол.; про проект у 20 млн дол., який окупився всього за 4 місяці. Інший приклад — річна економія 700 тис. дол. за рахунок впровадження Data Mining у мережі універсамів у Великобританії. Data Mining становить велику цінність для керівників і аналітиків у їх повсякденній діяльності. Ділові люди усвідомили, що за допомогою методів Data Mining вони можуть отримати відчутні переваги в конкурентній боротьбі.

Розглянемо деякі бізнес-приклади Data Mining.

Роздрібна торгівля. Підприємства роздрібно́ї торгівлі сьогодні збирають докладну інформацію про кожну окрему покупку, використовуючи кредитні картки з маркою магазину і комп'ютеризовані системи контролю. Ось типові задачі, які можна розв'язувати за допомогою Data Mining у сфері роздрібно́ї торгівлі:

- аналіз купівельної корзини (аналіз схожості), призначений для виявлення товарів, яких покупці прагнуть придбати. Знання купівельної корзини необхідне для поліпшення реклами, вироблення стратегії

створення запасів товарів і способів їх розкладки в торгових залах;

- дослідження часових шаблонів допомагає торговим підприємствам приймати рішення про створення товарних запасів. Воно дає відповіді на питання типу “Якщо сьогодні покупець придбав відеокамеру, то через який час він найімовірніше купить нові батареї і плівку?”;

- створення прогнозуючих моделей дає можливість торговим підприємствам дізнаватися про характер потреб різних категорій клієнтів із певною поведінкою, наприклад таких, що купують товари відомих дизайнерів або таких, що відвідують розпродаж. Ці знання потрібні для розробки точно спрямованих, економічних заходів щодо просування товарів.

Банківська справа. Досягнення технології Data Mining використовуються в банківській справі для вирішення таких поширених завдань:

- виявлення шахрайства з кредитними картками. Шляхом аналізу минулих транзакцій, які згодом виявилися шахрайськими, банк виявляє деякі стереотипи такого шахрайства;

- сегментація клієнтів. Розбиваючи клієнтів на різні категорії, банки роблять свою маркетингову політику більш цілеспрямованою і результативною, пропонуючи різні види послуг різним групам клієнтів;

- прогнозування змін клієнтури. Data Mining допомагає банкам будувати прогнозні моделі цінності своїх клієнтів, і відповідно обслуговувати кожну категорію.

Телекомунікації. В галузі телекомунікацій методи Data Mining допомагають компаніям енергійніше просувати свої програми маркетингу і ціноутворення, щоб утримувати наявних клієнтів і залучати нових. Серед типових заходів відзначимо такі:

- аналіз записів про докладні характеристики викликів. Призначення такого аналізу — виявлення категорій клієнтів зі схожими стереотипами користування їх послугами і розробка привабливих наборів цін і послуг;

- виявлення лояльності клієнтів. Data Mining можна використовувати для визначення характеристик клієнтів, які, один раз скориставшись послугами цієї компанії, з великою часткою вірогідності залишаться їй вірними. У результаті засоби, що виділяються на маркетинг, можна витратити там, де віддача найбільша.

Страховання. Страхові компанії протягом кількох років накопичують великі обсяги даних — це значне поле діяльності для методів Data Mining:

- виявлення шахрайства. Страхові компанії можуть знизити рівень шахрайства, відшукуючи певні стереотипи у заявах про виплату страхового відшкодування, що характеризують взаємини між юристами,

лікарями і заявниками;

- **аналіз ризику.** Шляхом виявлення поєднань чинників, пов'язаних зі сплаченими заявами, страховики можуть зменшити свої втрати за зобов'язаннями. Відомий випадок, коли в США потужна страхова компанія виявила, що суми, сплачені за заявами людей, які перебувають у шлюбі, вдвічі перевищують суми за заявами самотніх людей. Компанія відреагувала на це нове знання переглядом своєї загальної політики надання знижок сімейним клієнтам.

Управління виробництвом, менеджмент якості. Шляхом аналізу даних автоматизованого виробництва і відхилень від нього можна ідентифікувати проблеми на етапах виробництва як з погляду якості, так і з погляду збереження темпу виробництва. На підставі такої встановленої інформації можна, наприклад, у виробничий процес ввести етап додаткового контролю, завдяки якому вже в процесі виробництва будуть виявлені розроблені вироби, які після закінчення виробничого процесу не пройдуть вихідний контроль.

Молекулярна генетика і гена інженерія. Мабуть, найгостріше завдання виявлення закономірностей в експериментальних даних постало в молекулярній генетиці і генній інженерії. Тут вона формулюється як визначення так званих маркерів, під якими розуміють генетичні коди, контролюючі ті або інші фенотипічні ознаки живого організму. Такі коди можуть містити сотні, тисячі і більше зв'язаних елементів. На розвиток генетичних досліджень виділяються великі кошти. Останнім часом у цій галузі виник особливий інтерес до застосування методів Data Mining. Відомо декілька великих фірм, що спеціалізуються на застосуванні цих методів для розшифрування геному людини і рослин.

Медицина. Відомо багато експертних систем для встановлення медичних діагнозів. Вони побудовані головним чином на основі правил, що описують поєднання різних симптомів різних захворювань. За допомогою таких правил дізнаються не тільки, на що хворий пацієнт, а й як потрібно його лікувати. Правила допомагають вибрати засоби медикаментозної дії, визначати показання/протипоказання, орієнтуватися в лікувальних процедурах, створювати умови найбільш ефективного лікування, передбачати результати призначеного курсу лікування тощо. Технології DataMining дають можливість виявляти в медичних даних шаблони, які становлять основу зазначених правил.

Прикладна хімія. Методи DataMining отримують широкого застосування в прикладній хімії (органічній і неорганічній). Тут нерідко постає питання про з'ясування особливостей хімічної будови тих або інших сполук, що визначають їх властивості. Особливо актуальне таке завдання при аналізі складних хімічних сполук, опис яких включає сотні і тисячі структурних елементів і їх зв'язків.

Можна навести ще багато прикладів різних галузей знань, де методи

DataMining відіграють провідну роль. Особливість цих галузей полягає в їх складній системній організації. Вони належать до суперскладного рівня організації систем, закономірності якого не можуть бути достатньо точно описані мовою статистичних або інших аналітичних математичних моделей. Дані у цих галузях неоднорідні, гетерогенні, нестационарні і часто відрізняються високою розмірністю.

1.3 Етапи та методи пошуку нових знань

Важливо розуміти, що побудова моделі інтелектуального аналізу даних є складовою масштабнішого процесу, який включає всі етапи, починаючи з визначення базової проблеми, яку модель вирішуватиме, до розгортання моделі в робочому середовищі. Цей процес може задаватись за допомогою таких шести базових кроків (рис. 1.3).

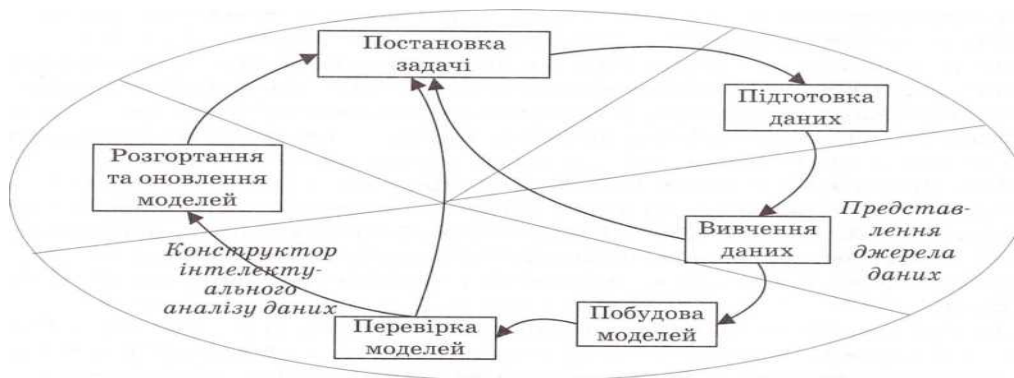


Рисунок - 1.3 Етапи інтелектуального аналізу даних

1. Постановка задачі.
2. Підготовка та огляд даних:
 - оцінювання;
 - об'єднання і очищення;
 - відбір;
 - перетворення.
3. Побудова моделей:
 - оцінка та інтерпретація;
 - зовнішня перевірка.
4. Використання моделей.
5. Нагляд за моделлю.

Хоча він має циклічний характер, кожен крок не обов'язково веде безпосередньо до наступного. Створення моделі інтелектуального аналізу даних є динамічним ітеративним процесом. Виконавши огляд даних, користувач може виявити, що наявних даних недостатньо для створення необхідних моделей інтелектуального аналізу даних, що, відповідно, веде

до необхідності пошуку додаткових даних. Можна розробити декілька моделей і зрозуміти, що вони не розв'язують сформульованої задачі. Отже, потрібна зміна її характеристик.

Може виникнути потреба в оновленні вже розгорнутих моделей за рахунок нових даних, що надійшли. Таким чином, важливо зрозуміти, що створення моделі інтелектуального аналізу даних є процесом і що кожен крок такого процесу може повторюватись стільки разів, скільки необхідно для створення ефективної моделі. Розглянемо докладніше кожен з цих етапів.

Визначення проблеми. Для того, щоб повніше використовувати всі переваги інтелектуальних технологій, необхідно чітко уявити мету майбутнього аналізу. Модель будують залежно від мети. Якщо необхідно збільшити прибуток торгової організації, то для цілей “збільшення кількості продажу” і “збільшення ефективності реклами” необхідно будувати різні моделі. На цьому ж етапі визначаються способи оцінювання результатів майбутнього проекту і можливі витрати на його реалізацію.

Підготовка та огляд даних. Це найтриваліший етап, який може займати від 50 до 85 % часу всього процесу пошуку нового знання, бо необхідно визначити джерела одержання даних. Це можуть бути дані, накопичені самою організацією або зовнішні дані від загальнодоступних джерел (відомості про погоду або перепис населення) або приватних джерел (різні архівні дані, бази нотаріальних контор тощо).

Оцінювання даних. При побудові моделі необхідно пам'ятати одне правило, що стосується коректності вхідних даних: “Якщо на вхід задачі надходить “сміття”, то і результатом теж буде “сміття”. Вхідні дані можуть перебувати в одній базі або в декількох. Перед “завантаженням” даних у сховище необхідно врахувати, що різні джерела даних можуть бути спроектовані під певні задачі і, відповідно, виникають проблеми, пов'язані з об'єднанням даних: різні формати подання числових даних (наприклад, цілі або дійсні); різне кодування (наприклад, різний формат дат); різні способи зберігання даних; різні одиниці вимірювання (дюйми і сантиметри); а також частота збирання даних і дата останнього оновлення. Навіть якщо дані перебувають в одній базі, то однак треба звертати увагу на пропущені значення і значення нереальної величини, так звані “викиди”. Аналітик повинен завжди знати, як, де і за яких умов збираються дані, і бути впевненим, що всі дані, які використовуються для проведення аналізу, визначені однаковою мірою.

Об'єднання і очищення даних. На цьому етапі відбувається побудова сховища даних для подальшої обробки, тобто наповнення сховища або додавання до нього даних, відібраних на попередніх етапах. У цей же час виконується очищення, тобто виправлення всіх виявлених помилок. Існують різні аспекти очищення даних. Всі вони спрямовані на пошук і

виправлення помилок, допущених на етапі збирання інформації. Помилкою в даних можуть вважатися: пропущене значення, неможлива подія (неправильно набране значення — “викид”). Корекція відбувається на основі здорового глузду, використання правил або із залученням експерта, добре обізнаного з предметною галуззю. Запис у базі даних, в якому є помилка, повинен бути виправлений або, в спірних випадках, виключений із подальшого розгляду. Після перевірки дані перетворюються і форматуються відповідно до результатів оцінювання. Це робиться для більшої зручності спостереження за ними. Дані дискретних подій перетворюються на спеціально розроблену або стандартну форму, в якій відбивається час і опис подій. Якщо користувачі легко розбиратимуться в цій формі, вони зможуть швидко вивчити події, які були в основі побудови цієї форми. Може здатися, що цей крок дублює етап збирання даних, але насправді це два зовсім різних етапи. На першому з них відбувається добір даних для прискорення машинної обробки інформації без втрати якості, на другому — дані доводяться до вигляду, зручному для візуального контролю користувача. Людина, яка проводить аналіз, може повніше уявити собі вхідні дані. Це необхідно для різноманітних звітів, коли потрібно коротко охарактеризувати вхідні дані, вживані для аналізу.

Відбір даних. Якщо сховище сформоване і визначені типи моделей, які будуть побудовані для роз’язку задач, відбувається добір даних, необхідних саме для цих моделей. Мається на увазі не тільки зменшення кількості записів у базі за певною умовою, а й зміна кількості полів, злиття різних таблиць в одну, або, навпаки, створення на основі однієї таблиці декількох. Тобто перетворення відбувається у “трьох вимірах”: за кількістю записів та полів і структурою.

Перетворення даних. Слугує для збагачення отриманої бази, тобто додавання різних зв’язків на основі наявних полів (не просто “ціна” і “кількість”, а їх утворення — “загальна сума”, не борг і дохід, а відношення боргу до доходу), додавання інтервалів (за номером місяця можна поставити номер кварталу, а відсоток виконання плану можна доповнити характеристиками “добре”, “задовільно”), додавання критичних значень (максимум, середнє, мінімум).

Побудова моделі є ітераційним процесом, тобто необхідно побудувати низку моделей для пошуку однієї, що найбільше відповідає поставленим цілям.

Моделі можна розділити на дві групи:

- контрольовані (моделі класифікації, регресії, прогнозування часових послідовностей);
- Після того, як визначений тип моделі, необхідно обрати алгоритм побудови моделі або технологію пошуку знань.
- Суть процесу побудови контрольованої моделі зводиться до пошуку залежностей на одній (“навчання моделі”) і перевірки цих

залежностей на іншій частині даних (оцінка точності). Модель вважається побудованою, якщо завершується цикл “навчання” і перевірок. Якщо точність моделі при чергових ітераціях не поліпшується, то це свідчить про завершення побудови моделі. Оскільки “навчальні” і тестові дані перебувають в одній базі даних, то часто виникає необхідність у третьому наборі даних — контрольному, який обирається з таких даних, що не перетинаються з “навчальними” і тестовими. Він потрібний для незалежного оцінювання точності моделі. Як правило, всі три набори даних належать множині даних, необхідній для реалізації певного проекту.

- Найбільш відомий тестовий метод — проста оцінка, за якого розподіл даних на два набори відбувається довільно. Відношення кількості тестових даних до кількості даних, на основі яких відбувається побудова моделі, повинне бути в межах від 5 до 33 % . Після побудови моделі її використовують для передбачення значень на тестовому наборі. Мірою точності моделі вважають відношення кількості вдалих результатів до загальної кількості прикладів у тестовому наборі.

- Якщо для побудови моделі використовується не дуже велика база даних, то застосовується так звана перехресна оцінка точності. В цьому випадку дані довільно поділяються на дві приблизно рівні частини. Після цього модель будуватиметься на одній з них, а інша використовується для визначення точності. Потім частини бази міняються ролями. Отримані дві незалежні оцінки точності об’єднуються (як середнє арифметичне або іншим способом) для якнайкращої оцінки точності моделі, побудованої на всій базі.

- Для ще менших баз (у декілька тисяч записів) використовується перехресна оцінка точності. У цьому випадку база ділиться на n приблизно рівних непересічних груп. Далі перша з цих груп стає тестовим набором, а інші об’єднуються, і на їх основі відбувається побудова моделі. Отримана модель використовується для передбачення значень для тестового набору і таким чином виходить перше значення точності. Аналогічно набувають всі n незалежних значень точності. Середнє з них є точністю всієї моделі.

- Ще один спосіб використовується для визначення точності в маленьких базах даних. У цьому випадку модель будуватиметься на основі даних всієї бази. Після цього довільно з записів бази створюється безліч тестових наборів (мінімум 200, а іноді навіть більше 1000). Один запис може бути наявним у різних тестових наборах. Для будь-якого з них визначається точність. Середнє з них є точністю всієї моделі.

Після того, як побудова моделі завершена, можна її коригувати, використовуючи інші параметри або навіть змінити алгоритм її побудови, оскільки ніколи не можна сказати, який алгоритм та технологія пошуку знань дасть кращі результати. Не можна бути впевненим, що якась технологія працюватиме найкраще. Часто доводиться будувати велику

кількість моделей і оцінювати кожну з метою визначення кращої. Окрім цього, для різних моделей необхідна різна підготовка даних і неминуче повторення кроків. Все це збільшує час пошуку кращої моделі, тому необхідно застосовувати технології паралельних обчислень.

Оцінка та інтерпретація. Після побудови моделі необхідно оцінити результати і пояснити (інтерпретувати) їх значущість. При оцінці моделі обчислюється точність, але треба пам'ятати, що це значення правильне лише для даних, на яких модель побудована і бути готовим, що нові дані, до яких надалі застосовуватиметься модель, можуть відрізнитися від результатних.

Зовнішня перевірка. Висока точність моделі не є гарантією того, що модель правильно відображає реальне середовище. Однією з причин є наявність так званих неявних припущень у моделі. Тобто сам по собі коефіцієнт інфляції не може бути частиною моделі, що пояснює схильність покупців до покупки чи того іншого товару, але різка зміна цього коефіцієнта з 3 до 20 % вже, напевно, може пояснити таку поведінку. Інша причина — це існування неминучих проблем із даними, що призводять до некоректності моделі, тому дуже важливо перевірити її у реальному середовищі. Наприклад, якщо модель використовується для відбору кандидатів для цільової реклами, то можна зробити тестову розсилку для її перевірки на невеликому обсязі даних. Якщо вона використовується для передбачення ризику неповернення кредиту, то слід випробувати цю модель на невеликій кількості претендентів на позику. Чим більше ризик, пов'язаний з некоректністю моделі, тим більш важливо провести попередні експерименти для перевірки моделі перед її експлуатацією.

Використання моделі. Після побудови і оцінки моделі, її можна використовувати різними способами. Наприклад, аналітик може подивитися групи, які визначила модель кластеризації, графіки її ефективності або отримані правила. Іноді аналітик може її використовувати для вибору деяких записів із бази даних для проведення додаткового аналізу. Базуючись на результатах використання моделі, аналітик може рекомендувати дії, які можна починати в діловій сфері. Проте часто технології інтелектуальних обчислень — це частина автоматизованої системи (наприклад, пошук кредитних ризиків, визначення можливості втрати клієнтів та ін.), тобто модель вбудовується в систему, яку аналітик або менеджер можуть застосовувати для ухвалення рішення. З іншого боку, її можна включати в систему, що генерує деяку дію (наказ), коли прогнозована величина починає виходити за межі якихось значень. Загалом методи інтелектуальних обчислень — це невелика, хоч і важлива частина кінцевого програмного продукту. Процедура пошуку знання за допомогою цих методів може об'єднуватися з знаннями експертів і застосовуватися до даних у базах.

Спостереження за моделлю. Коли модель починає працювати в реальному середовищі, то необхідно вимірювати її точність на реальних даних. Проте, навіть якщо модель працює добре, і можна вважати, що робота на цьому закінчується, все одно необхідно продовжувати спостереження за нею. Всі системи мають властивість розвиватися, і отримані дані (їх структура, точність, періодичність) теж змінюються. Зовнішня змінна, така як коефіцієнт інфляції, своєю зміною теж може впливати на поведінку людей і на чинники, що впливають на цю зміну. Час від часу модель необхідно піддавати повторному тестуванню і навіть перебудові.

Простим способом спостереження діяльності моделі є графіки розбіжностей між величинами, що передбачаються, і реальними значеннями. Вони прості для побудови та розуміння і можуть вбудовуватися в програмні продукти. Така автоматизована система може стежити сама за собою і сповіщати користувача, коли величина цих розбіжностей починає виходити за певний граничний рівень.

Можна виокремити шість методів виявлення і аналізу знань:

- класифікація;
- регресія;
- прогнозування часових послідовностей (рядів);
- кластеризація;
- асоціація;
- послідовність.

Перші три використовуються в основному для передбачення, тоді як останні зручні для описання наявних закономірностей у даних.

Класифікація — найпоширеніша модель інтелектуального аналізу даних. З її допомогою виявляються ознаки, що характеризують групу, до якої належить той або інший об'єкт. Це робиться за допомогою аналізу вже класифікованих об'єктів і формулювання деякого набору правил. Наприклад, у багатьох видах бізнесу проблемою є втрата постійних клієнтів. Класифікація допомагає виявити характеристики “нестійких” покупців і створити модель, яка передбачає, хто саме схильний піти до іншого постачальника. Використовуючи її, можна визначити ефективні види знижок та інші вигідні пропозиції, що діють для різних покупців. Завдяки цьому можна утримати клієнтів, витративши стільки грошей, скільки необхідно.

Певний ефективний класифікатор може використовуватися для класифікації нових записів у базі даних у вже існуючих класах і в цьому випадку він набуває характеру прогнозу. Наприклад, класифікатор, що вміє ідентифікувати ризик віддачі позики, може бути використаний для ухвалення рішення у конкретному випадку. Тобто класифікатор використовується для прогнозування вірогідності повернення позики.

Регресійний аналіз використовується, коли відношення між змінними

можна виразити кількісно у вигляді деякої комбінації цих змінних. Отримана комбінація використовується для передбачення значення, якого може набувати цільова (залежна) змінна, що обчислюється на заданому наборі значень вхідних (незалежних) змінних. У простому випадку для цього використовуються стандартні статистичні методи, такі як лінійна регресія, але більшість реальних моделей не укладаються в її рамки. Наприклад, обсяги продажу або фондові ціни складні для передбачення, оскільки можуть залежати від комплексу відношень змінних.

Прогнозування часових послідовностей. Основою для будь-яких систем прогнозування слугує історична інформація, що зберігається в інформаційних сховищах у вигляді часових рядів. Якщо можна побудувати математичну модель і знайти шаблони, що адекватно відображають цю динаміку, є вірогідність, що з їх допомогою можна передбачати і поведінку системи в майбутньому. Прогнозування часових послідовностей дає змогу на основі аналізу поведінки часових рядів оцінювати майбутні значення прогнозованих змінних. Ці моделі повинні мати особливі ознаки часу: ієрархію періодів (місяць — квартал — рік), особливі відрізки часу (п'яти, шести або семиденний робочий тиждень), сезонність, свята та ін.

Кластеризація відрізняється від класифікації тим, що класи заздалегідь не задані і за допомогою моделі кластеризації засоби інтелектуальних обчислень самостійно створюють однорідні групи даних.

Асоціація належить до аналізу структури і застосовується, коли декілька подій пов'язані між собою. Класичний приклад аналізу структури покупок визначає придбання будь-якої кількості товарів як окрему економічну операцію (транзакцію). Оскільки велика кількість покупок відбувається в супермаркетах, а покупці для зручності використовують корзини, куди і складають весь товар, то для пошуку асоціацій слугує аналіз вмісту корзини. Метою підходу є пошук трендів (однакових ділянок) серед великої кількості транзакцій, які можна використовувати для пояснення поведінки покупців. Така інформація може бути використана для регулювання запасів, зміни розміщення товарів на території магазину і ухвалення рішення з проведення рекламної кампанії для збільшення продажу або для просування певної продукції. Наприклад, дослідження, проведене в супермаркеті, може показати, що 65 % людей купують картопляні чипси, беруть і “кока-колу”, а за наявності знижки на такий комплект — “кока-колу” купують у 85 % випадків. Маючи такі дані, менеджерам легко оцінити, наскільки дієва надана знижка.

Хоча цей підхід взятий виключно з роздрібною торгівлі, він може також застосовуватися у фінансовій сфері для аналізу портфеля коштовних паперів і пошуку наборів фінансових послуг, які клієнти часто купують разом. Це може використовуватися для створення деякого набору послуг як частини кампанії зі стимулювання продажу.

Послідовність має місце, якщо існує коло пов'язаних у часі подій.

Традиційний аналіз структури покупок має справу з набором товарів, які становлять одну транзакцію. Варіант такого аналізу трапляється, якщо є додаткова інформація (номер кредитної карти клієнта або номер його банківського рахунку) для скріплення різних покупок в єдину часову серію. У такій ситуації важливе не лише співіснування даних у межах однієї транзакції, а і порядок, в якому ці дані з'являються в різних транзакціях і час між ними. Правила, що встановлюють ці зв'язки, можуть бути використані для визначення типового набору попереднього продажу, який можуть зумовити наступний продаж певного товару. Після придбання будинку в 45 % випадків протягом місяця купується нова кухонна плита, а протягом наступних двох тижнів 60 % мешканців — холодильник.

1.4. Основні моделі інтелектуальних обчислень

Розглянемо основні види моделей, які використовуються для пошуку нового знання на основі даних інформаційного сховища. Метою інтелектуальних технологій є пошук нового знання, яке користувач може надалі застосувати для поліпшення результатів своєї діяльності. Результат моделювання — це виявлення відношень у даних.

На практиці широкого застосування набули такі види (алгоритми) інтелектуальних обчислень:

- нейронні мережі;
- дерева рішень;
- системи роздумів на основі аналогічних випадків;
- алгоритми визначення асоціацій і послідовностей;
- нечітка логіка;
- генетичні алгоритми;
- еволюційне програмування;
- візуалізація даних;
- комбіновані методи.

Нейронні мережі — це системи з архітектурою, що умовно імітують роботу нейронів. Математична модель нейрона є деяким універсальним нелінійним елементом із можливістю широкої зміни і налаштуванням його характеристик. Нейронні мережі є сукупністю пов'язаних між собою прошарків нейронів, які отримують вхідні дані, здійснюють їх обробку і генерують на виході результат. Між вузлами видимих вхідного і вихідного прошарків може бути певна кількість прихованих прошарків. Нейронні мережі реалізують непрозорий процес. Це означає, що побудована модель, як правило, не має чіткої інтерпретації. Багато пакетів, які реалізують алгоритми нейронних мереж, застосовуються у сфері обробки комерційної інформації при розпізнаванні образів, розшифруванні рукописного тексту, інтерпретації кардіограм. Апаратним або програмним реалізаціям алгоритмів нейромереж називаються нейрокомп'ютером. Його основними

особливостями є такі:

- нейрокомп'ютери дають стандартний спосіб розв'язку багатьох нестандартних задач. І неважливо, що спеціалізована машина краще розв'яже завдання одного класу. Важливіше, що один нейрокомп'ютер розв'яже і цю задачу, й іншу, і третю, і не треба кожного разу проектувати спеціалізовану ЕОМ, нейрокомп'ютер зробить все сам і не гірше;

- замість програмування застосовується навчання. Нейрокомп'ютер навчається, потрібно лише формувати навчальну множину. Таким чином, робота програміста замінюється новою роботою вчителя. Краще це чи гірше? Важко визначити. Програміст наказує машині виконати всі елементи роботи, вчитель створює “навчальне середовище”, до якого пристосовується нейрокомп'ютер. З'являються нові можливості для роботи;

- нейрокомп'ютери ефективні там, де потрібен аналог людської інтуїції, зокрема для розпізнавання образів, читання рукописних текстів, підготовки аналітичних прогнозів, перекладу з однієї мови на іншу тощо. Саме для таких завдань зазвичай важко скласти явний алгоритм;

- нейронні мережі дають змогу створити ефективне програмне і математичне забезпечення для комп'ютерів із високим ступенем розпаралелювання обробки;

- нейрокомп'ютери “демократичні”, вони прості, як текстові процесори, тому з ними може працювати будь-який, навіть недосвідчений користувач.

Дерева рішень — метод, широко вживаний у галузі фінансів і бізнесу, де частіше зустрічаються задачі числового прогнозу. В результаті застосування цього методу для навчальної вибірки даних створюється ієрархічна структура правил класифікації типу “Якщо..., то...”, що мають вигляд дерева. Для того, щоб вирішити, до якого класу зарахувати певний об'єкт або ситуацію, треба методом “найближчого сусіда”. Системи роздумів на основі аналогічних випадків дають гарні результати у різних завданнях. Головний їх мінус у тому, що вони взагалі не створюють будь-яких моделей або правил, які узагальнюють попередній досвід. У виборі рішення вони базуються на всьому масиві доступних історичних даних, тому неможливо сказати, на основі яких конкретних чинників ці системи будують свої відповіді.

Алгоритми виявлення асоціацій визначають правила про окремі предмети, які з'являються разом в одній транзакції, наприклад в одній покупці. Послідовність — це теж асоціація, але залежна від часу. Асоціація записується як $A > B$, де A називається передумовою, B — наслідком. Частота появи кожного окремого предмета або групи предметів, визначається дуже просто — підраховується кількість появ цього предмета у всіх подіях (покупках) і ділиться на загальну кількість подій. Ця

величина вимірюється у відсотках і має назву поширеність. Низький рівень поширеності (менше одного тисячного відсотка) свідчить про неістотність асоціації.

Для визначення важливості кожного отриманого асоціативного правила необхідно отримати величину, яка має назву “довірчість A до B ” (взаємозв’язок A і B). Ця величина показує, як часто з появою A з’являється B і розраховується як відношення частоти появи (поширеності) A і B разом до поширеності A . Тобто, якщо довірчість A до B дорівнює 20 %, то це означає, що при покупці товару A в кожному п’ятому випадку купують і товар B . Якщо поширеність A не дорівнює поширеності B , то і довірчість A до B не дорівнює довірчості B до A . Насправді покупка комп’ютера частіше призводить до придбання дискет, ніж купівля дискети до покупки комп’ютера.

Ще однією важливою характеристикою асоціації є потужність асоціації. Чим більше потужність, тим сильніший вплив, який поява A робить на появу B .

$$\text{Потужність} = \{ \text{Довірчість } A \text{ до } B \} : (\text{Поширеність } B).$$

Деякі алгоритми пошуку асоціацій спочатку сортують дані і лише після цього визначають взаємозв’язок і поширеність. Єдиною розбіжністю таких алгоритмів є швидкість або ефективність пошуку асоціацій. Це важливо у зв’язку з величезною кількістю комбінацій, які необхідно перебрати для пошуку більш значущих правил. Алгоритми пошуку асоціацій можуть створювати свої бази даних поширеності, довірчості і потужності, до яких можна звертатися при запиті. Наприклад: “Знайти всі асоціації, в яких для товару X довірчість більше 50 % і поширеність не менше 2,5 %”. При пошуку послідовностей додається змінна часу, що дає змогу працювати з серією подій для пошуку послідовних асоціацій упродовж деякого періоду часу.

Підсумовуючи цей метод аналізу, необхідно сказати, що випадково може виникнути така ситуація, коли товари в супермаркеті будуть згруповані за допомогою знайдених моделей, але це, замість очікуваного прибутку, дасть зворотний ефект. Таке може відбутися через те, що клієнт довго не ходитиме по магазину у пошуках бажаного товару, купуючи при цьому ще щось, що впадає в око, і те, що він ніколи не планував купувати.

Нечітка логіка застосовується для наборів даних, де належність даних до якої-небудь групи є вірогідністю в інтервалі від 0 до 1. Чітка логіка маніпулює результатами, які можуть бути або істиною, або ні. Нечітка логіка застосовується у тих випадках, коли існує “може бути” в доповненні до “так” чи ні”. Галуззю впровадження алгоритмів нечіткої логіки є будь-які аналітичні системи, зокрема:

- нелінійний контроль за процесами (виробництво);
- удосконалення стратегій управління і координації дій, наприклад складне промислове виробництво;

- самонавчальні системи (або класифікатори);
- дослідження ризикованих і критичних ситуацій;
- розпізнавання образів;
- фінансовий аналіз (ринки цінних паперів);
- дослідження даних (корпоративні сховища).

У Японії цей напрям переживає бум. Тут функціонує спеціально створена лабораторія (Laboratory for International Fuzzy Engineering Research, LIFE). Програмою організації є створення найближчих людині обчислювальних пристроїв. LIFE об'єднує 48 компаній, серед яких: *Hitachi, Mitsubishi, NEC, Sharp, Sony, Honda, Mazda, Toyota*. З іноземних учасників LIFE можна виділити: *IBM, Fuji Xerox*, до діяльності LIFE також виявляє цікавість *NASA*.

Потужність та інтуїтивна простота нечіткої логіки як методології вирішення проблем гарантує її успішне використання у вбудованих системах контролю й аналізу інформації. При цьому відбувається підключення людської інтуїції і досвіду оператора. На відміну від традиційної математики, яка потребує на кожному кроці моделювання точних і однозначних формулювань закономірностей, нечітка логіка пропонує зовсім інший рівень мислення, завдяки чому творчий процес моделювання відбувається на високому рівні абстракцій, за якого постулюється лише мінімальний набір закономірностей.

Недоліками нечітких систем є такі:

- відсутність стандартної методики конструювання цих систем;
- неможливість математичного аналізу наявними методами.

Генетичні алгоритми є потужним засобом розв'язання різних комбінаторних задач і проблем оптимізації. Проте генетичні алгоритми увійшли зараз до стандартного інструментарію методів інтелектуальних обчислень. Цей метод названий так тому, що певною мірою імітує процес природного добору в природі. Наприклад, потрібно знайти розв'язок задач, найбільш оптимальних в аспекті деякого критерію, де кожен розв'язок цілком описується певним набором чисел або величин нечислової природи. Скажімо, якщо нам треба вибрати сукупність фіксованої кількості параметрів ринку, що істотно впливають на його динаміку, це буде набір імен цих параметрів. Про цей набір можна говорити як про сукупність хромосом, що визначають якості індивіда, — даного розв'язку поставленої задачі. Значення параметрів, що визначають розв'язок, називаються генами. Пошук оптимального розв'язку при цьому схожий на еволюцію популяції індивідів, представлених наборами хромосом.

В еволюції діють три механізми: по-перше, добір найсильніших наборів хромосом, яким відповідають найбільш оптимальні розв'язки; по-друге, схрещування виробництво нових індивідів за допомогою змішування хромосомних наборів відібраних індивідів; і, по-третє, мутації — випадкові зміни генів у деяких індивідів популяції. В результаті зміни

покоління виробляється розв'язання поставленої задачі, яке вже не може бути далі покращене.

Генетичні алгоритми мають два слабких місця. По-перше, постановка задачі не дає можливості проаналізувати статистичну значущість отриманого за їх допомогою розв'язку, по-друге, ефективно сформулювати завдання, визначити критерій добору хромосом можуть тільки фахівці. Через ці чинники, генетичні алгоритми треба розглядати швидше як інструмент наукового дослідження, ніж засіб аналізу даних для практичного застосування у бізнесі і фінансах.

Еволюційне програмування — наймолодша галузь інтелектуальних обчислень. Гіпотези про вид залежності цільової змінної від інших формулюються системою у вигляді програм певною внутрішньою мовою програмування. Якщо це універсальна мова, то теоретично за її допомогою можна виразити залежність будь-якого вигляду. Процес побудови таких програм відбувається як еволюція в світі програм (цим метод трохи схожий на генетичні алгоритми). Якщо система знаходить програму, яка точно виражає шукану залежність, вона починає вносити до неї невеликі модифікації і відбирає серед побудованих таким чином дочірніх програм ті, які підвищують точність. Система “вирощує” декілька генетичних ліній програм, що конкурують між собою щодо точності визначення шуканої залежності. Спеціальний транслявальний модуль перекладає знайдені залежності з внутрішньої мови системи на зрозумілу користувачеві мову (математичні формули, таблиці та ін.), роблячи їх досяжними. Для того, щоб зробити отримані результати зрозумілішими для користувача не математика, є великий арсенал різноманітних засобів візуалізації виявлених залежностей.

Пошук залежності цільових змінних від інших проводиться у формі функцій будь-якого непевного вигляду. Наприклад, в одному з найбільш вдалих алгоритмів цього типу — методі групового обліку аргументів (МГОА) залежність шукають у формі поліномів. Причому складні поліноми замінюються декількома простими, що враховують лише деякі ознаки (групи аргументів). Зазвичай використовуються попарні об'єднання ознак. Цей метод не має великих переваг порівняно з нейронними мережами з готовим набором стандартних нелінійних функцій, але отримані формули залежності, в принципі, можна аналізувати та інтерпретувати (хоча на практиці складно).

Програми візуалізації даних у певному значенні не є засобом аналізу інформації, оскільки вони тільки представляють її користувачеві. Проте візуальне уявлення, скажімо, відразу чотирьох змінних наочно узагальнює величезні обсяги даних (рис. 1.4).

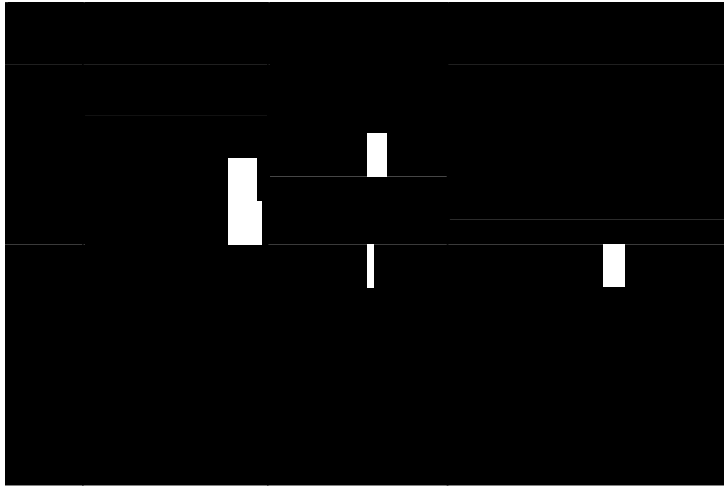


Рисунок 1.4 - Візуалізація показників діяльності інвестиційних фондів

Комбіновані методи. Часто виробники об'єднують зазначені підходи. Об'єднання алгоритмів нейронних мереж і технології дерев рішень сприяє побудові точнішої моделі і підвищенню швидкості. Для вирішення кожної проблеми потрібно шукати оптимальний метод.

1.5 Засоби програмної підтримки інтелектуального аналізу даних

Сьогодні ми є свідками активного розвитку технологій інтелектуального аналізу даних, виникнення яких пов'язане, в першу чергу, з необхідністю аналітичної обробки надвеликих обсягів інформації, що накопичується в сучасних сховищах даних. Можливість використання добре відомих методів математичної статистики і машинного навчання для вирішення подібних завдань відкрило нові можливості перед аналітиками, дослідниками, а також тими, хто ухвалює рішення, — менеджерами і керівниками компаній. Аналітики передбачили, що з 2005 р. такі технології стали дуже привабливою галуззю для ІТ-інвестицій. Згідно з результатами дослідження компанії *Forrester Research*, майже 30 % європейських фірм розглядали можливість впровадження цього програмного забезпечення в 2007 р. За даними компанії *International Data Corporation (IDC)*, у країнах Східної і Південно-Східної Азії (за винятком Японії) у 2005—2011 рр. середньорічний темп зростання рішень становив 21 %.

На ринку програмного забезпечення Data Mining є величезна різноманітність продуктів, що належать до цієї категорії. І не розгубитися в них досить важко. Для вибору продукту слід ретельно вивчити поставлені задачі і позначити ті результати, які необхідно отримати.

На початку 90-х років минулого століття ринок Data Mining

нараховував близько десяти постачальників. У середині 90-х років кількість постачальників, представлених компаніями малого, середнього і великого бізнесу, налічувало більше 50 фірм. Нині до аналітичних технологій, зокрема до Data Mining, виявляється величезний інтерес. На цьому ринку працює безліч фірм, орієнтованих на створення інструментів Data Mining, а також комплексне впровадження Data Mining, OLAP і сховищ даних. Інструменти Data Mining у багатьох випадках розглядаються як складова BI-платформ, до складу яких також входять засоби побудови сховищ і вітрин даних, засоби обробки несподіваних запитів (ad-hoc query), засоби звітності (reporting), а також інструменти OLAP. Розробкою в секторі Data Mining світового ринку програмного забезпечення зайняті як всесвітньо відомі лідери (IBM, Microsoft), так і нові компанії, що розвиваються.

Інструменти Data Mining можна представити або як самостійний розв'язок, або як доповнення до основного продукту. Останній варіант реалізується багатьма лідерами ринку програмного забезпечення. Так, вже стало традицією, що розробники універсальних статистичних пакетів, на додаток до традиційних методів статистичного аналізу, включають у пакет певний набір методів

Data Mining — це такі пакети, як SPSS (SPSS, Clementine), Statistica (StatSoft), SAS Institute (SAS Enterprise Miner). Деякі розробники OLAP-рішень також пропонують набір методів Data Mining, наприклад сімейство продуктів Cog- nos. Є постачальники, які включають Data Mining-рішення у функціональність СУБД: це Microsoft (Microsoft SQL Server), Oracle, IBM (IBM Intelligent Miner for Data).

Інструменти Data Mining можна оцінювати за різними критеріями. Оцінка програмних засобів Data Mining з погляду кінцевого користувача визначається шляхом оцінки набору його характеристик. Їх можна поділити на дві групи: бізнес і технічні характеристики. Цей поділ є достатньо умовним, і деякі характеристики можуть потрапляти одночасно до обох категорій.

Інтуїтивний інтерфейс. Інтерфейс — середовище передачі інформації між програмним середовищем і користувачем, діалогова система, яка дає змогу передати людині всі необхідні дані, отримані на етапі формалізації і обчислення. Він припускає розташування різних елементів, у т. ч. блоків меню, інформаційних полів, графічних блоків, блоків форм, на екранних формах. Для зручності роботи користувача необхідно, щоб інтерфейс був інтуїтивним.

Інтуїтивний інтерфейс дає можливість користувачеві легко і швидко сприймати елементи інтерфейсу, завдяки чому діалог “програмне середовище користувач став простішим і доступнішим. Поняття інтуїтивного інтерфейсу включає також поняття знайомого навколишнього середовища і наявність виразної нетехнічної термінології (наприклад, для

повідомлення користувача про помилку).

Зручність експорту імпорту даних. При роботі з інструментом Data Mining користувач часто застосовує різноманітні набори даних, працює з їх різними джерелами. Це можуть бути текстові файли, файли електронних таблиць та баз даних. Інструмент Data Mining повинен мати зручний спосіб завантаження (імпорту) даних. Після закінчення роботи користувач також повинен мати зручний спосіб вивантаження (експорту) даних у зручне для нього середовище. Програма має підтримувати найбільш поширені формати даних. Додаткова зручність для користувача створюється при нагоді завантаження і вивантаження певної частини (за вибором користувача) полів, що імпортуються або експортуються.

Наочність і різноманітність отримуваної звітності. Ця характеристика припускає одержання звітності в термінах предметної області, а також в якісно спроектованих вихідних формах у тій кількості, яка може надати користувачеві всю необхідну результативну інформацію.

Зручність і простота використання. Істотно полегшує роботу користувача можливість використовувати програм “Майстер”.

Можливості візуалізації. Наявність графічного подання інформації істотно полегшує інтерпретованість отриманих результатів.

Наявність значень параметрів, заданих за умовчанням. Для користувачів-початківців, — це істотна характеристика, оскільки при виконанні багатьох алгоритмів користувач має ставити завдання або вибирати велику кількість параметрів. Особливо багато їх в інструментах, що реалізують метод нейронних мереж. У нейросимуляторах найчастіше наперед задані значення основних параметрів, інколи недосвідченим користувачам навіть не рекомендується змінювати ці значення. Якщо таких значень немає, користувачеві доводиться використати безліч варіантів, перш ніж отримати прийнятний результат.

Кількість методів і алгоритмів. У багатьох інструментах Data Mining реалізовано відразу декілька методів, що дають можливість розв’язувати одну або декілька задач. Якщо для розв’язання однієї задачі (класифікації) передбачена можливість використання декількох методів (дерев рішень і нейронних мереж), користувач має можливість порівнювати характеристики моделей, побудованих за допомогою цих методів.

Можливості пошуку, сортування, фільтрації. Така можливість корисна як для вхідних даних, так і для вихідної інформації. Застосовується сортування за різними критеріями (полями) з можливістю накладання умов. За умови фільтрації вхідних даних з’являється можливість побудови моделі Data Mining на одній із вибірок набору даних. Фільтрація вихідної інформації корисна з погляду інтерпретації результатів. Так, іноді при побудові дерев рішень результати виходять великими за обсягом, і тут можуть виявитися корисними функції

фільтрації, пошуку і сортування. Додаткова зручність для користувача — кольорове підсвічування деяких категорій записів.

Захист, пароль. Дуже часто за допомогою Data Mining аналізується конфіденційна інформація, тому наявність пароля доступу в систему є бажаною характеристикою для інструменту.

Описані характеристики є критеріями функціональності, зручності, безпеки інструменту Data Mining. При виборі інструменту слід керуватися потребами, а також задачами, які необхідно вирішити. Наприклад, якщо точно відомо, що фірмі необхідно розв'язувати виключно задачі класифікації, то можливість розв'язання інструментом інших завдань зовсім не є критичною. Проте слід враховувати, що впровадження Data Mining при серйозному підході потребує серйозних фінансових вкладень, тому необхідно враховувати всі можливі задачі, які можуть виникнути в перспективі.

Ринок інструментів Data Mining визначається широтою цієї технології і внаслідок цього — величезним різноманіттям програмного забезпечення. Найбільш популярна група інструментів містить такі категорії:

- набори інструментів;
- класифікація даних;
- кластеризація і сегментація;
- інструменти статистичного аналізу;
- аналіз текстів (Text Mining), пошук відхилень (Information Retrieval);
- інструменти візуалізації.

Набори інструментів. До цієї категорії належать універсальні інструменти, які включають методи класифікації, кластеризації і попередньої підготовки даних. Перша група:

- **Clementine.** Data Mining з використанням Clementine є бізнес-процесом, розробленим для мінімізації часу розв'язання задач. Clementine підтримує процес Data Mining: доступ до даних, перетворення, моделювання, оцінювання і впровадження. За допомогою Clementine Data Mining виконується з методологією CRISP-DM;

- **IBM Intelligent Miner for Data.** Інструмент пропонує останні Data Mining- методи, підтримує повний Data Mining-процес: від підготовки даних до презентації результатів. Підтримка мов XML і PMML;

- **KXEN (Knowledge extraction Engines).** Інструмент, що працює на основі теорії Вапника SVM. Розв'язує задачі підготовки даних, сегментації, тимчасових рядів і SVM-класифікації;

- **Oracle Data Mining (ODM).** Інструмент забезпечує GUI, PL/SQL-інтерфейси, Java-інтерфейс. Використовувані методи: байєсівська класифікація, алгоритми пошуку асоціативних правил, кластерні методи, SVM та інші;

- Polyanalyst. Набір, що забезпечує всебічний Data Mining, також включає аналіз текстів, дерево рішень, аналіз зв'язків. Підтримує OLE DB for Data Mining і DCOM-технологію;

- SPSS. Один з найбільш популярних інструментів, підтримує безліч методів Data Mining;

- Statistica Data Miner. Інструмент забезпечує всебічний, інтегрований статистичний аналіз даних, має потужні графічні можливості, управління базами даних, а також додатки розробки систем.

Друга група задач представлена інструментами, що реалізують такі рішення:

- Magnum Opus є швидким інструментом пошуку асоціативних правил у даних, підтримується операційними системами Windows, Linux і Solaris;

- Nuggets це набір, що включає пошук асоціативних правил та інші алгоритми;

- Artool, інструмент містить набір алгоритмів для пошуку асоціативних правил у бінарних базах даних (binary databases);

- DM-II System, інструмент включає алгоритм СВА для виконання класифікації на основі асоціативних правил і деяких інших характеристик.

Для розв'язання задач *кластеризації та сегментації* широкого застосування набули:

- ClustanGraphics 3 — ієрархічний кластерний аналіз “згори донизу”, який підтримує потужні графічні можливості;

- Starprobe, заснований на web-кросс-платформовій системі, включає методи кластеризації, нейронні мережі, дерева рішень, візуалізацію тощо;

- Visipoint — кластеризація методом карт Кохонена (Self-organizing Map clustering) і візуалізація;

- Autoclass C “навчання без вчителя” за допомогою байесівських мереж від NASA, працює на основі операційних систем Unix і Windows;

- Reckless є набором кластерних алгоритмів, заснованих на концепції тс-ближніх сусідів. Інструмент перед проведенням кластеризації виконує пошук та ідентифікацію шумів і викидів для зменшення їх впливу на результати кластеризації.

Для розв'язання задач оцінювання і прогнозування застосовуються:

- Alyuda Forecaster XL — інструмент, реалізований у вигляді Exce-надбудови і призначений для розв'язання задач прогнозування та оцінювання з використанням нейронних мереж;

- Excelneuralpackage, в якому реалізовано дві базові парадигми нейронних мереж — багатопршаровий перцептрон і мережі Кохонена.

Як бачимо, ринок програмного забезпечення Data Mining представлений безліччю інструментів і на ньому відбувається постійна

конкурентна боротьба за споживача. Така конкуренція породжує нові якісні розв'язки. Все більша кількість постачальників прагнуть об'єднати у своїх інструментах якомога більшу кількість сучасних методів і технологій. Data Mining-інструменти найчастіше розглядаються як складова ринку Business Intelligence, який, незважаючи на деякий загальний спад в індустрії інформаційних технологій, упевнено і постійно розвивається. Водночас деякі фахівці відзначають відставання наявного програмного забезпечення від теоретичних розробок у зв'язку зі складністю програмної реалізації деяких нових теоретичних розробок методів і алгоритмів Data Mining. У цілому можна резюмувати, що ринок Business Intelligence, зокрема ринок інструментів Data Mining настільки широкий і різноманітний, що будь-яка компанія може обрати для себе інструмент, який влаштує її за функціональністю і можливостями бюджету.

Більш докладніше розглянемо деякі комп'ютерні системи Data Mining, які набули значного поширення в практиці діяльності підприємств та організацій.

Програмний продукт SAS Enterprise Miner (розробник SAS Institute Inc.) є інтегрованою компонентою системи SAS, створеною спеціально для виявлення у величезних масивах даних інформації, яка необхідна для ухвалення рішень. Розроблений для пошуку і аналізу глибоко прихованих закономірностей у даних SAS, Enterprise Miner включає методи статистичного аналізу, відповідну методологію виконання проектів Data Mining (SEMMA) і графічний інтерфейс користувача. Важливою особливістю SAS Enterprise Miner є його повна інтеграція з програмним продуктом SAS Warehouse Administrator, призначеним для розробки та експлуатації інформаційних сховищ, та іншими компонентами системи SAS. Розробка проектів Data Mining може виконуватися як локально, так і в архітектурі "клієнт — сервер".

Пакет підтримує виконання всіх необхідних процедур у рамках єдиного інтегрованого рішення з можливостями колективної роботи і поставляється як розподілений клієнт-серверний додаток, що особливо зручно для здійснення аналізу даних у масштабах потужних організацій. Пакет SAS Enterprise Miner призначений для фахівців з аналізу даних, маркетингових аналітиків, маркетологів, фахівців з аналізу ризиків, фахівців із виявлення шахрайських дій, а також інженерів і учених, відповідальних за ухвалення ключових рішень у бізнесі або дослідницькій діяльності.

У пакеті SAS Enterprise Miner реалізований підхід, який ґрунтується на створенні діаграм процесів обробки даних, що дає змогу уникнути необхідності ручного кодування і прискорити розробку моделей завдяки методиці Data Mining SEMMA. Графічний інтерфейс пакета є інтерфейсом типу "вказати і клацнути". З його допомогою користувачі можуть виконати

всі стадії процесу DataMining від вибору джерел даних, їх дослідження і модифікації до моделювання і оцінки якості моделей із подальшим застосуванням отриманих моделей як для обробки нових даних, так і для підтримки процесів ухвалення рішень.

Пакет SAS Enterprise Miner пропонує різні інструменти для підготовки даних, які дають можливість, наприклад, зробити вибірку або розбивку даних, здійснити вставку неотриманих значень, провести кластеризацію, об'єднати джерела даних, усунути зайві змінні, виконати обробку мовою SAS за допомогою спеціалізованого вузла SAScode, здійснити перетворення змінних і фільтрацію недостовірних даних. Пакет устаткований функціями описової статистики, а також розширеними засобами візуалізації, яка дає можливість досліджувати надвеликі обсяги даних, поданих у вигляді багатовимірних графіків, і проводити графічне порівняння результатів моделювання. Також він надає набір інструментів та алгоритмів прогностичного і описового моделювання, що включають дерева рішень, нейронні мережі, нейронні мережі, що самоорганізуються, методи міркування, засновані на механізмах пошуку в пам'яті (memorybased reasoning), лінійну і логістичну регресію, кластеризацію, асоціації, тимчасові ряди і багато чого іншого. Інтеграція різних моделей і алгоритмів у пакеті Enterprise Miner дає змогу проводити послідовне порівняння моделей, створених на основі різних методів, і залишатися при цьому в рамках єдиного графічного інтерфейсу. Вбудовані засоби оцінки формують єдине середовище для порівняння різних методів моделювання, як з погляду статистики, так і з погляду бізнесу, даючи можливість виявити найбільш відповідні методи для наявних даних. Результатом є якісний аналіз даних, виконаний з урахуванням специфічних проблем конкретного бізнесу.

Отримані моделі можна публікувати для сумісного використання в рамках підприємства за допомогою репозитарію моделей, що є першою на ринку системою управління моделями. Пакет надає низку вбудованих оціночних функцій, що дають змогу порівняти результати різних методів моделювання як у термінах бізнесу, так і з використанням статистичної діагностики. Це дає можливість вимірювати ефективність моделі в термінах її прибутковості. Підсумком робіт з інтелектуального аналізу даних є розгортання створеної моделі — це завершальна стадія, на якій реалізується економічна віддача від проведених досліджень. Процес застосування моделі до нових даних, відомий як скоринг, часто потребує ручного написання або перетворення програмного коду. Пакет SAS Enterprise Miner автоматизує процес підбору коефіцієнтів і надає готовий програмний код для скорингу на всіх стадіях створення моделі, підтримує створення різних програмних середовищ для розгортання моделі мовами SAS, C, Java і PMML.

Система Poly analyst призначена для автоматичного і

напівавтоматичного аналізу числових баз даних і витягання з сирих даних практично корисних знань. Polyanalyst знаходить багатофакторні залежності між змінними в базі даних, автоматично будує і тестує багатомірні нелінійні моделі, що виражають знайдені залежності, виводить класифікаційні правила за навчальними прикладами, знаходить в даних багатомірні кластери, будує алгоритми рішень. Розробник системи Polyanalyst — російська компанія Megaputer Intelligence (“Мегапьютер”). За своєю природою Polyanalyst є клієнт-серверним додатком. Користувач працює з клієнтською програмою Polyanalyst Workplace. Математичні модулі виокремлені в серверну частину — Polyanalyst Knowledge Server. Така архітектура надає природну можливість для масштабування системи.

Математичні модулі (Exploration Engines) та багато інших компонентів Polyanalyst виділені в окремі динамічні бібліотеки і доступні з інших додатків. Це дає можливість інтегрувати математику Polyanalyst в ті програмні ресурси, що існують в інформаційних системах, наприклад у CRM- або ERP-системах.

Основні риси призначеного для користувача інтерфейсу програми: широкі можливості маніпулювання з даними, графіка для представлення даних і візуалізації результатів, майстри створення об’єктів, наскрізний логічний зв’язок між об’єктами, мова символічних правил, інтуїтивне управління через drop-down і popup меню, докладна контекстна довідка. Одиницею Data Mining дослідження в Polyanalyst є проект. Проект об’єднує всі об’єкти дослідження, дерево проекту, графіки, правила, звіти і т. ін. Проект зберігається у файлі внутрішнього формату системи. Звіти досліджень подаються у форматі HTML і доступні через Інтернет (рис. 1.5).

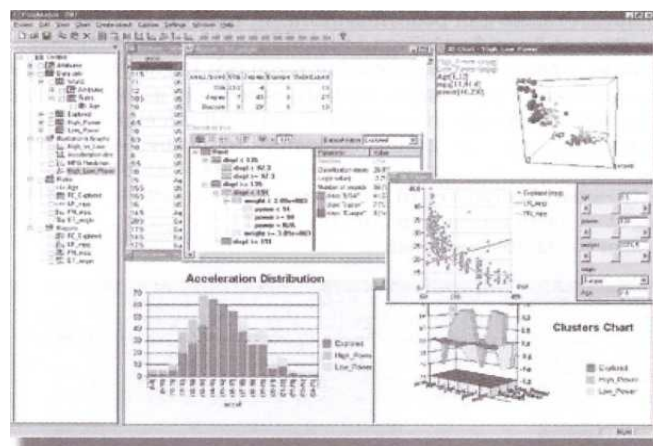


Рисунок 1.5 - Головне меню пакета Polyanalyst

Пакет Polyanalyst включає 18 математичних модулів, заснованих на різних алгоритмах Data і Text Mining. Більшість з цих алгоритмів є Know-Now компанії “Мегапьютер” і не мають аналогів в інших системах. До них належать: моделювання, прогнозування, кластеризація, класифікація,

текстовий аналіз.

Програмні продукти Cognos (розробник — компанія Cognos) — це інструменти інтелектуального або ділового аналізу даних (Business Intelligence Tools) або BI-інструменти. За їх допомогою розв'язують такі задачі.

Робота з запитами і звітами. Рішення в галузі роботи зі звітами орієнтовані на різні типи користувачів. Продукти відрізняються вимогами до рівня складності звітів і рівня навичок кінцевих користувачів:

- Decision Stream — засіб для створення вітрин даних (Data marts), оптимізованих на формування запитів і побудову звітів;
- Impromptu — засіб для роботи з запитами, а також з статичними звітами, що настраюються;
- Powerplay — засіб побудови багатомірних звітів;
- Impromptu Web Reports — засоби для роботи зі статичними звітами через web;
- Cognos Query — засіб для створення запитів, навігації і дослідження даних у т. ч. через web;

Visualizer — засіб для роботи з потужними візуальними звітами.

Аналіз даних. Засоби аналізу даних призначені для аналізу критичної інформації і виявлення значущих чинників. Цей процес охоплює повний набір аналітичних задач і завдань із побудови звітів, включаючи роботу зі звітами бізнес-рівня, можливість переходу до даних нижчого рівня, створення і перегляд уявлень з метою виявлення пріоритетів. Інтеграція засобів дозволяє зручно переходити від дослідження і аналізу даних за допомогою звітів бізнес рівня до дослідження й аналізу даних за звітами нижнього рівня:

- Powerplay — засіб багатовимірного (OLAP) аналізу і побудови бізнес-звітів;
- Impromptu — засіб для перегляду звітів із детальною інформацією нижчого рівня (для Windows);
- Impromptu Web Reports — засіб для перегляду звітів із детальною інформацією нижчого рівня (для Web);
- Visualizer — засіб візуального подання даних.

Візуалізація і виявлення пріоритетів. До розділу візуалізації інформації і виявлення пріоритетів можна віднести цілий спектр продуктів. З їх допомогою користувачеві стає доступна візуалізована інформація, подана в зручному вигляді для виявлення критичних чинників на великих масивах даних. У цих продуктах за основу береться можливість аналізу ключових чинників, що впливають на цю галузь знань (бізнесу) за допомогою широких можливостей з візуалізації даних. Правильно виявлені пріоритети є основою для ухвалення ефективних рішень:

- Visualizer — засіб для подання інформації у формі візуальних представлень з використанням візуальних елементів для виявлення

пріоритетів;

PowerPlay — засіб багатовимірного подання інформації;

- Impromptu — засіб для роботи зі звітами, що будуються;
- CognosQuery — засіб web- користувачів для побудови запитів.

Розвідка даних (datamining). Засоби розвідки і здобуття даних пропонують низку можливостей з автоматизованого перегляду даних, дозволяючи розкривати приховані тенденції, виявляти пріоритетні рішення і дії шляхом відображення тих чинників, які більше за інших впливають на досліджувані показники:

- Scenario — засіб сегментації і класифікації;
- 4Thought — засіб прогнозування (рис. 1.6);
- Visualazer як засіб візуалізації.

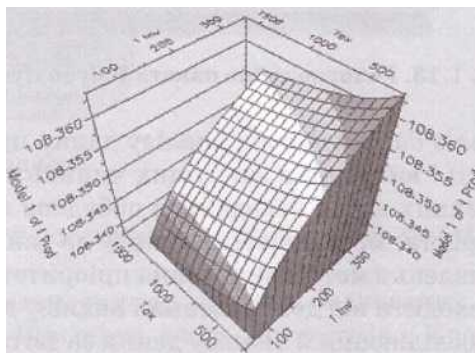


Рисунок 1.6 Інтерфейс програмного комплексу Cognos 4Thought

Захист інформації. Захист інформації відбувається шляхом використання єдиного для всіх застосувань компонента AccessManager, що дає змогу описувати класи користувачів і керувати ними для всіх типів аналітичних додатків Cognos. На додаток до AccessManager можуть бути використані також звичайні можливості забезпечення безпеки на рівні бази даних і операційної системи. На практиці можливе одночасне використання всіх трьох рівнів захисту інформації.

Опис метаданих. Як засіб опису метаданих може бути використаний єдиний для всіх CognosBI-продуктів компонент CognosArchitect. Привабливість використання єдиного для всіх засобів модуля полягає в можливості одноманітного подання бізнес-інформації. Одного разу сформульовані метадані стають доступними в будь-якому аналітичному застосуванні Cognos.

Система STATISTICADataMiner(розробник — компанія StatSoft) спроектована та реалізована як універсальний і всебічний засіб аналізу даних — від взаємодії з різними базами даних до створення готових звітів, що реалізують так званий графічно орієнтований підхід. Основою STATISTICADataMiner є браузер процедур DataMining, який містить більше 300 основних процедур, спеціально оптимізованих під задачі

DataMining, засоби логічного зв'язку між ними і управління потоками даних, що дає змогу конструювати власні аналітичні методи:

Добытчик даних - Мои процедуры Добытчик даних - геепроцедури

Добытчик даних - Чистка и фильтрация даних

Добытчик даних - Вурение и расслоение даних ▶

Добытчик даних - ^глубленные методы классификации ▶

Добытчик даних - Разведчик общих многомерных моделей >

Добытчик даних - Обобщенное прогнозирование »

Добытчик даних - Нейросетевой разведчик •

• ftНейронные сети

И\$! Анализ независимых компонент

об Обобщенные методы (кластерного анализа

1Ш Ассоциативные правила

**frSequence,- Assgdsbn,, andlinkAnalysisЛ Общие деревья классификации и регрессии .Л. Общие CHAID модели Интерактивные деревья*

Растущие деревья классификации и регрессии Ц§§ Обобщенные аддитивные модели

./Ч Многомерные адаптивные £плаины (MAP - сплайны)

И£:Гі Процедура обучения (Сайесовские, ближайшего соседа...)

Бустрое построение прогноза

²iУ Качество подгонки: классификации, прогноза

Щ Отсевание признаков

Объединение групп (классов)

Робочий простір STATISTICA Data Miner складається з чотирьох основних частин:

- Data Acquisition — збирання даних. У цій частині користувач ідентифікує джерело даних для аналізу: файл даних або запит із бази даних.

- Data Preparation, Cleaning, Transformation — підготовка, перетворення і очищення даних. Тут дані перетворюються, фільтруються, групуються.

- Data Analysis, Modeling, Classification, Forecasting — аналіз даних, моделювання, класифікація, прогнозування. Користувач може за допомогою браузера або готових моделей задати необхідні види аналізу даних, таких як прогнозування, класифікація, моделювання.

- Reports — результати. У цій частині користувач може проглянути, задати вигляд і побудувати результати аналізу (наприклад, робоча книга, звіт або електронна таблиця).

Засоби аналізу STATISTICA Data Miner можна поділити на п'ять основних класів:

- General Slicer/Dicer and Drill-Down Еxpіогег-розмітка —

розподіл і поглиблений аналіз. Набір процедур, що дає змогу розбивати, групувати змінні, обчислювати описові статистики, будувати дослідні графіки.

- General Classifier - класифікація. STATISTICA Data Miner включає повний пакет процедур класифікації: узагальнені лінійні моделі, дерева класифікації, регресійні дерева, кластерний аналіз.
- General Modeler/Multivariate Explorer — узагальнені лінійні, нелінійні і регресійні моделі. Цей елемент містить лінійні, нелінійні, узагальнені регресійні моделі та елементи аналізу дерев класифікації.
- General Forecaster — прогнозування. Включає моделі АРСС, сезонні моделі АРСС, експоненційне згладжування, спектральний аналіз Фур'є, сезонну декомпозицію, прогнозування за допомогою нейронних мереж.
- General Neural Networks Explorer — нейромережевий аналіз. У цій частині міститься якнайповніший пакет процедур нейромережевого аналізу.

Наведені вище елементи є комбінацією модулів *інших продуктів* StatSoft. Окрім них, STATISTICA Data Miner містить набір спеціалізованих процедур Data Mining, які доповнюють лінійку інструментів Data Mining:

Feature Selection and Variable Filtering (for very large data sets) — спеціальна вибірка і фільтрація даних (для великих обсягів даних). Цей модуль автоматично вибирає підмножини змінних із заданого файла даних для подальшого аналізу.

- Association Rules — правила асоціації. Модуль є реалізацією так званого апіорного алгоритму виявлення правил асоціації. Наприклад, результат роботи цього алгоритму міг би бути таким: клієнт після купівлі продукту А, у 95 випадках зі 100 протягом наступних двох тижнів після цього замовляє продукт В або С.

- Interactive Drill-Down Explorer — інтерактивний поглиблений аналіз. Є набором засобів для гнучкого дослідження великих наборів даних. На першому кроці ви задаєте набір змінних для поглибленого аналізу даних, на кожному наступному обираєте необхідну підгрупу даних для подальшого аналізу

- Generalized EM & k-Means Cluster Analysis - узагальнений метод максимуму середнього і кластеризація методом середніх. Цей модуль — розширення методів кластерного аналізу, призначений для обробки великих наборів даних і дає змогу кластеризувати як безперервні, так і категоріальні змінні, забезпечує всі необхідні функціональні можливості для розпізнавання образів

- Generalized Additive Models (GAM) - узагальнені адитивні моделі (GAM). Набір методів, розроблених і популяризованих Hastie і Tibshirani.

- **General Classification and Regression Trees (GTrees)** — узагальнюючі класифікаційні і регресійні дерева (GTrees). Модуль є повною реалізацією методів, розроблених Breiman, Friedman, Olshen і Stone (1984). Окрім цього, модуль містить різні доопрацювання і доповнення, такі як оптимізації алгоритмів для великих обсягів даних. Модуль є набором методів узагальненої класифікації і регресійних дерев.

- **General CHAID (Chi-square Automatic Interaction Detection) Models** — узагальнені CHAID-моделі (Хі-квадрат автоматичне виявлення взаємодії). Подібно до попереднього елемента, цей модуль є оптимізацією даної математичної моделі для великих обсягів даних.

- **Boosted Trees** — розширювані прості дерева. Останні дослідження аналітичних алгоритмів показують, що для деяких завдань побудови “складних” оцінок, прогнозів і класифікацій використання послідовно збільшуваних простих дерев дає точніші результати, ніж нейронні мережі або складні цілісні дерева. Цей модуль реалізує алгоритм побудови простих збільшуваних (розширюваних) дерев.

- **Multivariate Adaptive Regression Splines (Mar Splines)** — багатовимірні адаптивні регресійні сплайни (Mar Splines). Цей модуль заснований на реалізації методики, запропонованої Friedman (1991).

Система STATISTICA включає величезний набір різних аналітичних процедур, що робить її недоступною для звичайних користувачів, які погано орієнтуються в методах аналізу даних. Компанією StatSoft запропонований варіант роботи для звичайних користувачів, що володіють невеликими досвідом і знаннями в аналізі даних і математичній статистиці. Для цього, крім загальних методів аналізу, були вбудовані готові закінчені модулі аналізу даних, призначені для вирішення найбільш важливих і популярних завдань: прогнозування, класифікації, створення правил асоціації та ін.

Oracle Data Mining. У березні 1998 р. компанія Oracle оголосила про спільну діяльність з 7 партнерами, які є постачальниками інструментів Data Mining. Потім до Oracle8i включили засоби підтримки алгоритмів Data Mining. У червні 1999 р. Oracle купує Darwin (Thinking Machines Corp.) і у 2000—2001 році виходять нові версії Darwin, Oracle Data Mining Suite. У червні 2001 р. виходить Oracle9i Data Mining.

Oracle Data Mining є опцією або модулем в Oracle Enterprise Edition. Опція Oracle Data Mining (ODM) призначена для аналізу даних за допомогою методів, що відносяться до технології витягання знань, або Data Mining. ODM підтримує всі етапи технології витягання знань, включаючи постановку задачі, підготовку даних, автоматичну побудову моделей, аналіз і тестування результатів, використання моделей у реальних ситуаціях. Важливо, що моделі будуються автоматично на основі аналізу наявних даних про об’єкти, спостереження і ситуації за допомогою спеціальних алгоритмів. Основу опції ODM утворюють процедури, що

реалізують різні алгоритми побудови моделей класифікації, регресії, кластеризації. На етапі підготовки даних забезпечується доступ до будь-яких реляційних баз, текстових файлів, файлів формату SAS. Додаткові засоби перетворення і очищення даних дають змогу змінювати вигляд сценарію, проводити нормалізацію значень, виявляти невизначені або відсутні значення. На основі підготовлених даних спеціальні процедури автоматично будують моделі для подальшого прогнозування, класифікації нових ситуацій, виявлення аналогій. ODM підтримує побудову п'яти різних типів моделей. Графічні засоби надають широкі можливості для аналізу отриманих результатів, верифікації моделей на тестових наборах даних, оцінки точності й стійкості результатів. Уточнені і перевірені моделі можна включати в наявні додатки шляхом генерації їх описів на C++, Java, а також розробляти нові спеціалізовані застосування за допомогою того, що входить до складу середовища ODMзасобу розробки SoftwareDevelopmentKit (SDK).

Аналітична платформа Deductor. Склад і призначення аналітичної платформи Deductor(розробник — компанія BaseGroupLabs). Deductorскладається з двох компонентів: аналітичного додатка DeductorStudioі багатомірного сховища даних DeductorWarehouse. Архітектура системи Deductorподана рис. 1.7.

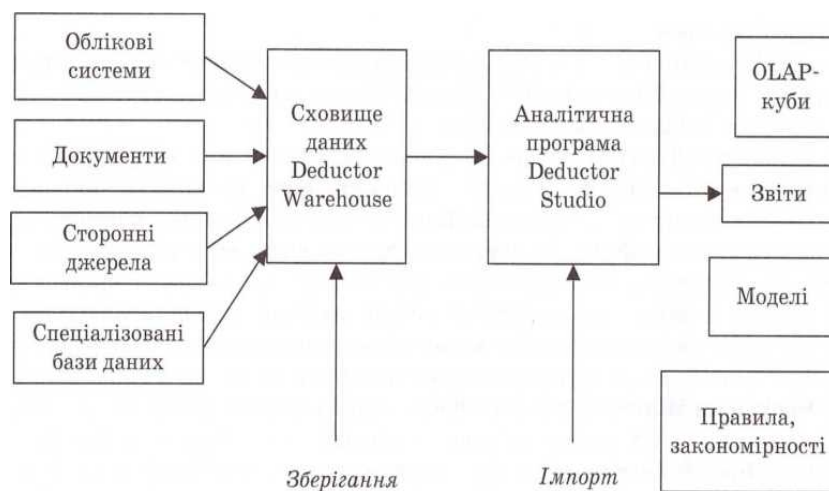


Рисунок 1.7 - Архітектура системи Deductor

Deductor Warehouse — багатовимірне сховище даних, що акумулює всю необхідну для аналізу наочної області інформацію. Використання єдиного сховища дає змогу забезпечити несуперечність даних, їх централізоване зберігання і автоматично створює всю необхідну підтримку процесу аналізу даних. Deductor Warehouse оптимізований для розв'язання саме аналітичних задач, що позитивно позначається на швидкості доступу до даних.

Deductor Studio — це програма, призначена для аналізу інформації з

різних джерел даних. Вона реалізує функції імпорту, обробки, візуалізації та експорту даних. Deductor Studio може функціонувати і без сховища даних, отримуючи інформацію з будь-яких інших джерел, але найбільш оптимальним є їх спільне використання.

Розглянемо цей процес детальніше. На початковому етапі в програму завантажуються або імпортуються дані будь-якого довільного джерела. Сховище даних Deductor Warehouse є одним із джерел даних. Підтримуються також інші сторонні джерела. Зазвичай у програму завантажуються не всі дані, а певна вибірка, необхідна для подальшого аналізу. Після здобуття вибірки можна отримати детальну статистику за нею, проглянути, який мають вигляд дані на діаграмах і гістограмах. Такий розвідувальний аналіз дає можливість ухвалювати рішення про необхідність перед обробки даних. Наприклад, якщо статистика показує, що у вибірці є порожні значення (пропуски даних), можна застосувати фільтрацію для їх усунення. Оброблені дані далі піддаються трансформації. Наприклад, нечислові дані перетворюються на числові, що необхідно для деяких алгоритмів. Безперервні дані можуть бути розбиті на інтервали, тобто виробляється їх дискретизація. До трансформованих даних застосовуються методи глибшого аналізу. На цьому етапі виявляються приховані залежності і закономірності в даних, на підставі яких будуються різні моделі. Модель є шаблоном, який містить формалізовані знання. Останній етап — інтерпретація — призначений для того, щоб з формалізованих знань отримати знання мовою наочної галузі.

Вся робота з аналізу даних у Deductor Studio базується на виконанні таких дій: імпорт даних, їх обробка, візуалізація, експорт. Відправною точкою для аналізу завжди є процедура імпорту даних. Отриманий набір даних може бути оброблений будь-яким із доступних способів. Результатом обробки також є набір даних, який, у свою чергу, знову може бути оброблений. Імпортований набір даних, а також дані, отримані на кожному етапі обробки, можуть бути експортовані для подальшого використання в інших, наприклад в облікових системах. Результати кожної дії можна відобразити різними способами: OLAP- куби (крос-таблиця, крос-діаграма), плоска таблиця, діаграма, гістограма, статистика, аналіз за принципом “що — якщо”, граф нейромережі, дерево ієрархічна система правил тощо.

Послідовність дій, які необхідно виконати для аналізу даних, називається сценарієм. Сценарій можна автоматично виконувати на будь-яких даних. Зразок типового сценарію зображено на рис. 1.8.

Deductor Warehouse — багатовимірне сховище даних, що акумулює всю необхідну для аналізу наочної області інформацію. Вся інформація в сховищі міститься в структурах типу “зірка”, де в центрі розташовані таблиці фактів, а “променями” є виміри. Така архітектура сховища найбільш адекватна завданням аналізу даних. Кожна “зірка” називається

процесом і описує певну дію. У Deductor Warehouse може одночасно зберігатися безліч процесів, що мають загальні виміри.



Рисунок 1.8- Типовий сценарій DeductorStudio

Що є сховищем Deductor Warehouse? Фізично — це реляційна база даних, яка містить таблиці для зберігання інформації і таблиці зв’язків, що забезпечують цілісне зберігання відомостей. Понад реляційною базою даних реалізований спеціальний прошарок, який перетворить реляційну систему на багатовимірну. Багатовимірна ідеологія використовується тому, що вона більша, ніж реляційна, відповідає ідеології аналізу даних. Завдяки цьому прошарку користувач оперує багатовимірними поняттями, такими як “вимір” або “факт”, а система автоматично виконує всі маніпуляції, необхідні для роботи з кореляційною СУБД.

Окрім консолідації даних, робота зі створення закінченого аналітичного рішення містить декілька етапів.

Очищення даних. На цьому етапі проводиться редагування аномалій, заповнення пропусків, згладжування, очищення від шумів, виявлення повторів і суперечностей.

Трансформація даних. Виробляється заміна порожніх значень, квантування, таблична заміна значень, перетворення до ковзаючого вікна, зміна формату набору даних.

Data Mining. Будуються моделі з використанням нейронних мереж, дерев рішень, карт, що самоорганізуються, асоціативних правил.

Програмний комплекс KXEN. Програмне забезпечення KXEN є розробкою однойменної французько-американської компанії, яка працює на ринку з 1998 року. Аббревіатура KXEN означає “Knowledge extraction Engines” — “двигунці” для витягання знань. Відразу слід сказати, що розробка KXEN має особливий підхід до аналізу даних. У KXEN немає дерев рішень, нейронних мереж та інших популярних засобів. KXEN — це інструмент для моделювання, який дозволяє говорити про еволюцію Data Mining і реінжинірингу аналітичного процесу в організації в цілому. В основі цих тверджень лежать досягнення сучасної математики і

принципово інший підхід до вивчення явищ у бізнесі. Слід зазначити, що все те, що відбувається усередині KXEN, дуже відрізняється (принаймні, за своєю філософією) від того, що ми звикли вважати традиційним Data Mining.

Бізнес-моделювання KXEN — це аналіз діяльності компанії та її оточення шляхом побудови математичних моделей. Він використовується в тих випадках, коли необхідно зрозуміти взаємозв'язок між різними подіями і виявити ключові рушійні сили та закономірності в поведінці об'єктів, що цікавлять нас, або процесів. KXEN охоплює чотири основні типи аналітичних задач:

- задачі регресії — класифікації (в т. ч. визначення вкладів змінних);
- задачі сегментації — кластеризації;
- аналіз часових рядів;
- пошук асоціативних правил (аналіз споживчої корзини).

Побудована модель в результаті стає механізмом аналізу, тобто частиною

бізнес-процесу організації. Головна ідея — на основі побудованих моделей створити систему “наскрізного” аналізу процесів, що відбуваються. Це дає можливість автоматично виробляти їх оцінку і будувати прогнози в режимі реального часу (в міру того, як ті або інші операції фіксуються обліковими системами організації).

У 1990 р. були отримані важливі результати в математиці і машинному навчанні. Ініціатором досліджень у цій галузі став Володимир Вапник, який опублікував свою Статистичну теорію навчання. Він був першим, хто “відчинив двері” до нових поглядів на декомпозицію помилки, що виявляється в процесі вжиття методів машинного навчання, виявив і описав структуру цієї помилки та на основі зроблених висновків відшукав спосіб структурування методів моделювання. Математичний апарат, закладений у KXEN, у процесі аналізу будує декілька конкуруючих моделей. Але цей процес здійснюється не випадково (перебиранням різних методів моделювання), а шляхом вивчення різних наборів моделей з опорою на Теорію мінімізації структурних ризиків В. Вапника (Structured Risk Minimization). Творці KXEN розробили механізм порівняння моделей з метою одержання найкращого співвідношення між їх точністю та надійністю, і вже цю оптимальну модель представили як результат аналізу користувачеві.

KXEN Analytic Framework за своєю суттю не є монолітним застосуванням, а виконує роль компонента, вбудованого в наявне програмне середовище. Цей “двигунець” може бути підключений до DBMS-систем (наприклад, Oracle або MS SQL-Server) через протоколи ODBS.

KXEN Analytic Framework є набором модулів для проведення

описового і передбачувального аналізу. Зважаючи на специфіку задач конкретної організації, конструюється оптимальний варіант програмного забезпечення KXEN. Завдяки відкритим програмним інтерфейсам, KXEN легко вбудовується в наявні системи організації. Тому форма подання результатів аналізу, з якою працюватимуть співробітники на місцях, може визначатися побажаннями замовника і особливостями його бізнес-процесу. *Компонент агрегації подій* (KXEN Event Log, KEL) призначений для агрегації подій за певні періоди часу. Застосування KEL дає змогу поєднати транзакційні дані з демографічними даними про клієнта.

Компонент використовується у випадках, коли “сирі” дані містять одночасно статичну інформацію (наприклад, вік, стать або професія індивіда) і динамічні змінні (наприклад, шаблони покупок або транзакції за кредитною картою). Дані автоматично агрегуються всередині визначених користувачем інтервалів без програмування на SQL або внесення змін у схему бази даних. Компонент KEL комбінує і стискає ці дані для того, щоб зробити їх доступними для інших компонентів KXEN. Перевагою використання цього компонента є можливість інтегрувати додаткові джерела інформації “на льоту” для того, щоб поліпшити якість моделі.

Компонент Кодування Послідовностей (KXEN Sequence Coder, KSC) дає змогу реагувати події в серії транзакцій. Наприклад, черговість “кляцань” клієнта, що фіксується на web-сайті, може трансформуватися в ряди даних для кожної сесії. Кожна колонка відображає конкретний перехід з однієї сторінки на іншу. Як і у випадку з KEL, нові колонки даних можуть додаватися до наявних даних про клієнтів і доступні для обробки іншими компонентами KXEN. Перевагою використання цього компонента є можливість застосовувати не задіяні раніше джерела інформації для того, щоб поліпшити якість прогнозуючих моделей.

Компонент погодженого кодування (KXEN Consistent Coder, K2C) дає змогу автоматично підготувати дані і трансформувати їх у формат, що відповідає умовам використання аналітичними додатками KXEN. Використання K2C дає можливість трансформувати номінальні і порядкові змінні, автоматично заповнювати відсутні значення і виявляти викиди. Перевагою використання цього компонента є можливість автоматизації підготовки даних, яка допомагає звільнити час безпосередньо для досліджень і моделювання.

Компонент робастної регресії (KXEN Robust Regression, K2R) використовує відповідний регресійний алгоритм для того, щоб побудувати моделі, що описують наявні залежності, і згенерувати прогнозуючі моделі. Ці моделі можуть потім застосовуватися для скорингу, регресії і класифікації. На відміну від традиційних регресійних алгоритмів, використання K2R дає змогу безпечно вправлятися з великою кількістю змінних (більше 10 000). Модуль K2R будує індикатори і графіки, які дозволяють легко переконатися в якості і надійності побудованої моделі.

Перевагою використання цього компонента є автоматизація процесу інтелектуального аналізу даних. Моделі допомагають деталізувати індивідуальні вклади змінних.

Компонент інтелектуальної сегментації (KXEN Smart Segmenter, K2S) дає змогу виявити природні групи (кластери) в наборі даних. Модуль оптимізований для того, щоб знаходити кластери, які відносяться до конкретно поставленого завдання. Він описує властивості кожної групи і вказує на її відмінності від всієї вибірки. Як і у випадку з іншими модулями, цей модуль також будує індикатори якості і надійності моделі. Перевагою використання цього компонента є автоматичне виявлення груп, важливих для того конкретного завдання, яке необхідно вирішити.

Машина опорних векторів KXEN (Support Vector Machine, KSVM) дає змогу виробляти бінарну класифікацію. Використання компонента, потрібного для розв'язання задач, що ґрунтуються на наборах даних із невеликою кількістю спостережень і великою кількістю змінних. Це робить модуль ідеальним для розв'язання задач в областях з дуже великою розмірністю, таких як медицина і біологія. Перевагою використання цього компонента є можливість вирішення завдань, які раніше вимагали написання спеціальних програм за допомогою промислового програмного забезпечення.

Компонент аналізу часових рядів (KXEN Time Series, KTS) дає можливість прогнозувати значимі шаблони і тренди у часових рядах. Використовує хронологічні дані, що накопичилися, для того, щоб спрогнозувати результати наступних періодів. Модуль KTS виявляє тренди, періодичність і сезонність з метою, одержання точних і достовірних прогнозів. З'являється можливість підлаштуватися під шаблони бізнесу, що повторюються, і передбачати скорочення постачань до того, як вони стануться.

Компонент експорту моделей (KXEN Model Export, KMX) дає змогу створювати коди різного типу: SQL, C, YB, SAS, PMML і багато інших для вбудовування в наявні застосування і бізнес-процеси. Побудована модель у вигляді коду може бути передана на іншу машину для подальшого аналізу даних у пакетному або інтерактивному режимі. Використання цього модуля дає можливість виробляти аналіз нової інформації за допомогою моделі автономно, поза самою системою моделювання. Це істотно прискорює впровадження моделей у виробничий процес та не потребує створення спеціальних умов і програм для аналізу всієї бази даних за допомогою моделі. Користувач також може перекласти модель тією мовою, яку підтримує його комп'ютер.

1.6 Новітні напрями застосування Data Mining

Як зазначалося раніше, галузь використання Data Mining нічим не

обмежена — вона скрізь, де є будь-які дані. Сучасна практика інтелектуального аналізу даних виокремлює два основні напрями застосування систем Data Mining: як масового продукту і як інструменту для проведення унікальних досліджень.

Слід зазначити, що на сьогодні найбільшого поширення технологія Data Mining набула при розв'язанні бізнес-задач. Можливо, причина в тому, що саме в цьому напрямі віддача від використання інструментів Data Mining може становити, за деякими джерелами, до 1000 % і витрати на її впровадження можуть досить швидко окупитися. Зараз технологія Data Mining використовується практично у всіх сферах діяльності людини, де накопичені ретроспективні дані.

Серед новітніх напрямів застосування технологій інтелектуального аналізу даних слід видокремити web-задачі, практика розв'язання яких викликає поширений інтерес і набуває популярності.

Web Mining. Web Mining можна перекласти як “видобуток даних у web”. Web Intelligence готовий “розпочати нову главу” у стрімкому розвитку електронного бізнесу. Здатність визначати інтереси і переваги кожного відвідувача, спостерігаючи за його поведінкою, є серйозною і критичною перевагою конкурентної боротьби на ринку електронної комерції. Системи Web Mining можуть відповісти на багато питань, наприклад хто з відвідувачів є потенційним клієнтом web-магазину, яка група клієнтів приносить найбільший дохід, які інтереси певного відвідувача або групи відвідувачів.

Технологія Web Mining охоплює методи, здатні на основі даних сайту виявити нові, раніше невідомі знання, які надалі можна буде використовувати у практиці. Тобто технологія Web Mining застосовує технологію Data Mining для аналізу неструктурованої, неоднорідної, розподіленої і значної за обсягом інформації, що міститься на web-вузлах. Згідно з таксономією Web Mining, тут можна виділити два основні напрями: Web Content Mining і Web Usage Mining.

Web Content Mining означає автоматичний пошук і одержання якісної інформації зі джерел Інтернету, переобтяжених “інформаційним шумом”. Тут також йде мова про різні засоби кластеризації і анотування документів. У цьому напрямі виділяють два підходи: заснований на агентах та базах даних.

Підхід на основі агентів (Agent Based Approach) включає такі системи:

- інтелектуальні пошукові агенти (Intelligent Search Agents);
- фільтрація інформації — класифікація;
- персоніфіковані агенти мережі.

Приклади систем інтелектуальних агентів пошуку: Harvest (1994), FAQ-Finder (1995), Information Manifold (1995), OCCAM (1996) and ParaSite (1997), ILA (Information Learning Agent) (1995), ShopBot (1996).

Підхід, заснований на базах даних (Database Approach), включає такі системи:

- багаторівневі бази даних;
- системи web-запитів (Web Query Systems). Наприклад, W3QL (1995), WebLog (1996), Lorel (1995), UNQL (1995 and 1996), TSIMMIS (1994).

Другий напрям Web Usage Mining — це виявлення закономірностей у діях користувача web-вузла або їх групи. Аналізується така інформація: які сторінки переглядав користувач, послідовність їх перегляду, які групи користувачів можна виділити серед загальної їх кількості на основі історії перегляду web-вузла.

Web Usage Mining має такі складові:

- попередня обробка;
- операційна ідентифікація;
- інструменти виявлення шаблонів;
- інструменти аналізу шаблонів.

При використанні Web Mining перед розробниками виникає два типи завдань. Перший стосується збирання даних, другий — використання методів персоніфікації. У результаті збирання деякого обсягу персоніфікованих ретроспективних даних про конкретного клієнта, система накопичує певні знання про нього і може рекомендувати йому, наприклад певні набори товарів або послуг. На основі інформації про всіх відвідувачів сайта web-система може виявити певні групи відвідувачів і також рекомендувати їм товари або пропонувати товари в розсилках.

Завдання Web Mining можна розподілити на такі категорії:

- попередня обробка даних для Web Mining;
- виявлення шаблонів і відкриття знань з використанням асоціативних правил, тимчасових послідовностей, класифікації і кластеризації;
- аналіз отриманого знання.

Text Mining. Text Mining охоплює нові методи для виконання семантичного аналізу текстів, інформаційного пошуку і управління. Синонімом поняття Text Mining є KDT (Knowledge Discovering in Text — пошук або виявлення знань у тексті). На відміну від технології Data Mining, яка передбачає аналіз упорядкованої в деякі структури інформації, технологія Text Mining аналізує великі і надвеликі масиви неструктурованої інформації. Програми, що реалізують це завдання, повинні оперувати природною людською мовою і при цьому розуміти семантику аналізованого тексту. Один з методів, на якому засновані деякі Text Mining-системи, — пошук так званого підрядка в рядку.

Call Mining. За словами аналітиків “видобуток дзвінків” може стати популярним інструментом корпоративних інформаційних систем. Технологія Call Mining об’єднує розпізнавання мови, її аналіз і Data

Mining. Meta Call Mining — спрощення пошуку в аудіоархівах, що містять записи переговорів між операторами і клієнтами. За допомогою цієї технології оператори можуть виявляти недоліки в системі обслуговування клієнтів, вишукувати можливості збільшення продажу, а також виявляти тенденції у зверненнях клієнтів. Серед розробників нової технології Call Mining (“видобуток” і аналіз дзвінків) — компанії CallMiner, Nexidia, ScanSoft, Witness Systems.

У технології Call Mining розроблено два підходи — на основі перетворення мови на текст і на базі фонетичного аналізу. Прикладом реалізації першого підходу, заснованого на перетворенні мови, є система CallMiner. У процесі Call Mining спочатку використовується система перетворення мови, потім виконується її аналіз, під час якого залежно від змісту розмов формується статистика телефонних викликів. Отримана інформація зберігається в базі даних, в якій можливі пошук, витягання і обробка. Приклад реалізації другого підходу — фонетичного аналізу — продукція компанії Nexidia. За такого підходу мова розбивається на фонемі, що є звуками або їх поєднаннями. Такі елементи утворюють розпізнавані фрагменти. У процесі пошуку певних слів і їх поєднань система ідентифікує їх з фонемами.

Аналітики зазначають, що впродовж останніх років інтерес до систем на основі Call Mining значно зріс. Це пояснюється тим фактом, що менеджери вищої ланки компаній, що працюють в різних сферах, у т. ч. у галузі фінансів, мобільного зв'язку, авіабізнесу, не хочуть витратити багато часу на прослуховування дзвінків із метою узагальнення інформації або ж виявлення якихось фактів порушень. За словами Деніела Хонг, аналітика компанії Datamonitor: “Використання цих технологій підвищує оперативність і знижує вартість обробки інформації”. Типова інсталяція продукції від розробника Nexidia коштує від 100 до 300 тис. дол. Вартість впровадження системи CallMiner із перетворення мови і набору аналітичних застосувань становить близько 450 тис. дол.

На думку експертів, додатки Audio Mining і Video Mining отримують з часом набагато ширше використання, наприклад при індексації навчальних відеофільмів і презентацій у медіабібліотеках компаній. Проте технології Audio Mining і Video Mining перебувають зараз на рівні становлення, а практичне їх вживання — на початковій стадії.

Детальніше розглянемо можливі варіанти впровадження систем інтелектуального аналізу даних. Різні варіанти впровадження Data Mining мають свої сильні і слабкі сторони. Так, перевагами готового програмного забезпечення є готові алгоритми, технічна підтримка виробника, повна конфіденційність інформації, відсутність програмного коду, можливість придбання різних модулів і надбудов до використовуваного пакета, спілкування з його іншими користувачами тощо. Проте таке рішення має і слабкі сторони. Залежно від інструменту це може бути висока вартість

ліцензій на програмне забезпечення, неможливість додавати власні функції, складність підготовки даних, практична відсутність в інтерфейсі термінів наочної галузі та ін. Таке рішення потребує наявності висококваліфікованих кадрів, які зможуть якісно підготувати дані з аналізу, знають, які алгоритми слід застосовувати для вирішення яких завдань, зуміють інтерпретувати отримані результати у термінах розв'язуваних бізнес-задач. Не кожна компанія може утримати штат таких фахівців, а часто це навіть неефективно.

Уявімо ситуацію, коли менеджер зіштовхується “наодинці” з одним із продуктів, в якому реалізовані методи технології Data Mining (від найпростіших, таких, що включають 1—2 алгоритми, до повнофункціональних програмних комплексів, що пропонують десятки різних алгоритмів). Він має завдання — виявити найбільш перспективних потенційних клієнтів, а бачить перед собою лише набір математичних алгоритмів. Це і є “зворотний бік” використання готових інструментів. Отже, придбанню готового інструменту має передувати серйозна підготовка до впровадження Data Mining.

Розглянемо інший варіант впровадження, який припускає використання Data Mining-консалтингу, або так званої адаптації програмного забезпечення під конкретне завдання. За даними консалтингової компанії Meta Group, у світі не менше 85 % ринку Data Mining займають саме послуги, тобто консультації з ефективного впровадження цієї технології для вирішення актуальних бізнес-задач. В Інтернеті можна знайти перелік із більш ніж ста відомих компаній, що займаються консалтингом у сфері Data Mining. Наприклад, одна зі всесвітньо відомих консалтингових компаній у сфері Data Mining — компанія Two Crows, що спеціалізується на публікації звітності Data Mining, проводить освітні семінари, консультує користувачів і розробників Data Mining у всьому світі. Найвідоміша методологія Data Mining розроблена компанією Two Crows.

До консалтингових Data Mining-компаній належать і виробники готового програмного забезпечення: IBM Global Business Intelligence Solutions, SAS Institute, SPSS, Statsoft та ін. Деякі консалтингові компанії надають свої послуги на певних територіях. Це, наприклад, компанія Arvato Business Intelligence, яка забезпечує Data Mining консультування і моделювання у Франції, Німеччині, Іспанії та інших європейських країнах. Деякі консалтингові компанії спеціалізуються на наданні послуг у певних наочних галузях. Зокрема, компанія Blue Hawk LLC здійснює Data Mining і надає консультаційні послуги в сферах Direct Marketing і CRM. Деякі компанії надають послуги з використанням певних методів Data Mining. Компанія Bayesia надає консультування і “налаштування рішення” під клієнта на основі байєсівської класифікації. Компанія Visual Analytics забезпечує послуги з бізнес-консультування для пошуку шаблонів із

використанням візуального Data Mining.

Розглянемо переваги, які має цей варіант впровадження Data Mining порівняно з готовими програмними продуктами та їх самостійним використанням.

Висококваліфіковані фахівці. Для ефективного застосування технології Data Mining потрібні кваліфіковані фахівці, які зуміють якісно провести весь цикл аналізу. Поки що таких грамотних фахівців на просторах СНД та України небагато, і тому вони досить “дорогі”. Навчання власних фахівців, по-перше, достатньо ризиковане (їх задоволенням переманить конкурент), по-друге, потрібні чималі витрати. Клієнти, скориставшись послугами консалтингової компанії, отримують доступ до висококласних професіоналів компанії, економлячи при цьому значні кошти на пошукові або навчання власних фахівців.

Адаптованість. Готові продукти призначені для вирішення хоч і широкого, але все-таки стандартного і обмеженого кола завдань, тому адаптація продукту до умов конкретного бізнесу лягає на плечі співробітників компанії. Перед замовником знову постане згадана проблема кваліфікованих фахівців. Консалтингова компанія надає послуги, повністю адаптовані під бізнес замовника і його завдання.

Гнучкість інструменту. Його можливість швидко підлаштувати програмне забезпечення під потреби бізнесу:

- можливість вибору найбільш зручних понять, у межах яких повинні бути сформульовані знання або межі наочної області. Так, аналізуючи конкурентів, що діють на ринку, їх можна поділити на “сильних” і “слабких” або ж на “агресивних”, “спокійних” і “пасивних” — залежно від того, що цікавить аналітика в певний момент. Відповідно, знання будуть сформульовані у обраних замовником термінах, і в результаті він отримає рішення саме в тих термінах, які йому цікаві та зрозумілі;

- одержання осмислених і зрозумілих замовникові знань у природній формі. Використання адаптованого під конкретний бізнес програмного забезпечення позбавить користувача від необхідності вивчення формул або залежностей у математичній формі, і надасть знання у найбільш інтуїтивному вигляді.

Розглянемо приклад практичного використання розглянутого підходу, який був застосований компанією Snow Cactus для оцінки кредитоспроможності позичальника банку. Реалізація задачі “Видавати кредит?” відбувалася в системі dm-Score, адаптованій під конкретну бізнес-задачу програмного забезпечення кредитного скорингу. Система кредитного скорингу dm-Score (dm — від Data Mining) впроваджена в банку для аналізу кредитних історій і виявлення прихованих впливів параметрів позичальників на їх кредитоспроможність. Така система повинна вписуватися в інформаційний простір банку, тобто безпосередньо

взаємодіяти з базами даних, де зберігається інформація про позичальників і кредити, з автоматизованою банківською системою (АБС), іншим програмним забезпеченням, і працювати з ними як єдине ціле.

У процесі впровадження фахівці знайомляться з використовуваною в банку АБС системою автоматизації ритейлу, базами даних та ін., погоджують вимоги до системи скорингу — як функціональні, так і не функціональні, а також вивчають, які дані накопичені банком і які задачі вони допомагають вирішувати, здійснюють адаптацію системи відповідно до них. Однією з важливих переваг впровадження системи dm-Score є те, що у цьому процесі враховуються всі індивідуальні вимоги і побажання з боку банку. Важливо також зазначити, що в цьому випадку система dm-Score інтегрується в інформаційний простір банку, а не навпаки, тобто впровадження не потребуватиме будь-яких змін у наявних бізнес-процесах. Отже, у результаті впровадження банк отримує систему скорингу, яка враховує всі специфічні особливості і потреби клієнта.

Описувана система dm-Score дає можливість розв'язувати такі задачі:

- оцінка кредитоспроможності позичальника (скоринг позичальника);
- ухвалення рішення про видачу кредиту або відмову. При цьому система може пояснити фахівцеві банку, чому було ухвалено саме таке рішення;
- визначення максимального розміру кредиту (ліміту кредиту за кредитною картою) на основі скорингу позичальника;
- формування професійної думки про кредитний ризик за позиками;
- вироблення індивідуальних умов кредитування для кожного позичальника з урахуванням ризику для банку;
- прогнозування поведінки позичальника, тобто наявність і частота прострочень конкретного позичальника, середній розмір використовуваного кредиту за кредитною картою та ін.;
- оптимізація анкети позичальника (виключення несуттєвих питань без погіршення якості анкети);
- перевірка анкети конкретного позичальника на повноту і внутрішню не-суперечність;
- вирішення інших завдань, специфічних для конкретного банку.

Ця система робить висновки на основі даних, вже накопичених банком у процесі роботи на ринку роздрібного кредитування. При цьому у процесі впровадження система налаштовується саме на той набір даних, на який орієнтований конкретний банк. Іншими словами, система dm-Score готова працювати з тими даними, які наявні, і не потребує фіксації на якійсь конкретній жорстко заданій анкеті. Під час аналізу даних про

позичальників і кредити застосовуються різні математичні методи, які виявляють в них чинники і їх комбінації, що впливають на кредитоспроможність позичальників, і силу їх впливу. Виявлені залежності становлять основу для ухвалення рішень у відповідному блоці. Блок аналізу повинен періодично використовуватися для аналізу нових даних банку (приходять нові позичальники, поточні проводять виплати), для забезпечення актуальності системи і адекватності схвалюваних нею рішень. Блок ухвалення рішень використовується безпосередньо для одержання висновку системи dm-Score про кредитоспроможність позичальника, про можливість видачі йому кредиту, максимально допустимий розмір кредиту та ін. З цим блоком працює співробітник банку, який або вводить у нього анкету нового позичальника, або отримує її з торгової точки, де банк здійснює програму споживчого кредитування. Завдяки тісній інтеграції системи з інформаційним простором банку, результати роботи цього блоку передаються безпосередньо АБС і системі автоматизації рительно, які вже формують всі необхідні документи, ведуть історію кредиту тощо. Таким чином, і система dm-Score, і всі банківські системи працюють як одне ціле, підвищуючи продуктивність праці співробітників банку.

2 СХОВИЩА ДАНИХ ТА OLAP-ТЕХНОЛОГІЇ

2.1. Концепція сховищ даних

На початку вісімдесятих років минулого століття, в період бурхливого розвитку реєструвальних інформаційних систем, виникло розуміння обмеженості можливостей їх застосування для аналізу даних і побудови на їх основі систем інтелектуального аналізу даних. Реєструвальні системи створювалися для автоматизації рутинних операцій із ведення бізнесу — виписки рахунків, оформлення договорів, перевірки стану складу і т. п. Основними вимогами до систем було забезпечення транзакційності змін, що вносилися, і максимізація швидкості їх виконання. Саме ці вимоги визначили вибір реляційних моделі представлення даних “суть — зв’язок” як основного технічного при побудові реєструвальних систем.

Для менеджерів і аналітиків були потрібні системи, які б давали можливість:

- Аналізувати інформацію в часовому аспекті;
- Формувати довільні запити до системи; доробляти великі обсяги даних;
- Обробляти великі обсяги даних;
- Інтегрувати дані з різних реєструвальних систем.

Очевидно, що реєструвальні системи не задовольняли жодної з вищезгаданих вимог. У цій системі інформація актуальна тільки на момент звернення до бази даних, в інший момент часу за тим же запитом можна отримати абсолютно інший результат. Інтерфейс реєструвальних систем розрахований на проведення тільки певних операцій, а можливості одержання результатів на нерегламентований (ad-hoc) запит дуже обмежені. Обробка значних масивів даних за обсягом також не можлива через налаштування СУБД на виконання коротких транзакцій і неминучого уповільнення роботи решти користувачів. Відповіддю на цю потребу стала поява нової технології організації баз даних — технології сховищ даних (Data Warehouse).

За спостереженнями дослідницької компанії Forrester Research, більшість крупних компаній стикаються з такою проблемою: вони накопичують величезну кількість інформації, яка ніколи не використовується. Практично в будь-якій організації реально функціонує безліч транзакційних систем, орієнтованих на оперативну обробку даних (кожна для конкретного класу задач) і безперервно поповнюють численні бази даних. Окрім цього, часто підприємства володіють величезними обсягами інформації, що зберігається в так званих успадкованих системах. Всі ці дані розподілені мережами персональних комп’ютерів, зберігаються на мейнфреймах, робочих станціях і серверах. Таким чином, інформація є,

але вона розосереджена, неузгоджена, неструктурована, часто надмірна і не завжди достовірна. Тому в більшості організацій ці дані до цих пір не можуть бути використані для ухвалення критичних бізнес-рішень. На вирішення цієї суперечності й спрямована концепція сховищ даних (Data Warehouse).

Зберігання даних є наріжним каменем систем бізнес-аналітики. 87 % компаній мають одне або більше функціонуючих сховищ даних, причому більше половини використовують декілька подібних. Фактично кожні чотири компанії мають двадцять або більше сховищ, логічно інтегрованих, 18 % цих сховищ має обсяг у терабайт або більше, 30 % респондентів чекає, що обсяг їх сховища даних зросте вдвічі або навіть у декілька разів протягом наступних двох або трьох років. Різноманітність джерел даних, що наповнюють сховища, дивує. Природно, кожне сховище використовує транзакційні системи як джерело даних. Проте відомо декілька незвичайних джерел даних, такі як web-сайти електронної комерції (51 %), прив'язки до баз даних, розташованих у замовників або бізнес-партнерів (49 %), сервіси розширеного доступу до даних, які забезпечують характеристичну інформацію про замовників (47 %), витягання інформації з web-сайтів (41 %), сервіси пошуку/виявлення (34 %) і канали нових даних від крупних нових агентств (33 %). Нарешті, 47 % компаній мають одну або більше вітрин даних, і 49 % використовують накопичувачі операційних даних.

Білл Інмон, автор концепції, у своїй класичній статті “Що таке сховища даних” (1996 р.) визначає сховища даних як **“предметно орієнтовані, інтегровані, незмінні, такі, що підтримують хронологію наборів даних, організовані з метою підтримки управління”**. Він розглядає сховища як “єдине і одноосібне джерело істини”, “центр всесвіту” систем підтримки прийняття рішень.

(СППР). “Зі сховищ даних, — пише він, — інформація постачається в різні відділи, фільтруючись відповідно до заданих настройок СППР. Ці окремі бази даних для ухвалення рішень називаються вітринами даних”.

В основі концепції сховищ даних — ідея об'єднання корпоративних даних, розсіяних по системах оперативної обробки даних, історичних архівах та інших зовнішніх джерелах. Ці джерела можуть містити дані, які не використовуються безпосередньо в системах обробки інформації, але які є життєво необхідними для СППР: законодавча база (включаючи податкові прогнози), плани розвитку галузей, статистичні дані, електронні довідники. Як показує практика, рішення, прийняте на основі лише внутрішніх даних, найчастіше виявляється некоректним.

Мета концепції сховищ даних — з'ясувати відмінності в характеристиках даних в операційних і аналітичних системах, визначити вимоги до даних, що містяться у сховищі, загальні принципи і етапи його побудови, основні джерела даних, дати рекомендації з вирішення

потенційних проблем, що виникають при їх вивантаженні, очищенні, узгодженні, транспортуванні і завантаженні до цільової бази даних сховища.

Предметом концепції сховищ даних є не аналіз даних, а власне дані, тобто концепція їх підготовки для подальшого аналізу. Водночас концепція сховища даних визначає не просто єдиний логічний погляд на корпоративні дані, а реалізацію єдиного інтегрованого їхнього джерела.

Розглянемо порівняльні характеристики даних, що застосовуються в інформаційних системах, орієнтованих на їхню операційну і аналітичну обробку (табл. 2.1).

Таблиця 2.1 - Порівняння характеристик даних

Характеристика	Операційні дані	Аналітичні дані
Частота оновлення	Висока, маленькими порціями	Низька, великими порціями
Джерела даних	В основному внутрішні	В основному зовнішні
Обсяги даних, що зберігаються	Сотні мегабайт, гігабайти	Гігабайти та терабайти
Вік даних	Поточні (за період від декількох місяців до одного року)	Поточні та історичні (за період у декілька років, десятки років)
Призначення	Фіксація, оперативний пошук і перетворення даних	Зберігання деталізованих і агрегованих історичних даних, аналітична обробка, прогнозування і моделювання

Згідно з цією таблицею, сховища даних мають переваги порівняно з використанням оперативних систем або баз даних:

- на відміну від оперативних систем, сховище даних містить інформацію за весь необхідний часовий інтервал — аж до декількох десятиліть — в єдиному інформаційному просторі, що робить такі сховища ідеальною основою для виявлення трендів, сезонних залежностей та інших важливих аналітичних показників;

- як правило, інформаційні системи підприємства зберігають і подають аналогічні дані по-різному. Наприклад, одні і ті самі показники можуть зберігатися в різних одиницях вимірювання. Одна й та сама продукція або одні й ті самі клієнти можуть іменуватися по-різному. У системах сховищ невідповідності в даних усуваються на етапі збирання інформації і завантаження її в єдину базу даних. При цьому організуються єдині довідники, всі показники в яких визначаються у однакових одиницях вимірювання;

— дуже часто оперативні системи внаслідок помилок операторів містять деяку кількість неправильних даних. На етапі розміщення у сховищі даних інформація заздалегідь обробляється. Дані за спеціальною технологією перевіряються на відповідність заданим обмеженням і за необхідності коригуються (очищаються). Технологія забезпечує побудову аналітичних звітів на основі надійних даних і своєчасного сповіщення адміністратора сховища про помилки у вхідній інформації;

— універсалізація доступу до даних. Сховище даних надає унікальну можливість отримувати будь-які звіти про діяльність підприємства на основі одного джерела інформації. Це дає змогу інтегрувати дані, що вводяться і вже накопичені в різних оперативних системах, і легко їх порівнювати. При цьому в процесі створення звітів користувач не зв'язаний відмінностями у доступі до цих оперативних систем;

— прискорення одержання аналітичних звітів. Одержання звітів за допомогою засобів, що надаються оперативними системами,— спосіб неоптимальний. Ці системи витрачають багато часу на агрегацію інформації (розрахунок сумарних, середніх, мінімальних, максимальних значень). Крім того, в поточній базі оперативної системи містяться тільки найнеобхідніші і останні за часом дані, тоді як інформація за минулі періоди поміщається в архів. Якщо дані доводиться отримувати з архіву, тривалість побудови звіту зростає вдвічі. Слід також враховувати, що сервер оперативної системи часто не забезпечує необхідної продуктивності при одночасній побудові складних звітів і введенні інформації. Це може катастрофічно позначатися на роботі підприємства, оскільки оператори не зможуть оформлювати накладні, фіксувати відвантаження або одержання продукції в той час, коли виконується побудова чергового звіту. Сховище даних дає змогу вирішити ці проблеми. По-перше, робота сервера сховища не заважає роботі операторів. По-друге, у сховищі крім детальної інформації містяться і наперед розраховані агреговані значення. По-третє, у сховищі архівна інформація завжди доступна для включення у звіти. Все це допомагає значно скоротити час створення звітів і уникнути проблем в оперативній роботі;

— побудова довільних запитів. Інформацію в сховищі даних недостатньо тільки централізувати і структурувати. Аналітикові потрібні засоби візуалізації цієї інформації, інструмент, за допомогою якого легко отримувати дані, необхідні для ухвалення своєчасних рішень. Одна з головних вимог будь-якого аналітика — простота формування звітів і їх наочність. У разі оперативних систем побудова звітів часто позбавлена гнучкості; щоб створити новий звіт, доводиться залучати фахівців інформаційного відділу, які об'єднують дані декількох систем. У разі ж використання сховища даних вирішення проблеми надає технологія OLAP (On-Line Analytical Processing). Ця технологія забезпечує доступ до даних у

термінах, звичних для аналітика. Технологія OLAP базується на концепції багатовимірного представлення даних. Справді, кожне числове значення, що міститься в сховищі даних, має до декількох десятків атрибутів (наприклад, кількість продажу певним менеджером у певному регіоні на певну дату і т. п.). Можна вважати, що робота йде з багатовимірними структурами даних (багатовимірними кубами), в яких числові значення розташовані на перетині декількох вимірювань. Саме цей підхід використовується в OLAP-системах. Вони надають гнучкі засоби навігації багатовимірними структурами — так звані OLAP-маніпуляції. З їх допомогою аналітик може отримувати різні зрізи даних, “крутити” їх.

Як бачимо, з наведених переваг використання технології сховищ даних, велика їх частина може істотно спростити, підвищити швидкість і якісно поліпшити процес DataMining. Отже, комплексне впровадження цих технологій дає розробникам і користувачам незаперечні переваги над використанням розрізнених баз даних різних інформаційних систем при створенні систем підтримки прийняття рішень.

Існує три істотні проблеми, пов’язані зі сховищами даних. Вони полягають в управлінні “брудними” даними, оптимальному виборі джерела даних, а також у продуктивності і масштабованості операцій, заснованих на скануванні.

Наслідки “брудних” даних проявляються у повсякденному житті: ми отримуємо безліч поштових відправлень з орфографічними помилками в іменах, послань з різними варіантами одного і того ж імені, листів, адресованих людям, які давно переїхали, банківські повідомлення про численні витратні операції (тоді як гроші з рахунку знімалися лише один раз) тощо. До “брудних” даних належать відсутні, неточні або непотрібні дані з погляду практичного застосування (наприклад, подані у неправильному форматі, який не відповідає стандарту). “Брудні” дані можуть з’явитися з різних причин: помилка при введенні даних; використання інших форматів або одиниць вимірювання; невідповідність стандартам; відсутність своєчасного оновлення; невдале оновлення всіх копій; невдале видалення записів-дублікатів тощо. Очевидно, що результати запитів, одержання даних або бізнес-аналізу над сховищем, що містить велику кількість “брудних” даних, не можуть вважатися надійними і корисними. Представлені сьогодні на ринку засоби очищення даних (наприклад, продукти компаній Vality/AscentialSoftware, TrilliumSoftware і FirstLogic) допомагають виявляти і автоматично коригувати деякі найбільш важливі типи даних, особливо імена і адреси людей (із використанням національного каталогу імен і адрес). Проте цим засобам належить пройти ще довгий шлях, оскільки сьогодні вони не вміють працювати зі всіма типами “брудних” даних, і не всі компанії використовують навіть наявні засоби. До того ж більшість підприємств не впроваджують надійні методики і процеси, які гарантують високу якість

даних у сховищі. Недостатня увага, що приділяється якості даних, зумовлена нерозумінням типів і обсягу “брудних” даних, що містяться у сховищах; їхньому впливу на ухвалення рішень і виконувannya дій; а також тим фактом, що продукти очищення даних, представлені на ринку, майже не рекламуються або дуже дорого коштують. Для того щоб почати приділяти необхідну увагу якості даних у своїх сховищах, підприємствам перш за все потрібно розібратися в різноманітті можливих “брудних” даних, джерелах їх виникнення і методах їх виявлення та очищення.

Поза сумнівом більшість організацій сьогодні не докладають достатніх зусиль для забезпечення високої якості даних у своїх сховищах. Для цього потрібно мати процес, методології і ресурси для відстежування і аналізу якості даних, методологію для запобігання або виявлення та очищення “брудних” даних, для оцінки їх вартості та витрат на забезпечення високої якості даних.

У EwhaWomen’sUniversity розроблений прототип інструментального засобу DAQUM (DataQualityMeasurement), призначений для відстежування більшості типів “брудних” даних і приписування їх різним типам кількісної міри якості даних залежно від особливостей програм.

Іншою проблемою сховищ даних є проблема вибору джерел даних. Сьогодні проектувальники сховищ даних проектують схему бази цільового сховища даних із використанням засобів моделювання баз даних. Схема бази даних складається з таблиць, стовпців (полів) таблиць, типів даних і обмежень стовпців, а також зв’язків між таблицями. Проектувальники також визначають відображення (перетворення) схем джерел інформації на схему цільового сховища даних. Але як проектувальники можуть переконатися в тому, що сховище містить всі дані, потрібні програмним комплексам і не має тих, які додаткам не потрібні? Відповіді на ці питання ґрунтуються на припущених досвідчених проектувальників, яким доводиться виявляти потреби даних (таблиці і стовпці), опитуючи розробників програмних додатків, бізнес-аналітиків і адміністраторів баз даних. Після початкового створення сховища часто виявляється, що в ньому не містяться дані, потрібні для одержання відповідей на деякі запити, і є ті, що не потрібні додаткам. Хоча вартість зберігання може бути відносно невеликою, всі поля, потрібні і непотрібні, зберігаються в одних і тих самих записах, прочитуються і записуються спільно, що уповільнює швидкість вибірки, збільшує час обробки, а також приводить до неефективного використання середовища зберігання.

Є багато пропозицій стосовно моделювання сховища даних у вигляді репозитарія результатів всіх виконуваних запитів. Автори цих пропозицій намагаються знайти алгоритми, які обиратимуть (для завантаження в сховищі даних) підмножину початкових даних, що мінімізує загальний час відповідей на запити. Деякі автори намагаються також мінімізувати вартість оновлення сховища даних. Іншими словами, вони ґрунтуються на

припущенні, що всі запити до сховища даних можна наперед дізнатися або передбачити, як і всі можливі зміни сховища даних, а, отже, і початкових джерел даних.

Ідеальний спосіб вибірки даних для завантаження в сховище полягає в тому, що визначаються всі запити, які генеруватимуться всіма програмними додатками, що виконуються над сховищем даних, і визначаються таблиці та поля, які фігурують у цих запитах. Визначення всіх запитів до створення сховища даних є важким завданням. Проте це може стати можливим після початкового створення сховища даних за рахунок реєстрації протягом розумного проміжку часу всіх запитів, що надходять від додатків. Аналіз зареєстрованих запитів використовують для упорядкування сховища даних і видалення тих, до яких додатки не мають доступу.

Потенційно корисним і практичним є засіб, який аналізує потреби додатків у даних, автоматично порівнює ці потреби зі схемами джерел даних і видає рекомендації щодо складу оптимального піднабору джерел даних, які потрібно завантажити в сховище, щоб в ньому були всі потрібні дані і не було зайвих. Таким засобом є MaxCentra. Його функціонування спирається на наявність попередньої побудованої бази знань ключових слів, яка представляє потреби програмних додатків у даних. Ключові слова в основному є неявними вказівками таблиць і полів, до яких здійснюватиметься доступ при виконанні запитів, що генеруються додатком. Такий список ключових слів може бути забезпечений бізнес-аналітиками або розробниками програмних додатків, або отриманий автоматично шляхом аналізу запитів від додатків, що виконуються над не оптимізованим сховищем даних. MaxCentra відштовхується саме від цього і підтримки та сприяння проектувальників дає змогу визначити оптимальну схему бази даних для сховища даних.

Для сховищ даних існує також проблема продуктивності та масштабованості. У системах реляційних баз даних для вибірки невеликої кількості потрібних записів без повного сканування таблиці або бази даних використовуються різні методи доступу, такі як індекси на основі хешування. Такі методи доступу ефективні при вибірці за одним ключовим полем (або за невеликою кількістю полів), коли результати є незначною частиною таблиці. Прикладами таких запитів є: “Знайти всіх співробітників віком 25 років” або “Знайти всіх інженерів-програмістів віком 25 років”. Для швидкої відповіді на такі запити можна створити індекси: на стовпці “Вік” таблиці “Співробітники” або на стовпцях “Вік” і “Посада” таблиці “Співробітники” відповідно. Проте методи доступу в загальному випадку не допомагають при відповіді на запити, результатами яких є значна частина таблиці. Прикладами є запити: “Знайти всіх співробітників жіночої статі” або “Знайти молодих співробітників”. Крім того, методи доступу не дають користі, якщо значення стовпця часто

змінюється, оскільки такі зміни потребують перебудови методів доступу.

Крім подібних “простих” запитів є два класи операцій, для яких методи доступу в системах реляційних баз даних не працюють. До першого класу належать операції “агрегації”, що передбачають групування всіх записів таблиці і застосування до згрупованих записів агрегатних функцій (середнього значення, загального числа, суми, мінімального або максимального значення). Цей тип операцій важливий у таких програмних додатках, як аналіз даних web- журналів, сегментації даних про замовників і т. ін. На ринку є продукти MaxScanі AbInitio, призначені для вирішення проблем продуктивності і масштабованості при виконанні цього типу операцій. У MaxScan застосовується метод зберігання таблиць за полями, методи хешування для групування і агрегації записів, а також методи паралельної обробки. Продуктивність і масштабованість при виконанні агрегатних операцій у 10—20 разів перевищують показники систем реляційних баз даних. Пакет AbInitio є засобом ETL, за якого в механізмі трансформації даних використовуються методи підвищення продуктивності.

До іншого класу належать операції “переміщення файлів”, що читають і/ або записують файли повністю. Цей тип операцій важливий на етапі “перетворення даних”, що потребує великих тимчасових витрат при створенні їх сховища або на етапі “підготовки даних” при автоматичному витяганні знань (здобутті даних) з наявних джерел. Етап перетворення даних включає трансформацію формату і подання даних у заданих полях (зміна одиниці вимірювання, зміна формату дати і часу, зміна абрєвіатури і т. ін.), злиття двох або більше полів в одне, розщеплення поля на два або більше, сортування таблиці, побудову узагальненої таблиці з таблиць, що містять деталізовані дані, створення нової таблиці шляхом з’єднання двох або більше таблиць, розщеплення таблиці на дві або більше тощо. До етапу підготовки даних відноситься перетворення даних заданого поля в цифровий код (у нейронних мережах), безперервних цифрових даних у заданому полі на категоричні дані (наприклад, вік, що перевищує 60 років, вважається “пенсійним”), додавання до запису нового поля, взяття з таблиці зразків даних, реплікація в таблиці деяких записів (для досягнення бажаного розподілу записів) і т. ін.

Сьогодні операції переміщення файлів майже повністю залежать від послідовних операцій систем реляційних баз даних над файлами, тобто читання одного або більше файлів, створення тимчасового файла і запису результуючого файла (файлів). Частота виконання подібних операцій і обсяг використовуваних даних може зробити виправданим застосування сервера перетворення/ підготовки даних. В ідеалі такий сервер може складатися з декількох процесорів, які працюють паралельно. Незалежно від конфігурації процесора на ньому повинні виконуватися програмні засоби перетворення/підготовки даних, спроектовані для паралельної

обробки. Коли це виправдано, потрібно застосовувати конвеєрну паралельну обробку, за якої одержання часткових результатів однієї операції ініціюють виконання іншої без потреби очікування завершення першої операції. Конвеєрна паралельна обробка усуває потребу в записі в тимчасовий файл повних результатів однієї операції і в їх читанні наступною операцією, що дає можливість заощадити декілька процесів обробки файлів.

Технологія сховищ даних виникла у США, а потім отримала визнання в усьому світі. Перші сховища даних з'явилися ще на початку 80-х років минулого століття (тоді вони називалися бази атомарних даних), але тільки до кінця 80-х років була повною мірою усвідомлена необхідність інтеграції корпоративної інформації і належного управління нею, а також з'явилися технічні можливості для створення відповідних систем, спочатку названих “сховищами інформації” (Information Warehouse, IW), які пізніше отримали свою сучасну назву “сховища даних” (Data Warehouse, DW). На початку 90-х років, з появою технологій витягання, перетворення і завантаження даних (Extraction, Transmission and Loading, ETL) і оперативної аналітичної обробки (On-Line Analytic Processing, OLAP) почалося активне поширення сховищ даних у комерційному секторі. Цьому процесу сприяло також опублікування першої книги Білла Інмона (W.H. Inmon. Building the Data Warehouse, QED/Wiley, 1991—312 p.), який отримав загальне визнання як “батько концепції сховища даних”. Незабаром технологія сховищ даних перетворилася на розвинуту архітектурну будову, відому як фабрика корпоративної інформації (Corporate Information Factory, CIF).

Іншим важливим напрямом, який сприяв виникненню сховищ даних, була поява систем підтримки ухвалення рішень (Decision Support Systems, DSS) та інформаційних систем для вирішень (Executive Information Systems, EIS). Непретендуючи на володіння власним інтелектом, на відміну від всього того, що обговорювалося в 60—70-ті роки і що залишилося вічними ідеями, системи DSS і EIS виявилися практично корисними: вони стали прототипами сучасних

систем розподілу даних, онлайнаної аналітики, нової дисципліни корпоративного управління на основі знань. Системи DSS були спочатку налаштовані на менеджмент середньої ланки, а EIS намагалися дати більш загальне і багатовимірне бачення поля даних — для керівників корпоративного рівня. Саме ці дві технологічні ідеї, що глибоко перетиналися між собою, стали, найімовірніше, прямими предками сучасної концепції сховищ даних.

Розглянемо типові проблеми, що вирішуються за допомогою сховищ даних. До них належать, зокрема, аналіз клієнтської бази, аналіз продажу і доходів, а також управління пасивами і активами та ін.

Аналіз клієнтської бази дає змогу сформувати цільові сегменти

клієнтів і використовувати цю інформацію під час продажу банківських продуктів та послуг. Цільові сегменти формуються на основі демографічних і фірмо графічних відомостей, фінансових показників (наприклад, обороту або прибутку), галузевих ознак та інших параметрів клієнтів. Одним із найбільш важливих питань є виділення сегментів прибуткових клієнтів із метою їх подальшого утримання. Зокрема, за рахунок детальнішої сегментації підрозділи маркетингу починають краще розуміти потреби клієнтів і можуть використовувати ці дані при проведенні маркетингових кампаній. Аналіз клієнтської бази і сегментація дають можливість наблизитися до реалізації концепції індивідуального маркетингу й ефективніше застосовувати систему управління взаєминами з клієнтами.

Аналіз продажу допомагає виявляти тенденції, планувати продаж за продуктами, клієнтами, підрозділами і, виходячи з результатів збуту, будувати механізми стимулювання клієнтських та продуктових підрозділів. Завдяки використанню сховища даних можна отримати інтегроване уявлення про результати продажу і взяти цю інформацію на озброєння при формуванні планів.

Аналіз доходів актуальний для будь-якого банку, причому найбільше затребуваний аналіз у розрізі клієнтів. Дуже важливо мати уявлення про розподіл доходів за продуктами і послугами, каналами надання послуг і підрозділами банку. Аналіз доходів у розрізі клієнтів і продуктів дає змогу формувати “унікальні” пропозиції для кожного “унікального” клієнта з метою максимізації прибутку в довгостроковій перспективі. Він сприяє формуванню цінової політики банку, виокремленню сегментів, продуктів і послуг, які стратегічно важливі для нього. Наприклад, банк ChaseManhattanBank має сховище обсягом більше 560 Гбайт, компанія MastercardOnline— 1,2 Тбайт. Коли всі дані містяться в єдиному сховищі, вивчення зв'язків між окремими елементами даних може бути плідним, а результатом аналізу стають нові знання.

Управління активами і пасивами. За допомогою сховища даних можна проводити ефективний аналіз активів і пасивів та управляти не тільки ними, а й миттєвою ліквідністю банку на основі інструментального і портфельного підходів. Ці завдання вирішуються за мінімальних витрат на підготовку спеціальних даних і з урахуванням лише обмеженого обсягу інформації, що збирається з джерел у філіях. Програмний комплекс забезпечує завантаження з інформаційних джерел семи типів і дає змогу формувати декілька десятків звітів.

Є різні типи сховищ даних, які мають свою специфіку.

Фінансові сховища даних. У більшості випадків фінансові сховища даних — це сховища, які організації будують у першу чергу. Створення фінансового сховища — дуже привабливе рішення, оскільки:

- фінансові дані завжди зберігаються в центрі “мозку”

організації;

- у більшості організацій фінансові дані становлять найменші обсяги даних з-поміж інших;
- фінанси охоплюють всі аспекти функціонування компанії;
- фінансові дані за своєю природою мають структуру, на яку безпосередньо впливає повсякденна практика обробки фінансової інформації.

З цих причин фінанси стають переважаючою галуззю побудови корпоративного сховища даних. Проте фінансові сховища даних мають серйозні, властиві тільки цьому типу сховищ, недоліки. Перший з них полягає в тому, що за очікуванням організації, відомості з фінансових сховищ із точністю до однієї копійки збігатимуться з даними реального фінансового середовища. Часто можна почути, що “це фінансове сховище явно не працює, тому що в звіті, який я отримав вчора, було зазначено, що доходи становлять 145,99832 доларів, а коли я виконав той же звіт у фінансовому сховищі даних, то отримав величину, яка дорівнює 139,76201 доларам. Цьому сховищу просто не можна довіряти”. Очікування того, що інформація у фінансовому сховищі даних повинна повністю відповідати цифрам із поточного фінансового звіту, є глибоко помилковим. Люди (тобто фінансові працівники), які так думають, просто не розуміють, що, при переході даних з операційного середовища у фінансове сховище даних, відбувається трансформація. А коли дані перетікають з середовища додатків до реального середовища компанії, їх розглядають в іншому вимірі. При такому переході даних з одного середовища в інше:

- змінюються звітні періоди. В операційному середовищі звітний період закінчується в кінці місяця, у середовищі сховища даних закінчується на корпоративному календарі;
- змінюються схеми групування і кодування рахунків. У операційному середовищі дані розраховуються відповідно до одного плану бухгалтерських рахунків, а у фінансовому середовищі всієї компанії може бути абсолютно інший набір схеми групування і кодування;
- змінюються класифікації даних. Так, в операційному середовищі Північна Америка складається всього з 48 континентальних штатів, у глобальному сховищі даних Північна Америка включає також Канаду, Мексику, Аляску і острови Карибського басейну;
- змінюються валюти. Операційні грошові кошти відповідають тій валюті, в якій вони перебувають в обігу: гривні, євро, фунти стерлінгів, долари тощо. У глобальному середовищі всі гроші зводяться до однієї загальної валюти: долари або євро.

Сховища даних у галузі страхування.

Сховища даних у галузі страхування за деякими невеликими винятками схожі на інші сховища. Перший виняток (і це особливо справедливо відносно страхування життя) полягає в тому, що тривалість

існування наявних сховищ дуже велика. Такі сховища містять дані, які є старими, дуже старими. Причина, з якої страхові компанії вимушені цікавитися такими даними, — обережна обробка даних. Практично для кожної справи є доказ, що діяльність, якою організація займалася в 1950 р., практично не пов'язана з сьогоdnішнім заняттям. І часто цей доказ є правдоподібним.

Друга відмінність цих сховищ визначається датами, які зберігаються в цьому бізнесі. Середовище страхування вирізняється наявністю величезної кількості дат, пов'язаних із бізнесом, ніж із будь-яким іншим видом діяльності. Так, у сфері роздрібно́ї торгівлі є декілька важливих дат: дата продажу, дата появи на складі, можливо, дата виробництва. У банківській справі суттєвою є дата транзакції. У телекомунікації — це дата телефонного дзвінка. У страхуванні є дати різних типів.

Третя відмінність полягає в тому, що ці сховища даних використовують свій робочий цикл ділової активності. Більшість організацій має обмежений і короткий економічний цикл. Так, у банках — це роздруковування чека. У торгівлі — купівля виробу. У телефонній компанії — дзвінок. Проте у страхуванні ним може бути заявка на страхове відшкодування, яка можливо буде задоволена через п'ять років. Або закриття поліса може супроводжуватися двомісячним відстроченням. Резюмуючи, можна сказати, що швидкість, з якою функціонує страхування, відрізняється від швидкості, характерної для інших галузей. Ця різниця у швидкості відбивається на сховищі даних. У інших сховищах транзакції просто збираються і обробляються. У галузі страхування транзакція може відкладатися на невизначений термін, а її різні частини можуть відбиватися в сховищі даних. Результатом є особливий підхід при проектуванні і впровадженні таких сховищ даних.

Сховища даних для управління людськими ресурсами. Сховища даних для управління людськими ресурсами мають істотні відмінності від інших. Перша відмінність — кількість предметних областей. Таке сховище даних неминуче має одну важливу предметну область — це працівник. Практично все інше підпорядковане цій галузі або має другорядне значення. Більшість інших сховищ даних мають декілька базових предметних областей. Проте основна відмінність сховищ даних для управління людськими ресурсами полягає в тому, що такі сховища взагалі використовують дуже мало транзакцій. Так, є дата, коли суб'єкт стає працівником, дата, коли людина звільняється, а також річні надбавки і підвищення. Але, крім транзакцій фонду заробітної плати та інших рідкісних, згенерованих працівником транзакцій, у такому сховищі практично більше нічого і немає. Порівняйте сферу управління людськими ресурсами з комунікацією або банківським середовищем, і різниця в числі транзакцій стане очевидною. Ця різниця в темпах транзакцій між тією, що розглядається й іншими сферами діяльності, є причиною виникнення

певних труднощів, які полягають у тому, що в галузі управління людськими ресурсами спостерігається тенденція до об'єднання операційної обробки людських ресурсів і обробки людських ресурсів для систем ухвалення рішення в одне середовище. В інших галузях спокуса зробити таку архітектурну помилку зовсім невелика.

Глобальні сховища даних. Глобальні сховища даних призначені для глобального представлення корпорації. Розрізняють три типи таких сховищ:

- географічно превалююча обробка даних. Наприклад, необхідно інтегрувати бізнес у Гонконзі з бізнесом у Парижі, який у свою чергу потрібно інтегрувати з Ріо-де-Жанейро, а той — з Нью-Йорком;
- обробка даних, що функціонально превалює. Виробнича діяльність повинна бути інтегрована з постачанням, яке необхідно інтегрувати з продажем, потім — з дослідженнями та ін.;
- галузева превалююча обробка даних. Наприклад, потрібно інтегрувати друкарську справу з консалтингом, який підлягає інтеграції з бізнесом у сфері медичного устаткування, а той — зі спеціалізацією в галузі програмного забезпечення.

Особливість глобального сховища даних полягає в тому, що на глобальному рівні мало загальних вимірників. Єдиний загальний вимірник — це гроші. І інтеграція бізнесу може бути досягнута тільки з його допомогою. Інші ж вимірники можуть мати або не мати сенсу на глобальному рівні. Клієнт, продукт, постачальник, транзакція — всі ці класичні предметні галузі можуть бути як наявними, так і відсутніми в глобальній інтегрованій сфері — глобальному сховищі даних. Крім цього, глобальне сховище даних схильне до того, від чого інші сховища захищені — до руйнівної дії змін. Якщо в інших сховищах зміни базових даних трапляються нечасто, то для цього типу сховищ вони відбуваються постійно і в самій основі. Так, у будь-який момент може бути відкрите нове родовище нафти, наприклад у Венесуелі. У наступну хвилину в Перу спалахне революція. А потім, завдяки розвитку технології, стануть доступними поклади нафти в Луїзіані. Одразу це потягне санкції ОПЕК. У Мексиці буде змінено законодавство. І так далі. Якщо розглядати ситуацію в глобальному аспекті, то видно, що зміни мають постійний характер. Тому структура і технологія, використовувана для розміщення і обслуговування глобального сховища даних, повинна дозволяти підтримувати ці безперервні зміни.

Сховища даних із можливостями Data Mining/ Data Mining і Exploration. Сховища даних, що підтримують технологію Data Mining і Exploration, є гібридом класичних сховищ. Такі сховища використовуються для виконання потужної статистичної обробки даних, є дуже детальними, глибоко історичними, оптимізованими для статистичного аналізу. Крім того, для них характерна орієнтація на певний проект. Це означає, що, на

відміну від всіх інших типів сховищ даних, їх перестають використовувати відразу після закінчення аналізу, заради якого вони створювалися. Ще одна важлива відмінність сховищ даних із можливостями DataMining/DataMining Exploration полягає в тому, що ці сховища дуже часто включають зовнішні дані. Такі дані дуже корисні з погляду забезпечення бізнес-перспективи, яку не так легко побачити без їх участі.

Сховища даних у галузі телекомунікації. Відмітна особливість цих сховищ полягає в тому, що вони значною мірою визначаються даними, згенерованими в деталях на рівні дзвінка. Зрозуміло, у галузі телекомунікації наявні безліч інших типів даних. Але жодна інша галузь сховищ даних не зумовлюється такою мірою розміром однієї предметної області — деталями на рівні дзвінка. Є багато способів зберігання деталей на рівні дзвінка:

- зберігання деталей на рівні дзвінка тільки за декілька місяців;
- зберігання безлічі деталей на рівні дзвінка, розміщених на різних носіях;
- резюмування або агрегація деталей на рівні дзвінка;
- зберігання тільки відібраних деталей на рівні дзвінка та ін.

На жаль, незважаючи на різноманітність методів обробки, для цього сховища даних обробка може бути виконана тільки над деталями на рівні дзвінка. А робота на підсумковому або агрегованому рівні просто неможлива.

2.2 Технології побудови сховищ даних

Ідея, покладена в основу технологій інформаційних сховищ, полягає в тому, що проводити оперативний аналіз безпосередньо на базі інформаційних систем неефективно. Натомість, всі необхідні для аналізу дані витягуються з декількох традиційних баз даних (в основному реляційних), перетворюються і потім поміщаються в одне джерело даних — сховище даних

У процесі розміщення дані:

- очищуються — усунення непотрібної інформації;
- агрегуються — обчислення сум, середніх;
- трансформуються — перетворення типів даних, реорганізація структур зберігання;
- об'єднуються із зовнішніх і внутрішніх джерел — приведення до єдиних форматів;
- синхронізуються — відповідність одному моменту часу.

Сьогодні технології побудови сховищ даних є основою для створення повноцінних інтелектуальних систем аналізу даних, орієнтованих на рішення слабо структурованих задач прийняття рішень, оскільки вони містять дані, що мають такі властивості.

Цілісність і внутрішній взаємозв'язок.

Хоча дані розміщуються з різних джерел, але вони об'єднані єдиними законами іменування, способами вимірювання атрибутів та ін. Це має велике значення для корпоративних організацій, в яких одночасно можуть експлуатуватися різні за своєю архітектурою обчислювальні системи, що подають однакові дані по-різному. Наприклад, можуть використовуватися декілька різних форматів подання дат, або один і той же показник може називатися по-різному, наприклад “вірогідність доведення інформації” і “вірогідність одержання інформації”. У процесі розміщення такі невідповідності усуваються автоматично.

Предметна орієнтованість. Локальні бази даних містять мегабайти інформації, абсолютно не потрібної для аналізу (адреси, поштові індекси, ідентифікатори записів та ін.). Така інформація не вноситься у сховище, що обмежує спектр даних, що розглядаються при ухваленні рішення, до мінімуму.

Відсутність часової прив'язки. Оперативні системи охоплюють невеликий інтервал часу, що досягається за рахунок періодичної архівації даних. Сховища даних, навпаки, містять історичні дані, накопичені за великий інтервал часу (роки, десятиліття).

Доступність лише для читання. Модифікація даних не проводиться, бо вона може призвести до порушення цілісності сховища даних. Оскільки не потрібно мінімізувати час занурення, то структура сховища може бути оптимізована для обробки певних запитів, що досягається за рахунок денормалізації реляційної схеми, попередньої агрегації і побудови найбільш доречних індексів.

Інтегрованість означає, що дані відповідають вимогам всього підприємства, а не одній функції бізнесу. Цим сховище даних гарантує, що однакові звіти, які згенерували для різних аналітиків, міститимуть однакові результати.

Незмінність означає, що, потрапивши один раз у сховище, дані там зберігаються і не змінюються, вони можуть лише додаватися.

Всі дані, які містяться в сховищі, можна розділити на такі категорії: метадані (дані про дані), агреговані та детальні.

Важливою особливістю систем інтелектуального аналізу даних на основі сховищ даних є метадані. Це різні системні словники, що дають змогу контролювати склад і структуру інформації у сховищі, керувати процесами завантаження і розрахунків тощо. Без прошарку управлінських даних сховищу з часом загрожує перетворення на велике електронне звалище.

Ключовою відмінністю інформації сховища є не тільки спосіб його наповнення (з різних зовнішніх і внутрішніх джерел), а й модель зберігання даних, особливо це стосується агрегованої інформації. У сховищі інформація розміщується в денормалізованому вигляді у формі

класичної “сніжинки” або “зірки”. Такий підхід дає змогу істотно знизити час реагування бази даних при виконанні запитів. Не вдаючись до технологічних особливостей проектування, зазначимо, що відбувається це за рахунок деякої надмірності зберігання даних. Така форма зберігання інформації в корпоративних сховищах є загальносвітовою практикою, разом з тим частина детальних даних цілком може зберігатися в нормалізованих таблицях. Вимога до спеціальних структур зберігання обумовлена деяким спеціальним способом використання даних, про що доцільно сказати докладніше. Використовувати інформацію, накопичену в сховищі, можна як за допомогою традиційних звітів, так і з використанням динамічних запитів до бази даних. Існують також абсолютно специфічні способи використання інформації, призначені спеціально для аналітичних завдань. До них належать так звані OLAP-технології (On-line Analytical Processing) і технології інтелектуального аналізу даних. З практичного погляду рішучий крок, який робить OLAP-технологія, полягає в тому, щоб, відмовившись від зайвої спільності, зробити процес аналізу максимально швидким. У рамках цієї технології передбачається, що склад і структура показників для аналізу відомі наперед і змінюються дуже рідко (для деяких видів систем — практично не змінюються). Користувач може виконувати над даними в такому багатовимірному поданні набір OLAP-операцій підйому (консолідації за деякими напрямками), спуску (деталізації за деяким напрямком), повороту (зміни напрямку сортування). Детальні дані в системах сховищ даних найчастіше є джерелом для інтелектуального аналізу.

Отже, даним в інформаційному сховищі даних, організованім в інтегровану цілісну структуру, властиві природні внутрішні зв'язки, вони набувають нових властивостей, що дає можливість отримати статус інформації.

Розглянемо характеристики інтеграції даних у сховищі даних. Як відомо, метою інтеграції даних є одержання єдиної та цільної картини бізнес-даних. Для її досягнення застосуємо модель, яка включає додатки, продукти, технології та методи:

- додатки — рішення, створені постачальниками відповідно до вимог клієнтів, які використовують один або більше продуктів інтеграції даних;
- продукти — готові комерційні рішення, що підтримують одну або більше технологій інтеграції даних;
- технології реалізують один або більше методів інтеграції даних;
- методи — підходи до інтеграції даних, незалежні від технологій.

Є три основні методи інтеграції даних: консолідація, федералізація і поширення.

Консолідація даних. При використанні цього методу дані збираються з декількох первинних систем та інтегруються в одне постійне місце зберігання. Таке місце зберігання може бути використане для підготовки звітності та проведення аналізу, як у випадку застосування сховища даних, або як джерело даних для інших додатків, як у випадку впровадження операційного складу даних. При використанні цього методу зазвичай спостерігається деяка затримка між моментом оновлення інформації в первинних системах і часом, коли ці зміни з'являються в кінцевому місці зберігання. Залежно від потреб бізнесу таке відставання може становити декілька секунд, годин або багато днів. Термін “*режим, наближений до реального часу*” часто використовується для описання кінцевих даних, оновлення яких відстає від джерела на декілька секунд, хвилин або годин. Дані, що не відстають від джерела, вважаються даними “*в режимі реального часу*”, але цього важко досягнути при використанні методу консолідації даних.

Перевагою консолідації даних є здійснення трансформації значних обсягів даних (реструктуризація, узгодження, очищення і/або агрегація) в процесі їх передачі від первинних систем до кінцевих місць зберігання. Деякі труднощі, пов'язані з цим підходом, — це значні обчислювальні ресурси, потрібні для підтримки процесу консолідації даних, а також істотні ресурси пам'яті, необхідні для підтримки кінцевого місця зберігання. Але з урахуванням постійного вдосконалення апаратних засобів це не проблема.

Консолідація даних — це основний підхід, який використовується програмними додатками сховищ даних для побудови і підтримки оперативних складів даних і корпоративних сховищ. Консолідація даних також може бути застосована для створення залежної вітрини даних, але в цьому випадку в процесі консолідації використовується тільки одне джерело даних (наприклад, корпоративне сховище). У середовищі сховищ даних однієї з найпоширеніших технологій підтримки консолідації є технологія ETL(витягання, перетворення і завантаження — Extract, Transform and Load). Ще одна поширена технологія консолідації даних — управління змістом корпорації (Enterprise Content Management — ECM). Більшість рішень ECM спрямовані на консолідацію і управління неструктурованими даними, такими як документи, звіти і web-сторінки.

Федералізація даних. Процес забезпечує єдину віртуальну картину одного або декількох первинних файлів даних. Якщо бізнес-додаток генерує запит до цієї віртуальної картини, то процесор федералізації даних витягує їх з відповідних первинних складів, інтегрує так, щоб вони відповідали віртуальній картині і вимогам запиту, і відправляє результати бізнес-додатку, від якого прийшов запит. Наголосимо, що процес федералізації даних завжди полягає у *витяганні* даних із первинних систем на підставі зовнішніх вимог. Всі необхідні перетворення даних

здійснюються при їх витяганні з первинних файлів. Інтеграція корпоративної інформації (EnterpriseInformationIntegration, Eli) — це приклад технології, яка підтримує федеральний підхід до інтеграції даних. Один із ключових елементів федеральної системи — це метадані, які використовуються процесором федералізації даних для доступу до первинних даних. У деяких випадках ці метадані можуть складатися виключно з визначень віртуальної картини, які ставляться у відповідність (“межуються”) первинним файлам. У більш передових рішеннях метадані також можуть містити детальну інформацію про кількість даних, що перебувають у первинних системах, а також про шляхи доступу до них. Така розширена інформація може допомогти федеральному рішенню оптимізувати доступ до первинних систем.

Вважається, що основна перевага федерального підходу — це забезпечення доступу до поточних даних і позбавлення від необхідності консолідувати первинні дані в новому сховищі даних. Але слід пам’ятати, що федералізація даних не пристосована для витягання і узгодження великих їхніх масивів або для тих додатків, де існують серйозні проблеми з якістю даних у первинних системах. Ще один істотний чинник — потенційний вплив на продуктивність і додаткові витрати на доступ до численних джерел даних під час виконання програми.

Федеральна архітектура дуже корисна для потужних транснаціональних корпорацій і є зручним підходом для підтримки балансу між необхідністю автономії місцевих підрозділів компанії та їх гнучкості, з одного боку, і стандартизації і централізованого контролю, які здійснює центральний офіс, — з іншого. При цьому федеральним сховищем може бути як єдине фізичне федеральне сховище, так і федерація дрібніших спеціалізованих сховищ даних.

Поширення даних. Додатки поширення даних здійснюють копіювання даних з одного місця в інше. Ці додатки зазвичай працюють в оперативному режимі і проводять переміщення даних до місць призначення, тобто залежать від певних подій. Оновлення в первинній системі можуть передаватися в кінцеву систему синхронно або асинхронно. Синхронна передача потребує, щоб оновлення в обох системах відбувалися під час однієї фізичної транзакції. Незалежно від використовуваного типу синхронізації, метод розповсюдження гарантує доставку даних до системи призначення. Така гарантія — це ключова відмітна ознака поширення даних. Більшість технологій синхронного поширення даних підтримують двосторонній обмін даними між первинними і кінцевими системами. Прикладами технологій, що підтримують поширення даних, є інтеграція корпоративних додатків (EnterpriseApplicationIntegration, EAI) і тиражування корпоративних даних (EnterpriseDataReplication, EDR).

Великою перевагою методу поширення даних є те, що він може бути

використаний для переміщення даних у режимі реального часу або близькому до нього. Інші переваги полягають у гарантованій доставці і двосторонньому поширенні даних. Метод поширення даних може також використовуватися для врівноваження робочого навантаження, створення резервних копій і відновлення даних, зокрема у разі надзвичайних ситуацій. Практичне застосування цього методу вирізняється чималою різноманітністю як у плані продуктивності, так і відносно можливостей реструктуризації й очищення даних. Деякі корпоративні продукти поширення даних можуть підтримувати переміщення і реструктуризацію крупних масивів даних, тоді як продукти EAI часто мають обмежені можливості пересування великої кількості даних і їх реструктуризації. Одна з причин такої відмінності — той факт, що в центрі архітектури тиражування корпоративних даних лежать дані, а в центрі технології EAI — повідомлення або транзакції.

Гібридний підхід. Методи, які використовуються додатками інтеграції даних, залежать як від потреб бізнесу, так і від технологічних вимог. Часто додаток інтеграції даних використовує так званий гібридний підхід, який включає декілька методів інтеграції. Прикладом такого підходу є інтеграція даних про клієнтів (CustomerDataIntegration, CDI), метою якої є забезпечення узгодженої картини інформації про клієнтів. Найпростіший підхід до CDI— це створення консолідованого сховища даних про клієнтів, яке містить дані, отримані з первинних систем. Відставання інформації в консолідованому сховищі залежатиме від режиму консолідації даних (оперативний або пакетний) і від частоти оновлення цієї інформації. Інший підхід до CDI— це федералізація даних, коли визначаються віртуальні бізнес-подання даних про клієнтів у первинних системах. Ці подання використовуються бізнес-додатками для доступу до поточної інформації про клієнтів у первинних системах. За федерального підходу також може використовуватися довідковий файл метаданих для зв'язку інформації про клієнтів на основі загальних ключових елементів.

Гібридний підхід, що використовує як консолідацію, так і федералізацію даних, також може бути. Загальні дані про клієнтів (ім'я, адреса та ін.) можуть бути консолідовані в одному сховищі, а дані, які стосуються до певного первинного додатку (наприклад, замовлення), можуть бути федералізовані. Такий гібридний підхід може бути розширений за рахунок поширення даних. Якщо клієнт оновлює своє ім'я і адресу під час транзакції в інтернет-магазині, то ці зміни можуть бути відправлені до консолідованого сховища даних, а звідти поширені в інші первинні системи, такі як база даних про клієнтів роздрібного магазину.

На сьогодні сформувалося два основні підходи до архітектури сховищ даних. Це так звана корпоративна інформаційна фабрика (CorporateInformationFactory, CIF) Білла Інмона і сховище даних з

архітектурою шини (DataWarehouseBus, BUS) Ральфа Кімболла. Розглянемо кожний із них докладніше.

Corporate Information Factory. На рис. 2.1 представлений підхід, що використовується в сховищах даних із архітектурою CIF. Колись цей підхід був відомий під назвою корпоративного сховища даних (EnterpriseDataWarehouse, EDW). Робота такого сховища починається зі скоординованого витягання даних із джерел. Після цього завантажується реляційна база даних із третьою нормальною формою, що містить атомарні дані. Отримане нормалізоване сховище використовується для того, щоб наповнити інформацією додаткові репозиторії презентаційних даних, тобто підготовлених для аналізу. Ці репозиторії, зокрема, включають спеціалізовані сховища для вивчення і “здобуття” даних (DataMining), а також вітрини даних.

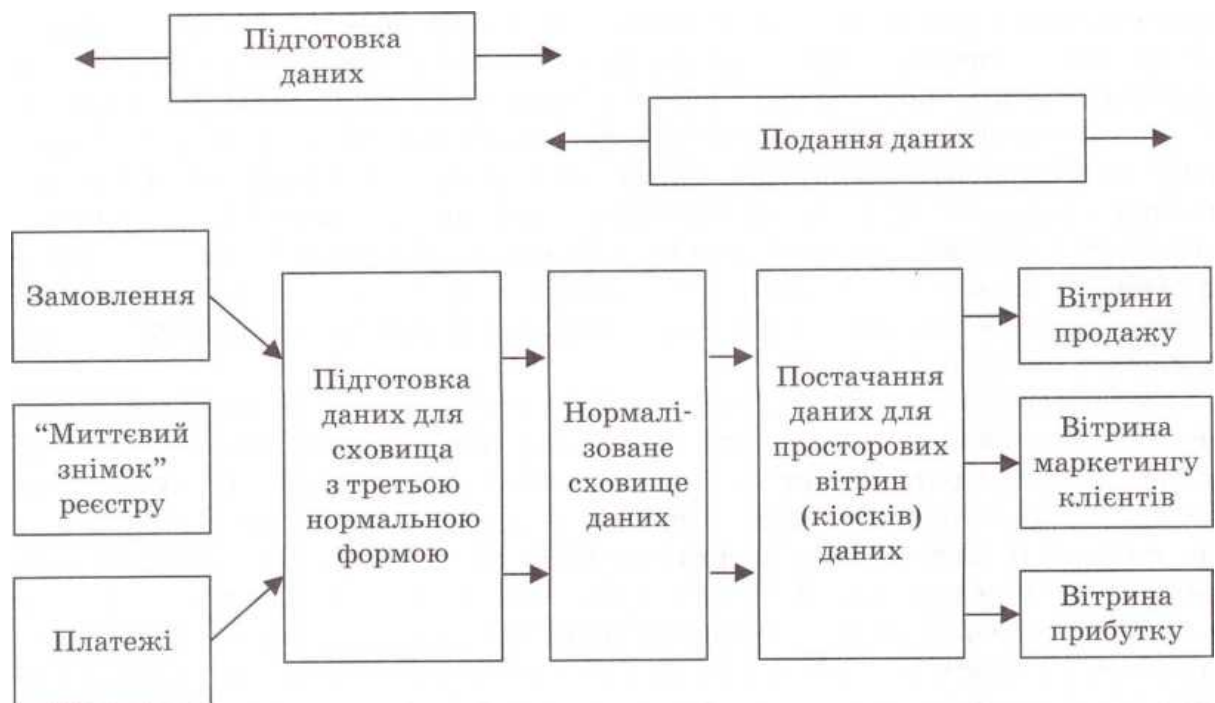


Рисунок 2.1 - Нормалізоване сховище даних із просторовими вітринами підсумкових даних(CIF)

За такого сценарію кінцеві вітрини даних створюються з метою обслуговування бізнес-відділів або для реалізації бізнес-функцій і використовують просторові моделі для структуризації сумарних даних. Атомарні дані залишаються доступними через нормалізоване сховище даних. Очевидно, що структура атомарних і сумарних даних за такого підходу істотно відрізняється. Відмітними характеристиками підходу Білла Інмона до архітектури сховищ даних можна назвати такі:

- використання реляційної моделі організації атомарних даних і просторової — для організації сумарних даних;

- використання ітеративного або “спірального” підходу при створенні великих сховищ даних, тобто “будівництво” сховища не відразу, а частинами. Це дає змогу за необхідності вносити зміни в невеликі блоки даних або програмних кодів і позбавляє від необхідності перепрограмувати значні обсяги даних у сховищі. Те ж саме можна сказати і про потенційні помилки: вони також будуть локалізовані в межах порівняльного невеликого масиву без ризику зіпсувати все сховище;

- використання третьої нормальної форми для організації атомарних даних, що забезпечує високий ступінь детальності інтегрованих даних і, відповідно, надає корпораціям широкі можливості для маніпулювання ними і зміни формату і способу подання даних у міру необхідності;

- сховище даних — проект корпоративного масштабу, що охоплює всі відділи і обслуговує потреби всіх користувачів корпорації;

- сховище даних — не механічна колекція вітрин даних, а фізично цілісний об’єкт.

DataWarehouseBus. На рис. 2.2 наведене схематичне зображення альтернативного підходу до архітектури сховищ даних, відомого як сховище з архітектурою шини, або підхід Ральфа Кімболла.



Рисунок 2.2 - Просторове сховище даних

У цій моделі первинні дані перетворюються на інформацію, придатну для використання, на етапі підготовки даних. При цьому обов’язково беруться до уваги вимоги до швидкості обробки інформації і якості даних. Як і в моделі Білла Інмона, підготовка даних починається зі скоординованого витягання даних із джерел. Низка операцій здійснюється

централізовано, наприклад підтримка і зберігання загальних довідкових даних, інші дії можуть бути розподіленими. Область уявлення просторово структурована, при цьому вона може бути централізованою або розподіленою.

Просторова модель сховища даних містить ту саму атомарну інформацію, що і нормалізована, але інформація структурована по-іншому з метою полегшення її використання і виконання запитів. Ця модель включає як атомарні дані, так і загальну інформацію (агрегати у пов'язаних таблицях або багато вимірних кубах) відповідно до вимог продуктивності або просторового розподілу даних. Запити в процесі виконання звертаються до все нижчого рівня деталізації без додаткового перепрограмування з боку користувачів або розробників додатка.

На відміну від підходу Білла Інмона, просторові моделі будуються для обслуговування бізнес-процесів (які, у свою чергу, пов'язані з бізнес-показниками), а не бізнес-відділами. Наприклад, дані про замовлення, які повинні бути доступні для загальнокорпоративного використання, вносяться до просторового сховища даних тільки один раз, на відміну від CIF-підходу, в якому їх довелося б тричі копіювати у вітрини даних відділів маркетингу, продажів і фінансів. Після того, як у сховищі з'являється інформація про основні бізнес-процеси, консолідовані просторові моделі можуть видавати їх перехресні характеристики. Матриця корпоративного сховища даних із архітектурою шини виявляє і підсилює зв'язки між показниками бізнес-процесів (фактами) і описовими атрибутами (вимірюваннями).

Підсумовуючи, можна назвати типові риси підходу Ральфа Кімболла:

- використання просторової моделі організації даних із архітектурою “зірка” (star scheme);
- використання дворівневої архітектури, яка включає стадію підготовки даних, не доступну для кінцевих користувачів, і сховище даних із архітектурою шини як таке. До складу останнього входять декілька вітрин атомарних даних, декілька вітрин агрегованих даних і персональна вітрина даних, але воно не містить одного фізично цілісного або централізованого сховища даних;
- сховищу даних із архітектурою шини властиві такі характеристики: воно просторове; включає як дані про транзакції, так і сумарні дані; включає вітрини даних, присвячені тільки одній предметній галузі або які мають тільки одну таблицю фактів (facttable); може містити безліч вітрин даних у межах однієї бази даних;
- сховище даних не є єдиним фізичним репозиторієм (на відміну від підходу Білла Інмона). Це “віртуальне” сховище, колекція вітрин даних, кожна з яких має архітектуру типу “зірка”.

Дотепер технології сховищ даних досягли рівня зрілості — і як дисципліна, і як ринок технологій. Попит на них високий як ніколи. Про

це, наприклад, свідчить той факт, що в 2007 р. ці засоби вперше потрапили в десятку головних пріоритетів директорів з інформаційних технологій (згідно з досвідом компанії GartnerGroupInc.). Багато корпорацій вже володіють інфраструктурою сховищ даних і нині займаються її удосконаленням і вирішенням проблем, з якими не впоралися на стадії впровадження. Частина корпорацій також прагне досягти подальших рівнів у розвитку цих засобів і акцентує увагу на 10 основних тенденціях, які сьогодні впливають на підхід корпорацій до впровадження сховищ даних. Ці тенденції включають як ті поліпшення, які компанії вносять до своєї стратегії та інфраструктури, так і нові технології та ініціативи, сприяючи подальшому розвитку сховищ даних.

Тенденція № 1: серйозне ставлення до якості даних. Мало хто займатиметься вирішенням проблеми якості даних тільки заради якості як такої. Так що ж примушує корпорації приймати конкретні заходи з поліпшення якості даних, а не просто говорити про це? По-перше, низька якість даних коштує грошей у тому сенсі, що призводить до зниження продуктивності, ухвалення неправильних бізнес-рішень і неможливості отримати бажаний результат, незважаючи на істотні інвестиції в корпоративні додатки. По-друге, низька якість даних може утруднити виконання вимог законодавства. За прогнозом компанії META Group щорічне зростання ринку програмного забезпечення і послуг у сфері якості даних становитиме 20—30 % аж до 2010 р. Цей прогноз збігається з вже наявними тенденціями: компанії дійсно мають намір робити конкретні дії з метою вирішення проблем якості даних. Враховуючи всю різноманітність засобів, представлених зараз на ринку, для будь-якої компанії дуже важливо правильно обрати такі методологію та інструменти, які нададуть можливість практично підійти до вирішення цих проблем.

Тенденція № 2: стандартизація і консолідація інфраструктури. Корпорації стали усвідомлювати проблему роз'єднаності своїх рішень у сфері сховищ даних зовсім нещодавно, буквально кілька років тому. По-перше, виявилось, що усунення дублюючих один одного інструментів сховищ даних або вітрин даних може знизити витрати на ліцензії і підтримку. Крім того, поліпшення доступу до інформації — це також істотна вигода від усунення незалежних структур даних, хоча вона і не піддається прямим кількісним оцінкам. Але здійснити реальну стандартизацію і консолідацію інфраструктури сховищ даних набагато складніше, ніж просто декларувати такий намір. Ця діяльність включає політичні й організаційні аспекти, які такі ж проблемні, як і технологічний бік процесу.

Тенденція № 3: використання зарубіжних трудових ресурсів у галузі сховищ даних. На відміну від американських компаній тенденція перенесення виконання проектів у галузі сховищ даних за межі країни з метою економії засобів не така актуальна для України, де заробітна плата і так менша, ніж у розвинутих країнах. Тому не зупинятимемося на цій

тенденції, зазначимо лише два основні моменти:

- 1) така тенденція справді спостерігається в США протягом декількох останніх років і активно обговорюється бізнес-аналітиками;
- 2) перші результати цього процесу достатньо суперечливі. Разом із економією засобів компанії зазнають труднощів, пов'язаних з якістю роботи, комунікаційними проблемами та ін. У результаті компанії не завжди отримують ту економію засобів, на яку вони розраховували.

Тенденція №4: стратегічний підхід до інформації. Поволі, але правильно корпорації починають розглядати інформацію як стратегічну складову бізнесу. Поки мало хто втілює ідею “інформація як капітал” на практиці, але в багатьох організаціях вже є співробітники, які усвідомлюють стратегічну цінність інформації. Менеджери вищої ланки також прихильно ставляться до цієї ідеї. Доказом такої тенденції є той факт, що сховища даних стають важливою частиною інших проектів із великими перспективами для бізнесу. Компанії можуть не впроваджувати систем інтелектуального аналізу даних у масштабах всієї корпорації, але вони використовують прийоми таких систем і сховища даних у ключових корпоративних проектах, які спрямовані на оптимізацію бізнес-процесів і зниження витрат.

Отже, з чого починається ставлення до інформації як до капіталу? Перший крок — це розробка інформаційної стратегії і архітектури, а також упровадження основоположних стандартів управління даними. Це потребує участі як ІТ-відділу, так і менеджерів, оскільки однаково важливі і технологічні, і організаційні, й політичні аспекти. ІТ-фахівці повинні працювати разом із менеджерами, щоб зрозуміти вимоги останніх і з'ясувати, які саме дані потрібні для вирішення основних питань бізнесу.

Тенденція № 5: виконання вимог законодавства як двигун розвитку сховищ даних. Для виконання законодавчих вимог необхідний доступ до даних. Як результат усвідомлення цього, у США зросли інвестиції в рішення, що надають доступ до інформації (або що забезпечують безпеку даних) і гарантують її точність. Необхідність виконання вимог законодавства стала основним стимулом для проектів, пов'язаних зі сховищами даних, позначивши реальні негативні наслідки відсутності доступу до якісних даних. Багато корпорацій можуть випробувати спокусу проігнорувати сховища даних і звернутися до своїх систем планування ресурсів підприємства (EnterpriseResourcePlanning, ERP) та фінансових систем, але прозорість і можливості контролю бізнесу, які надають рішення систем аналізу даних і сховищ даних, залишаються поза конкуренцією.

Тенденція № 6: подальший розвиток проблеми інтеграції корпоративних даних. Для передових корпорацій вже не стоїть питання, які технології обрати для інтеграції даних, — вони зробили свій вибір. Багато хто вибрав технології витягання, перетворення і завантаження

даних (Extraction, Transmission, Loading, ETL) як можливий спосіб для інтеграції корпоративних даних. Тепер перед цими корпораціями постало інше питання: як використовувати ті інструменти ETL, які вони вибрали, тобто проблема інтеграції даних піднімається на наступний рівень. Для повноцінної інтеграції даних організаціям необхідно здійснювати управління метаданими і довідковими даними, а також забезпечити якість даних. Фактично їм потрібний ще один інтегруючий засіб, який може поєднати вибрані ними інструменти і збільшити вартість кінцевої продукції компанії. Водночас багато корпорацій все ще перебувають у процесі пошуку відповідних інструментів для інтеграції даних, обираючи між ETL-технологіями, інтеграцією корпоративних додатків (EnterpriseApplicationIntegration, EAI) і web-сервісами. Але потрібно пам'ятати, що межі між цими технологіями стають все більш розпливчастими: у ETL-технологіях з'являються можливості інструментів EAI для роботи в режимі реального часу, а в засобах EAI розширюються можливості для перетворення даних. Навіть для тих корпорацій, які ще не визначилися з вибором інтеграційних технологій, буде корисно наперед продумати ті кроки, які доведеться зробити для впровадження вибраної технології та її успішного використання.

Тенденція № 7: навчання кінцевих користувачів. Рішення сховищ даних виявляться незатребуваними, якщо кінцевим користувачам не пояснити, які дані тепер їм доступні, і не переконати їх, що вони точні і повноцінні. Багато організацій вважають, що достатньо навчити кінцевих користувачів загальним прийомам роботи з цими інструментами. Але, на думку відомої аналітичної корпорації GartnerGroupInc., “набагато важливіше навчити користувачів інтелектуальному аналізу даних”. Користувач, який навчений працювати з інструментами, але не знає, як їх використовувати в контексті конкретного середовища сховища даних, не зможе отримати бажані аналітичні результати. Відповідно він або звернеться в IT-відділ, або взагалі відмовиться від використання цих засобів. Недостатнє визнання користувачами і усвідомлення компаніями цінності попередніх проєктів сховищ даних примушують багато організацій визнати значущість комплексних навчальних програм для кінцевих користувачів.

Тенденція № 8: управління довідковими даними. У кожній корпорації є набір даних, які забезпечують важливу інформацію, необхідну для ідентифікації і конкретного визначення ключових об'єктів, таких як споживачі, продукція, постачальники та ін. Ці дані називаються довідковими (masterdata), і зараз вони стають головною проблемою для багатьох організацій. Управління довідковими даними, на перший погляд, є простим. Справді, чи може бути складно для організації ідентифікувати своїх споживачів або визначити кожен вид своєї продукції? Як і у випадках, що належать до сфери інтелектуального аналізу даних і сховищ

даних, те, що здається зовсім простим, у реальності може виявитися складним. Поширення корпоративних додатків разом із переважанням у багатьох організаціях ізольованих структур даних призводить до того, що довідкові дані виявляються розсіяними по всій корпорації. Різні галузі бізнесу можуть по-різному визначати й ідентифікувати такі об'єкти, як “споживач” або “продукт”, і, можливо, зберігати їх довідкові дані в роз'єднаних базах даних. Таким чином, тенденція до інтеграції і оптимальної організації корпоративних систем зробила управління довідковими даними пріоритетним завданням. Але треба мати на увазі, що будь-яке технічне рішення, яке буде обране для цього завдання, повинне не тільки включати конкретні управлінські інструменти, а й брати до уваги організаційні та політичні аспекти, якимось: хто є “власником” цих довідкових даних і хто повинен визначати їх.

Тенденція № 9: поява нових учасників ринку сховищ даних. Збільшення ринку сховищ даних не пройшло непоміченим серед постачальників в інших галузях технології. Серед нових учасників цього ринку є як невеликі компанії, так і вже достатньо відомі фірми, але найбільш потужні з них — це постачальники систем планування ресурсів підприємства (ERP-системи) і систем управління відносинами з клієнтами (CustomerRelationshipManagement, CRM), наприклад такі американські компанії, як SAP, Oraclei Siebel. Ці компанії вбачають великі можливості у ринку сховищ даних, по-перше, оскільки неухильно зростає оборот цього ринку, а по-друге, тому що багато корпорацій виявили, що їх ERPі CRM-системи функціонують не зовсім так, як очікувалося, в сенсі можливостей доступу до даних та їх аналізу. Постачальники ERPі CRMсистем приходять на цей ринок, пропонуючи своїм клієнтам засоби інтелектуального аналізу даних і сховищ даних, щоб допомогти їм у отриманні необхідних даних, і статистика показує, що ці постачальники досягають успіху на нових теренах.

Тенденція № 10: використання сховищ даних як робочого інструменту. Компанії усвідомлюють стратегічну роль сховищ даних, але вони також хочуть використовувати інформацію, отриману за допомогою власних даних, для того, щоб ухвалювати і тактичні рішення. Наприклад, роздрібні компанії зацікавлені в тому, щоб зрозуміти, як вони можуть використовувати інформацію, отриману з ланцюга постачань, для ухвалення своєчасних рішень. Таким чином, сховища даних можуть відіграти роль як в оперативному функціонуванні корпорації, так і у визначенні стратегічного напрямку її розвитку.

2.3 Вітрини та кіоски даних

У найбільш загальному вигляді сховища даних можуть бути розподілені на два типи: корпоративні сховища даних

(EnterpriseDataWarehouses) і кіоски або вітрини, даних (DataMarts).

Корпоративні сховища даних містять інформацію, що стосується всієї корпорації і зібрана з безлічі оперативних джерел для консолідованого аналізу. Зазвичай такі сховища охоплюють цілу низку аспектів діяльності корпорації і використовуються для ухвалення як тактичних, так і стратегічних рішень. Корпоративне сховище містить детальну і узагальнену інформацію, його обсяг може становити від 50 Гбайт до одного або декількох терабайт. Вартість створення і підтримки корпоративних сховищ може бути дуже високою. Зазвичай їх створюють централізовані відділи інформаційних технологій, причому згори донизу, тобто спочатку проектується загальна схема, і тільки тоді починається заповнення даними. Такий процес може тривати декілька років.

Кіоски, або вітрини, даних містять підмножину корпоративних даних і будуються для відділів або підрозділів усередині організації (рис. 2.3), силами самого відділу і охоплюють конкретний аспект, який цікавить співробітників певного відділу. Кіоск даних може отримувати дані з корпоративного сховища (залежний кіоск) або, що більш поширене, дані можуть надходити безпосередньо з оперативних джерел (незалежний кіоск). Кіоски і сховища даних будуються за схожими принципами і використовують практично одні і ті самі технології.

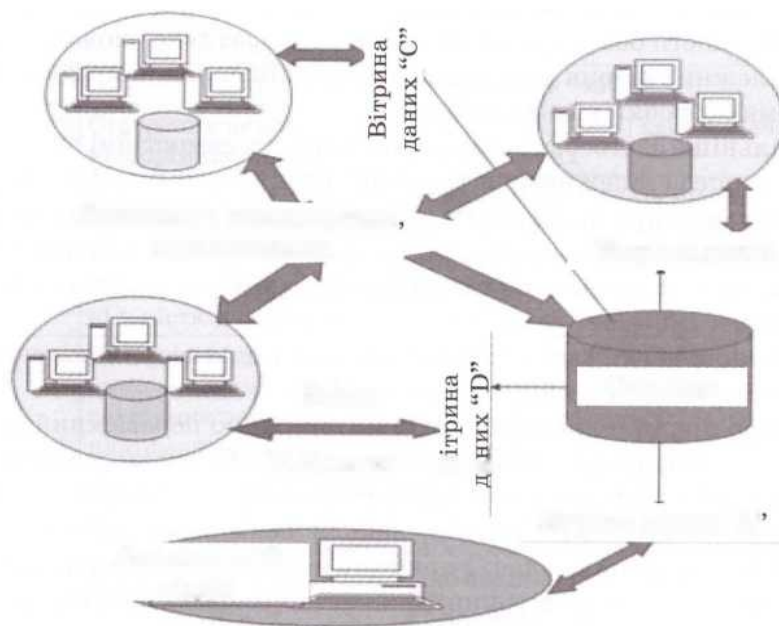


Рисунок 2.3 - Кіоски (вітрини) даних організації

Багато компаній, що усвідомлюють необхідність розробки корпоративного сховища даних, не в змозі впоратися з усіма завданнями виділення, стандартизації і об'єднання терабайт даних. Натомість вони вважають за краще будувати кіоски (або вітрини) даних (DataMarts) —

спеціалізовані сховища даних, присвячені тільки одному напрямку діяльності організації. Кіоск (вітрина) даних — це, найчастіше, найбільш керований різновид сховища даних. Його безперечний недолік полягає в тому, що без сховища даних, яке охоплювало б інформацію всього підприємства, неможливо порівнювати і аналізувати дані за всіма відділами і процесами. Фахівці компаній зрозуміли, що кіоски (вітрини) даних можуть прислужитися і навіть стати єдиною можливим рішенням для виконання термінових аналітичних завдань, але створення спеціалізованих кіосків без попередньої розробки корпоративної інфраструктури сховища даних може згодом призвести до великих ускладнень.

За класичним визначенням вітрина (кіоск) даних (DataMart) є підмножиною сховища даних, що відображає специфіку підрозділу (бізнес-об'єкт) і забезпечує підвищену продуктивність. Тобто вітрина є ланкою, на якій базується конкретна аналітична система для вирішення свого кола завдань. Проте можлива ситуація, коли деяка галузь діяльності підприємства практично не корелює з іншими, і можна побудувати відповідну вітрину даних автономно, без прив'язки до корпоративного сховища. Тоді вона поповнюватиметься даними безпосередньо з оперативних систем обробки транзакцій. Такі вітрини даних отримали назву незалежних на відміну від класичних залежних від сховища даних і поповнюваних із нього вітрин.

У низці випадків доцільно розгорнути вітрину (кіоск) даних замість повністю сформованого сховища. Вітрини даних накладають менші зобов'язання, вони дешевші, простіші в побудові і базуються на дешевших серверах, а не на мультипроцесорних комплексах. За такого підходу немає необхідності задіювати цілу інформаційну систему корпорації і підтримувати складні процедури синхронного оновлення вітрини даних при оновленні сховища. Але необхідно розуміти, що за такого підходу вітрини даних можуть перетворитися на цілі комплекси незалежних інформаційних баз даних, і природно буде поставлено завдання управління індивідуальними стратегіями пошуку, обслуговування і відновлення. З іншого боку, будувати єдине корпоративне сховище на основі множини незалежних вітрин даних значно вигідніше, ніж спираючись на розсіяні по системах обробки транзакцій дані.

То що доцільніше застосовувати: єдине сховище, самостійні вітрини (кіоски) даних, сховище із залежними вітринами або інші варіанти? Універсальної відповіді на питання про необхідність застосування того або іншого варіанта немає. У кожному випадку оптимальний варіант визначається вимогами бізнесу, інтенсивністю запитів, мережевою архітектурою, необхідною швидкістю реакції та іншими умовами.

Розглянемо аргументи “за” і “проти” використання вітрин (кіосків) даних відносно використання сховищ даних (табл. 2.2).

Говорячи про піврічні дослідження (з питання про переважний вибір

між кіосками даних і сховищами даних), аналітик Кевін Стрендж, керівник досліджень із програмних структур компанії GartnerGroup, зазначив, що кіоски не обов'язково мають невеликий розмір, деколи вони потребують розв'язання складних адміністративних задач, і причому не завжди обходяться дешевше, ніж сховища даних.

Таблиця 2.2 - Порівняльні аргументи застосування вітрин (кіосків) та сховищ даних

Критерії	Аргументи “за”
Вартість	Створення навіть декількох вітрин (кіосків) даних значно дешевше, ніж організація єдиного сховища даних
Терміни	На опис предметної області, взаємних зв'язків між даними, організацію сховища даних і розробку механізмів його поповнення може бути витрачено декілька років, тоді як опис будь-якого одного напряму діяльності підприємства без урахування різних зв'язків і з невеликою кількістю джерел надходження інформації потребує менше часу
Розміри	Оскільки вітрини (кіоски) даних зазвичай містять лише дані з певного кола питань і, отже, займають менше місця і потребують менше технічних ресурсів, то для них не так гостро стоїть питання апаратної платформи і вартості устаткування
Безпека	З вітринами (кіосками) даних зазвичай працює менша кількість користувачів, ніж зі сховищем даних. З'являється можливість контролю прав не тільки на рівні окремих таблиць і записів, а й на рівні доступу до всього додатка, що надійніше
	Аргументи “проти”
Дублювання даних	Різні вітрини даних можуть містити однакову інформацію, якщо цього потребують їх предметні області. Природно, дублювання інформації ставить перед користувачами і адміністраторами проблему синхронізації даних (тобто їх порівняння й уніфікації)
Розширення	Хоча виробники і стверджують, що вітрини (кіоски) даних поступово можна наростити до рівня сховищ даних (як модульні системи), на практиці це складно. Інакше вони коштували б стільки, скільки й самі сховища. Процес об'єднання незалежних (логічно і фізично) вітрин
Обмеженість	Вітрини (кіоски) даних задумані як склади даних, що містять інформацію з певної теми. Для великих компаній із широким колом вирішуваних задач і різноманітними інтересами вигідніше мати повноцінне сховище даних, оскільки воно зможе вміщувати всі необхідні для їх життєдіяльності відомості

Кіоски (вітрини) даних створюються спеціально під конкретний програмний додаток і вирішують специфічні задачі, скажімо, проводять аналіз, необхідний для роботи служби технічної підтримки користувачів, тоді як великі сховища даних не залежать від додатка.

Але кіоски даних можуть зберігати сотні гігабайт інформації і не обмежуються загальним стандартом у 50 Гбайт.

Кіоски добре підходять компаніям, які вимушені використовувати системи підтримки ухвалення рішення, але не мають достатнього досвіду для розробки повномасштабного сховища даних. За словами Стренджа, організації, що вже реалізували кіоски даних, поступово починають об'єднувати їх, і без сумніву, до кінця 2010 р. більшість організацій, які зараз створюють кіоски, прагнутимуть групувати їх у більш потужні структури.

Компанія “GartnerGroup” також дійшла висновку, що орієнтація сховищ даних на вебобмежувалася додатками для Intranetі не була розрахована на використання Internetчерез уразливість захисту і недостатній рівень продуктивності.

Прийоми моделювання кіосків (вітрин) даних відрізняються від прийомів моделювання сховищ даних через різні вимоги до структур даних. Якщо основним завданням сховища даних є зберігання консолідованої історичної інформації, то вітрина даних будується з урахуванням вимог із доступу до даних і подання інформації. Як правило, для моделювання вітрин (кіосків) даних використовуються такі типи моделей: схема “зірка” і “сніжинка”. Зупинимось докладніше на кожному з цих типів моделей.

Схема “зірка” — популярний тип моделі даних для вітрин даних. Ця модель характеризується наявністю таблиці фактів, оточеної пов'язаними з нею таблицями розмірностей. Запити до такої структури включають прості об'єднання таблиці фактів з кожною з таблиць розмірностей. Характеризується високою продуктивністю запитів. Проектується для виконання аналітичних запитів. Характеризується невеликим надлишком даних і високою порівняно з нормалізованими структурами продуктивністю. Деякі промислові СУБД та інструменти класу OLAP/Reportingуміють використовувати переваги схеми “зірка” для скорочення часу виконання запитів. Ця модель дає можливість отримати відповідь на широкий спектр аналітичних питань.

Розглянемо компоненти схеми “зірка”.

Розмірності. У технології багатовимірного моделювання розмірність — це аспект, у розрізі якого можна отримувати, фільтрувати, групувати і відображати інформацію про факти. Типові розмірності, що зустрічаються практично в будь-якій моделі:

- Клієнт;
- Продукт;
- Час;
- Географія;
- Співробітник.

Розмірності, як правило, мають багаторівневу ієрархічну структуру. Наприклад, розмірність ЧАС може мати таку структуру: РІК — КВАРТАЛ— МІСЯЦЬ — ДЕНЬ.

Факти. Зазвичай це числові величини, що зберігаються в таблиці фактів і є предметом аналізу. Приклади фактів: обсяг операцій, кількість проданих одиниць товару та ін. Факти мають низку властивостей, які охарактеризуємо.

1. Адитивні факти. Адитивність визначає можливість підсумування факту вздовж певної розмірності. Такі факти можна підсумовувати і групувати вздовж всієї розмірності на будь-яких рівнях ієрархії.

2. Напівадитивні факти. Це факти, які можна підсумовувати вздовж певної розмірності, і не можна вздовж інших. Прикладом може бути залишок на рахунку (або залишок товару на складі). Цю величину не можна підсумовувати вздовж розмірності ЧАС. Проте сума залишків за рахунками вздовж розмірності є предметом для аналізу.

Фахівці рекомендують моделювати напівадитивні факти так, щоб зробити їх більш адитивними. Наприклад, подати відсоток його складовими.

3. Неадитивні факти. Такі факти взагалі не можна підсумовувати. Приклад неадитивного факту — відношення (наприклад, виражене у відсотках).

Таблиці покриття. Використовуються з метою моделювання поєднання розмірностей, для яких відсутні факти. Наприклад, потрібно знайти кількість категорій продуктів, які сьогодні жодного разу не продавалися. Таблиця фактів продажу не може відповісти на це питання, оскільки вона реєструє лише факти продажу. Для того, щоб модель дала змогу відповісти на подібні питання, потрібна додаткова таблиця фактів (яка, по суті, не містить фактів).

Схема “сніжинка” використовується для нормалізації схеми “зірка”. Вона дещо скорочує надмірність у таблицях розмірності. Однією з переваг є швидше виконання запитів про структуру розмірності (запити типу “Вибрати всі рядки з таблиці розмірності на певному рівні”), які дуже часто виконуються при аналізі даних, і можуть затримувати хід аналізу. Проте основною відмінністю схеми “сніжинка” є не економія дискового простору, а можливість мати таблиці фактів із різним рівнем деталізації. Наприклад, фактичні дані на рівні дня, а планові — на рівні місяця.

Методика побудови вітрин (кіосків) даних із простої теоретичної дисципліни поступово перетворюється на складну науку, повну варіацій і напрямів. Якщо раніше було відомо лише про EDW (EnterpriseDataWarehouse), то тепер з’явилися вітрини даних, що поступово розвиваються (IncrementalArchitectedDataMart, ADM), розподілені вітрини (кіоски) (DistributedDataMart, DDM) та об’єднані вітрини даних (FederatedDataMart, FDM). Розглянемо деякі з цих нових напрямків.

Системи об’єднаних вітрин даних. У багатьох організаціях склалася практика реалізації численних сховищ даних. Хоча, за визначенням, є лише одне сховище даних, а всі інші об’єкти є його підмножиною або

вітринами (кіосками) даних, що поступово розвиваються, але не всі організації дотримуються цього правила. Отже, у компаніях є два, три, десять і навіть більше систем сховищ даних. Поширення сховищ даних привело до розвитку архітектури сховища даних підприємства, а саме до виникнення об'єднаних систем сховищ даних або вітрин (кіосків) даних.

Система об'єднаних вітрин даних характеризується спільним використанням загальних інформаційних ресурсів, усуваючи, таким чином, надмірність і гарантуючи достовірність інформації за всією організацією (рис. 2.4).

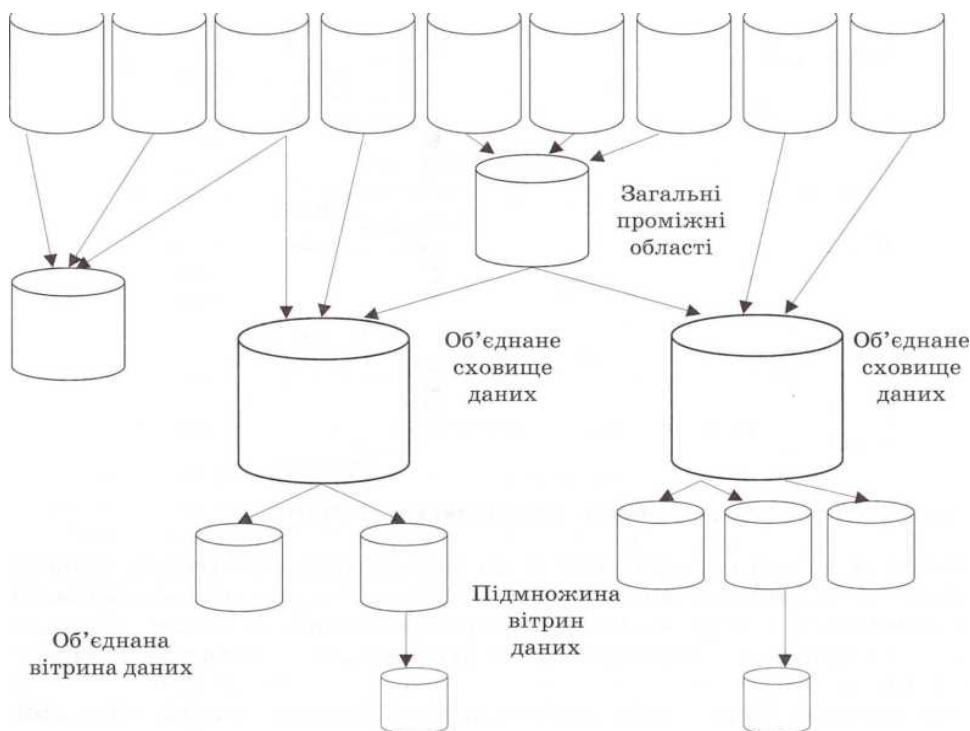


Рисунок 2.4 – Системи об'єднаних вітрин даних

Позитивними рисами об'єднаних вітрин даних є: загальна семантика бізнес-правил; один набір процесів витягання і бізнес-правил; децентралізовані ресурси і управління; паралельна розробка.

Недоліками такого архітектурного рішення є: необхідність координування робіт; труднощі в подоланні “політичних” моментів і вирішенні питань авторських прав; потрібна узгодженість серед різних відділів із питань архітектури, бізнес-правил і семантики; складне технічне середовище; наявність численних репозиторіїв метаданих.

Непроектовані вітрини даних. Виникнення непроектованих вітрин даних (Non-Architected Data Marts) пояснюється перш за все труднощами, пов'язаними з реалізацією систем EDW і FDW. “Брудні” та швидко отримувані набори даних не піддаються очищенню а, отже, не можуть використовуватися для подальшої інтеграції з будь-якими іншими

джерелами даних систем сховищ даних.

Дуже швидко вони перетворюються на застарілі системи, які лише додають проблем, а не вирішують їх. Для цих систем характерні численні процеси витягання, безліч бізнес-правил, невірогідність інформації (рис. 2.5).

Позитивними рисами непроєктованих вітрин даних є: висока продуктивність, низька вартість, недоліками — недостовірною інформацією; численні процеси витягання; велика кількість бізнес-правил; підвищена складність при інтеграції.

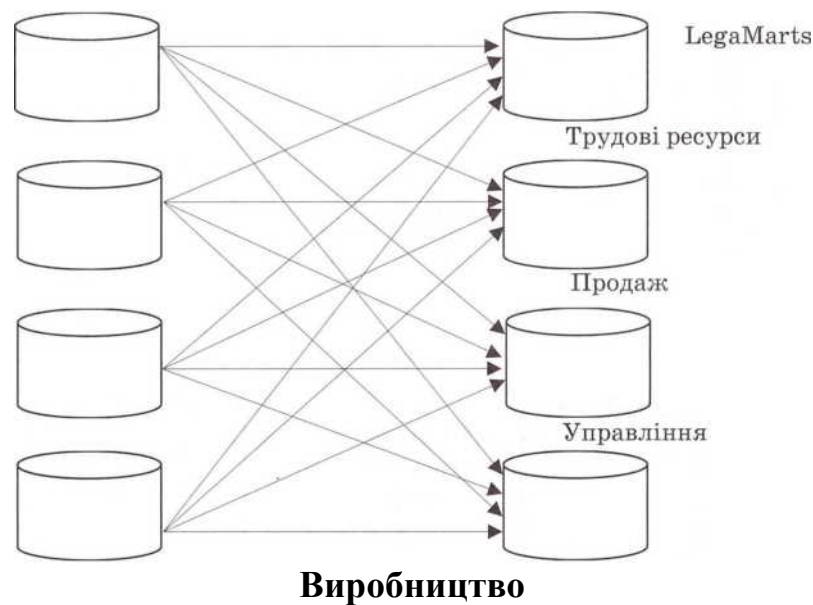


Рисунок 2.5 - Система непроєктованих вітрин даних (LegaMart)

Система вітрин (кіосків) даних, що поступово розвиваються. Ця архітектура є альтернативою сховища даних підприємства. Для наповнення таких вітрин зазвичай використовується інструментальний засіб класу підприємства, що реалізує стратегію “витягаєш один раз — наповнюєш багато” (рис. 2.6).

Переваги цих вітрин даних такі: єдиний набір процесів витягання; здійснимий масштаб. Недоліки: найбільш ефективні при використанні інструментального засобу класу підприємства; необхідність в архітектурі вітрин даних підприємства (EnterpriseDataMartArchitecture, EDMA).

Методика побудови вітрин (кіосків) даних виявилася напрямом ринку проєктів інтелектуального аналізу даних, що стрімко розвивається, швидко змінюється. Якщо раніше не було механізмів їх ефективного проєктування, і існував лише один спосіб їх створення, нині можна знайти значну кількість таких інструментів і низку технологій життєздатної архітектури таких систем. За умови вибору відповідної архітектури і

належного підходу до проекту можна побудувати систему сховища та вітрин даних, яка забезпечить не лише високе повернення інвестицій, а й значно підвищить ефективність функціонування всього підприємства.

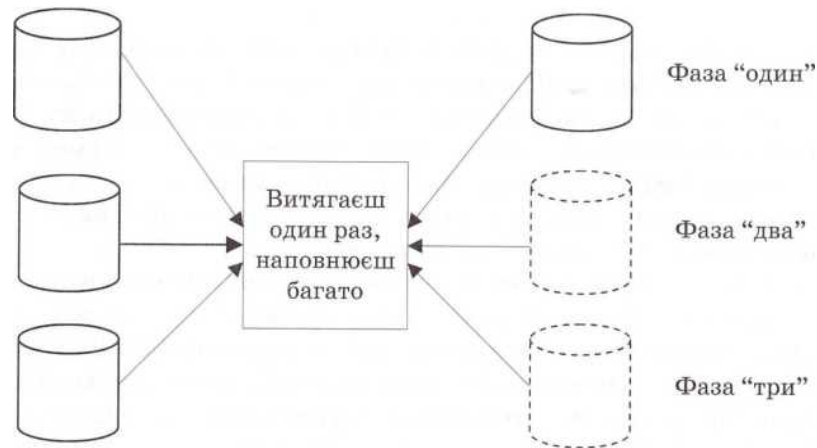


Рисунок 2.6. - Система вітрин даних, що поступово розвиваються

Декілька фірм пропонує системи побудови вітрин даних: Informatica (PowerMartSuite), Sagent Technology (Data Mart Solution) і Oracle (DataMartSuite). З метою ілюстрації процесу розробки вітрини даних розглянемо коротко склад і функціональність пакета DataMartSuite.

Пакет містить п'ять основних компонентів: Data Mart Designer, Data Mart Builder, Oracle7 Enterprise Server, Web Server і Discoverer 3.0. Data Mart Designer дає змогу описувати структуру вітрини і запам'ятовувати її в репозитарії. На виході DataMartDesigner видає опис мовою DDLSQL, який потім подається на вхід Oracle7 EnterpriseServer. У результаті створюється структура бази даних, що реалізовує вітрину даних. У процесі побудови вітрини користувач може застосовувати наявні описи структур або будувати вітрину "з нуля". Крім того, DataMartDesigner дає можливість будувати додатки для OracleWebServerна базі PL/SQL. DataMartBuilder витягує дані із зовнішніх джерел і заповнює вітрину. Цей компонент має наочний спеціалізований інтерфейс, що відображає потоки даних при заповненні сховища. DataMartBuilder здатний витягувати дані з реляційних СУБД і CSV-файлів. WebServer надає відкриту платформу для розробки web-додатків. Він включає WebRequestBroker (WRB), що реалізований на основі технології картриджів і дає змогу розробляти web-додатки, що вбудовуються в WebServer. Як засоби розробки можуть використовуватися Java, PL/SQL, LiveHTML, C і C++. Discoverer 3.0 — це засіб кінцевого користувача, що дає можливість генерувати звіти, а також виконувати деякі OLAP-операції з вітриною даних. Звіти, побудовані за допомогою Discoverer3.0, можна експортувати у форматі HTML, роблячи

їх доступними для web-браузерів. Discoverer3.0 також може створювати і підтримувати таблиці агрегованих даних. Крім того, DataMartSuite містить готовий додаток SalesAnalyzer.

2.4 OLAP-технологія

Вміння швидко і правильно ухвалювати рішення в сучасному бізнесі має принципове значення для досягнення успіху. Проте кількість інформації, що впливає на предмет рішення, інколи може бути величезною. Як вчинити в такій ситуації? Покласти на випадок або всетаки взятися за повномасштабний аналіз? Досвідчений керівник завжди вибере другий спосіб, тим більше, що сьогодні низка технологій, здатних спростити процес прийняття і моделювання рішень за великого обсягу “вхідної” інформації.

З метою спрощення роботи з багатоцільовими даними і умілого перетворення набору кількісних показників на якісні застосовується метод OLAP (On-Line Analytical Processing— оперативна аналітична обробка). Він, на відміну від інших способів автоматизації бізнес-діяльності, дає можливість отримати користувачеві “на виході” не готове чітко структуроване рішення, що видається після включення раніше налагодженого майстра обробки форм, а своєрідний матеріал для наочної і, якщо можна так сказати, творчої оцінки наявної ситуації. Тому сфера вживання OLAP-аналізу зазвичай обмежується менеджерським складом підприємств різних розмірів, якому доводиться часто займатися тактичними і стратегічними завданнями на зразок аналізу ключових показників діяльності і сценаріїв розвитку, маркетинговим і фінансово-економічним аналізом груп товарів або послуг, а також довгостроковим прогнозуванням роботи підприємства або його підрозділів. Для цього користувач OLAP-систем отримує потужний і гнучкий інструмент створення різних звітів за вибраними розрізами і напрямками. При цьому методика OLAP досконаліша за звичайні електронні таблиці, адже окрім простих функцій створення таблиць, графіків і діаграм, OLAP-системи дають можливість отримати узагальнені дані за самостійно обраними критеріями, вмить зануритися в деталі вибраних напрямів, відфільтрувати, відсортувати або відкинути непотрібні цифри або показники. Наприклад, якщо менеджеру продажів компанії потрібно отримати сезонні зведення динаміки продажу обраної категорії товарів, система запропонує йому різні дані про продаж за місяць, квартал, рік, а також знайде і проаналізує їх залежність від зазначених чинників, скажімо, часу проведення маркетингових акцій. Крім того, базуючись на одній лише статистиці продажу, OLAP-система може виявити ефективність роботи різних підрозділів компанії, у тому числі і в розрізі географічної ієрархії їх взаємодії. При цьому параметри, що характеризують успішність

підрозділів, обираються менеджером самостійно і у низці випадків можуть стати інструментом мотивації успішного персоналу.

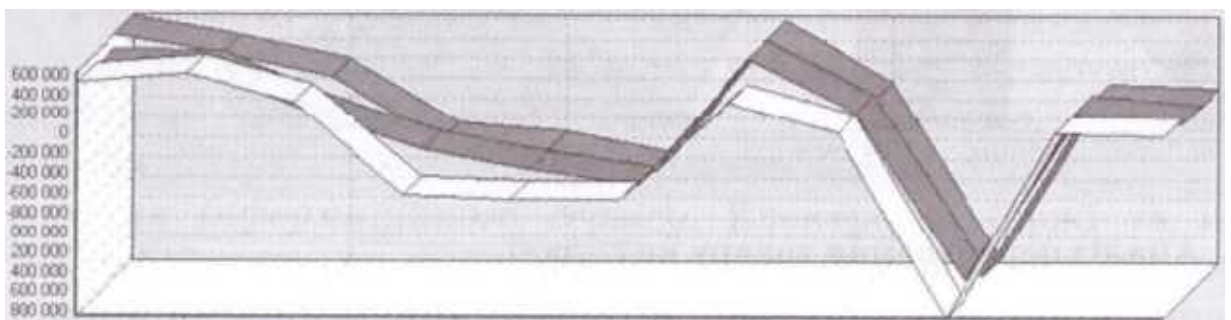
Щоб зрозуміти, як працює OLAP-система, досить розглянути її механізми. Найбільш показове поняття OLAP-технології — гіперкуб (метакуб), що є уявною фігурою в багатовимірному просторі, утвореному площинами даних, які важливі для діяльності підприємства. При цьому сама OLAP-система постає саме в ролі гіперкуба, здатного накопичувати всю інформацію, що цікавить керівника. Як ребра куба виступають різні категорії товарів або послуг, що пропонує компанія. Наприклад, ціна виробленого або конкурентного товару, компанії — учасники виробничого циклу, підрядники при організації послуг, обсяги продажу, географія самої компанії. Наголосимо, що градація різних осей квадрантів куба може мати різну структуру, а крім того, самі осі можуть бути взаємозалежними. Так, вісь часу може бути розбита за роками, кварталами, тижнями, а вісь доходів, або шкідливих викидів при виробництві — прологарифмована. Інформація, необхідна для аналізу в певний момент, вирізається з гіперкуба перетином площин даних, що використовуються при аналізі, ніби його шар або частина. При цьому розрізи можуть проходити як через весь куб, так і обмежуватися певними рамками і межами осей.

Технічно системи оперативного аналізу даних зазвичай функціонують у зв'язці зі сховищами та вітринами даних, а клієнтські OLAP-системи встановлюються на будь-яких призначених для користувача комп'ютерах корпоративної інформаційної системи. Рідше OLAP-модулі взаємодіють з іншими системами автоматизації, адже бази даних останніх досить часто мають дуже своєрідний вигляд і набір спеціальних показників. Утім, для сучасного українського підприємства характерна нетипова ситуація, коли є декілька систем автоматизації (для вирішення різних завдань) і, як наслідок, дані зберігаються розрізнено, а в результаті немає єдиного погляду на управлінську інформацію. Тому в процесі складання звіту беруть участь два фахівці — програміст, що забезпечує запити до баз даних, і економіст, що намагається за допомогою електронних таблиць звести ці дані в звіт, необхідний керівництву. Як показує практика, така модель взаємодії користувача звіту (керівника) і самих даних незмінно призводить до ефекту “зіпсованого телефону”, не кажучи вже про істотні витрати часу. В цьому аспекті використання OLAP-систем також виявляється вельми раціональним, адже використання декількох інформаційних систем незмінно приводить до “надлишку” даних, які можуть бути впорядковані OLAP-системою.

У чому ж відмінність OLAP-системи від сховища даних? З погляду користувача відповідь на це питання досить проста: в мірі предметної структурованості інформації. Працюючи з OLAP-додатком, користувач застосовує звичні економічні категорії і показники — види матеріалів і готової продукції, регіони продажу, обсяг реалізації, собівартість,

прибуток тощо. А для того, щоб сформувати будь-який, навіть дуже складний запит, користувачеві не доведеться вивчати SQL. При цьому відповідь на запит буде отримана протягом всього декількох секунд. Крім того, працюючи з OLAP-системою, економіст може користуватися такими звичними для себе інструментами, як електронні таблиці або спеціальні засоби побудови звітів. Таким чином, якщо сховище даних — в основному об'єкт уваги спеціаліста з інформаційних технологій, то OLAP без перебільшення можна назвати програмним засобом з арсеналу використовується по e-mail. Одержувач — співробітник планово-економічного відділу філії — відкриває куб в OLAPBrowser, аналізує, роздруковує і підшиває звіти. Для керівника філії створюється новий, узагальнений мікрокуб. У ньому додатково побудовані діаграми, що наочно показують структуру доходів і витрат, динаміку продажу за останній квартал у розрізі товарів та ін. Таку технологію, наприклад, використовує Укрсиббанк. За допомогою систем генерації і перегляду мікрокубів забезпечується інформаційна підтримка співробітників 20 філій банку, розподілених по всій Україні. Мікрокуби створюються в головному банку за даними бухгалтерського і управлінського обліку, які інтегровані в єдиному фінансовому сховищі.

Аналіз бюджетних даних. Для ведення фінансового планування і обліку фактичного виконання бюджетів підприємства застосовують прикладні модулі у складі комплексних ERP-систем Галактика, БЕСТ та ін., спеціалізовані програмні комплекси, наприклад Контур Корпорація, Бюджет холдингу, Інтальов, Бюджетне управління тощо. У всіх випадках для аналізу бюджетних планів, контролю виконання бюджету і аналізу відхилень фактичних показників від планових застосовується OLAP-система (рис. 2.7).



Риунок 2.7 - Аналіз бюджетних даних

Співробітники бюджетно-аналітичних і планово-економічних підрозділів випускають “план-факт”-звіти про виконання бюджетів за місяць, квартал, рік, аналізують бюджетні плани в розрізі центрів фінансової відповідальності і бізнес-напрямів, деталізують значення бюджетних статей. Наприклад, найбільший банк Казахстану “Банк

ТуранАлем” випускає всю бюджетну звітність у вигляді OLAP-звітів: фінансовий план та бюджет доходів і витрат у розрізі центрів фінансової відповідальності, бізнес-напрямів, банківських продуктів та ін.

Аналіз складських даних. Інформація про перебування і рух товарів на складі (товарні запаси, терміни зберігання товарів, постачальники і одержувачі продукції, накладні переміщення товарів) міститься в базі даних OLTP-модуля складського обліку. Аналіз цієї інформації дає відповіді на питання: “Скільки продукції було куплено замовником Івановим у третій декаді вересня?”, “Який оптимальний обсяг активних і резервних запасів за цією товарною позицією?”, “Чи існують сезонні коливання за цим типом товарів і яка їх амплітуда?”.

Деякі операторів працюють з одним програмним додатком, який розміщується на доступному сервері і безпосередньо звертається до бази даних системи складського обліку. За допомогою OLAP-звітів вони оперативнo контролюють поточну ситуацію: залишки товарних запасів у розрізі видів і партій товарів, термінів зберігання, стан відвантаження за одержувачами. Для віддалених підрозділів стан складу відбивається у мікрокубі, який розміщується на захищеній сторінці корпоративного веб-сайта.

Спільно з обліком складських даних розв’язується і задача обліку транспортної логістики. Для цього до облікової системи складу підключаються довідники транспорту, що обслуговує склади, ведеться облік його використання за кожним транспортним засобом. Наприклад, у ВАТ “Трьохгорна мануфактура” аналітична система дає змогу контролювати відвантаження товарів зі складів за артикулами.

Аналіз відвідуваності web-сайта. Web-сайт є серйозним маркетинговим інструментом для багатьох компаній. Аналіз поведінки відвідувачів сайта допомагає оцінити віддачу від маркетингових заходів і рекламних акцій, ефективність застосування on-line сервісів, інтерес до продуктів і послуг компанії та ін. Для аналізу використовуються дані log-файлів web-сервера, вивантажені в локальні або реляційні таблиці, або база даних сайта. Оскільки розміри таких баз, як правило, дуже великі, застосовується технологія мікрокубів. Генератор кубів щоночі за розкладом створює мікрокуб з актуальними показниками відвідуваності ресурсу. За його допомогою співробітники відділу маркетингу випускають звіти, які зберігаються в локальній мережі компанії. Запити до мікрокуба виконуються практично з “нульовим” часом очікування, мережевий трафік значно нижчий за рахунок стиснення інформації в мікрокубі. Таке рішення, наприклад, застосовується в компанії “1С: Парус” для аналізу використання співробітниками ресурсів Інтернету і аналізу дзвінків через мобільну мережу.

Публікація маркетингових досліджень. Маркетингові агентства збирають інформацію, обробляють її і продають результати зацікавленим

організаціям. Наприклад, певне агентство досліджує ринок продуктів глибокого заморожування і продає свої щомісячні звіти виробникам. Якщо в агентстві немає спеціалізованого ПЗ, результати досліджень можуть зводитися в Excel-файл. Цей файл засобами MSExcel можна зберігати в dbf-таблиці, з якої, у свою чергу, легко створити мікрокуб. Він продається виробникам продуктів за абонементом. Маркетологи підприємств отримують мікрокуб поштою і аналізують ринок, використовуючи програму OlapBrowser. Наприклад, англійська консалтингова компанія DecisionTreeConsulting виконує аналіз конкурентного середовища для брендів за замовленнями 200 найбільших компаній, таких як Sony, Toshiba, Panasonic, Nokia, Pioneer, Sanyo, Siemens, Phillips, HewlettPackard. Результати досліджень поставляються замовникам у вигляді мікрокубів. Індійська консалтингова компанія MARC, яка обслуговує фармацевтичні компанії Індії, також поширює результати досліджень замовникам у вигляді мікрокубів, надаючи їм віддалений доступ через Інтернет.

Створення інформаційного сервісу. Електронні біржі та інформаційні агентства публікують на своїх сайтах проспекти біржових індексів, котирування цінних паперів різних емітентів, рейтинги учасників фондового ринку за різними показниками та іншу інформацію у вигляді мікрокубів. Комерсанти знайомляться з актуальними даними з будь-якої точки земної кулі через Інтернет і за допомогою OlapBrowser проводять аналіз архівних і поточних біржових зведень та аналітичних довідок. Підтримка інформації в актуальному стані забезпечується за рахунок генерації мікрокубів за розкладом. Так, транснаціональна інвестиційна корпорація FidelityInvestment поставляє своїм клієнтам інформацію про цінні папери у вигляді мікрокубів через Інтернет. Світовий ринок OLAP-додатків стабільно зростає — на 16 % у 2008 р., і за прогнозами — приблизно на стільки ж у 2011—2013 роках. У абсолютних цифрах обсяг світового ринку OLAP, за даними “TheOLAPReport”, у 2007 р. становив 5,7 млрд дол., і як очікується, досягне показника в 7,31 млрд дол. до кінця 2011 р. Потрібно відзначити комерційну привабливість цього сектору адже він є нині одним із сегментів ІТ-ринку, що швидко розвиваються і зростають. Проте так було не завжди — показник зростання дуже коливався і демонстрував велику залежність від стану галузі в цілому — від більш ніж 40 % зростання (у 1995—1998 рр.) до близько 5 % у 2001—2002, у період кризи доткомів у США. Є всі підстави вважати, що і надалі ринок OLAP буде так само схильний до впливу загальногалузевих коливань кон’юнктури ІТ-ринку. Вочевидь, саме такі універсальні додатки надалі домінуватимуть на ринку (рис. 2.8).

Крім того, відзначимо такий важливий напрям, як інтеграція OLAP-додатків із системами управління базами даних і ERP-системами. Цей напрям інтеграції активно розвиває Microsoft— MicrosoftAnalysisServicesе додатком до MicrosoftSQLServer, а рішення для управління підприємством

— такі як Microsoft Dynamics AX (раніше — Microsoft Business Solutions Axapta) містять вбудовані OLAP-модулі. Схожої практики дотримуються й інші виробники СУБД і ERP-систем — Oracle і SAP. Вочевидь, що кількість і обсяг продажу інтегрованих OLAP-додатків з часом лише збільшуватиметься.

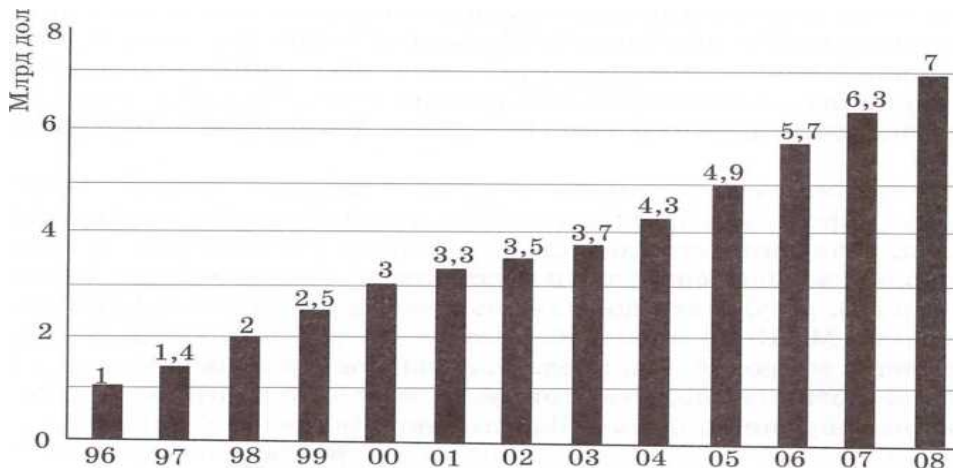


Рисунок 2.8 - Найбільші виробники OLAP-додатків

Одним із важливих чинників розвитку інтелектуального аналізу даних у сучасному інформаційному світі є рух OpenSource. Все більше компаній або використовують продукти з відкритими вихідними кодами, або самі беруть участь і підтримують розвиток таких програмних додатків. Майже для кожного типу програмних продуктів існують їх аналоги в сегменті OpenSource, і OLAP-додатки не є винятком. Вже зараз цілий набір додатків, що реалізують всі функції бізнес-аналітики (такі як Pentaho OpenBI Suite — повноцінна платформа для корпоративної системи аналізу даних, Palo-Server та ін.). Головна їх перевага можливість змінити додаток так, щоб він найбільше підходив під конкретні умови роботи, замість того, щоб змінювати структуру ухвалення рішень, що склалася. Це може залучити великі компанії, що використовують OLAP-додатки.

Також наслідком виникнення OpenSource OLAP-додатків стане їх поширення невеликі компанії, які до цього не застосовували дорогі OLAP-продукти через їх ціни, отримують таку можливість. Внаслідок цього можна очікувати появи ринку послуг з установки і підтримки OpenSource OLAP-додатків, де основними споживачами будуть компанії малого і середнього бізнесу.

Ринок OLAP-додатків тісно пов'язаний із ринком консалтингових послуг. Навчання, впровадження, підтримка є природним доповненням до програмного забезпечення для аналізу даних. Вже зараз на українському ринку успішно діють компанії, які пропонують комплексні рішення з консалтингу, навчання і впровадження цілого спектра продуктів. Проте до

того моменту, коли ринок OLAP неможливо буде відокремити від ринку консалтингових послуг, ще далеко. У довгостроковій перспективі компанії-виробники продовжуватимуть орієнтуватися на виробництво програмного забезпечення, залишаючи надання послуг партнерам.

Враховуючи названі чинники, що впливають на розвиток ринку OLAP-до- датків, можна передбачити, що в середньостроковій перспективі ринок зростатиме досить швидкими темпами, але конкуренція при цьому скорочуватиметься внаслідок процесу злиття компаній, що продовжується. Очікується більше комплексних рішень для аналізу даних і модулів для систем управління базами даних і ERP-систем, що реалізують функціональність OLAP. Ціна на OLAP-додатки знижуватиметься — як через тиск із боку лідерів ринку, так і через поширення OpenSourceOLAP-продуктів. Водночас потреба в послугах із впровадження і супроводу OLAP-систем збільшиться, що приведе до збільшення кількості спеціалізованих консалтингових компаній.

Ідея обробки багатовимірних даних не нова. Фактично вона бере початок з 1962 р., коли Кен Айверсон опублікував книгу “Мова програмування” (“A Programming Language”, APL). Вперше реалізація APL на практиці була здійснена в кінці 60-х років компанією IBM. APL — це математична мова з багатовимірними змінними і витонченими, хоч і досить абстрактними, операторами. Вона призначалася більше для описування багатовимірних перетворень, ніж для використання як практичної мови програмування. Так, у ній не приділялася увага таким простим питанням, як робота з файлами або виведення на друк. У дуже стислій нотації мови використовувалися грецькі символи. Насправді, тексти програм виходили вельми компактними. Вона стала відомою як “мова тільки для написання”, тому що було набагато легше переписати заново програму, ніж виправити раніше збережений текст. APL поглинала машинні ресурси і вимагала великих витрат. Програми виконувалися повільно і коштували дуже дорого. Проте, незважаючи на невдалий початок, APL не була відкинута. Вона використовувалася в багатьох ділових додатках 70-х, 80-х років, які функціонально схожі на сьогоденні OLAP-системи. Так, IBM розробила операційну систему для APL, названу VSPC, і деякі люди вважали її ідеальним середовищем для персонального використання задовго до появи електронних таблиць. У 80-х роках APL стала доступною на персональних машинах, але не отримала ринкового застосування. Альтернативою було програмування багатовимірних додатків із використанням масивів в інших мовах. Це було дуже важким завданням навіть для професійних програмістів, тому кінцевим користувачам залишалося чекати наступного покоління багатовимірних програмних продуктів.

У 1970 р. вперше з’явився прикладний багатовимірний програмний продукт, що використовувався у навчальних цілях — Express. Він у

повністю переписаному вигляді застосовується в сучасних OLAP-додатках, проте оригінальні концепції 70-х років залишилися далеко позаду. Сьогодні Express залишається однією з найбільш популярних OLAP-технологій, і компанії Oracle вдається підтримувати його на рівні сучасних вимог разом із багатьма новими продуктами з архітектурою “клієнт — сервер”.

Більше багатомірних продуктів виникло у 80-х роках. На початку десятиліття з'явився Stratagem, в новому обличчі — Acumate, який просувався на ринку до середини 90-х, але сьогодні, на відміну від Express, використовується дуже обмежено.

ComshareSystemW був багатомірним продуктом іншого стилю. Представлений у 1981 р., він першим використовував ідею гіперкуба і був більшою мірою орієнтований на кінцевого користувача в розробці фінансових додатків. Цей продукт привніс багато концепцій, які, правда, ще добре не пропрацювали: типу непроцедурних правил, повноекранного перегляду багатомірних даних, редагування даних, інтеграції з реляційними даними (у пакетному режимі). Проте ComshareSystemW був достатньо важкий для апаратного забезпечення того часу і менш програмований порівняно з іншими продуктами а, відповідно, був менш популярним у середовищі професіоналів. Він також ще використовується, але продається рідше, оскільки не має тих поліпшень, які очікувалися. В кінці 80-х Comshare випустив у середовищі DOS, а пізніше для Windows, продукт під назвою CommanderPrism, який використовував ті самі концепції, що були закладені в SystemW. Essbase, продукти компанії HyperionSolution, хоч і не є прямим нащадком SystemW, був, очевидно, під впливом його рішень за своєю орієнтацією на фінансові додатки і організацією гіперкуба з повними попередніми обчисленнями.

Іншим новаторським продуктом на початку 80-х був Metaphor. Він розроблявся для професійних маркетологів. Цей пакет також запропонував багато нових концепцій, які стали популярними тільки в 90-х роках, наприклад: клієнт-серверне обчислення, багатомірна обробка реляційних даних, розрахований на багато користувачів режим і об'єктна орієнтована розробка додатків. На жаль, стандартні персональні комп'ютери не забезпечували тих характеристик, які вимагав Metaphor, і постачальники були вимушені розробляти власні персональні машини і мережі. Надалі Metaphor став працювати вдало і на серійних персональних машинах, проте він ніколи не використовував стандартний графічний інтерфейс користувача (GUI).

У кінці 80-х серед інструментів кінцевого користувача для аналізу даних стали домінувати електронні таблиці. Перша багатомірна електронна таблиця з'явилася у вигляді Compete. Він просувався на ринок як дуже дорогий продукт для фахівців, але постачальники не забезпечили можливість охоплення ринку цим продуктом, і компанія

ComputerAssociates придбала права на нього разом з іншими продуктами класу “spreadsheet” (електронні таблиці), включаючи Supercalcі 20/20. Основним ефектом від придбання Competeкомпанією СА було різке зниження ціни і зняття захисту від копіювання, що, природно, сприяло його розповсюдженню. Проте він ще не був вдалим. Протягом декількох років Competez рідка можна було зустріти у вигляді навантаження в деяких комплектах постачання. Пізніше Competeбув покладений в основу Supercalc5, але багатовимірний аспект його не просувався.

Компанія Lotusбула наступною, хто спробував увійти на ринок багатовимірних електронних таблиць з продуктом Improv. Це гарантувало, що продаж 1—2—3 не знизиться, але коли Improvз часом був випущений під Windows, Excelвже став настільки серйозним конкурентом, що продаж Improvне вніс помітних змін у розподілі ринку. Lotus, подібно СА з Compete, знизила ціну на Improv, проте і цього було замало для просування на ринку, і нові розробки в цій галузі не здійснювалися. Виявилось, що користувачі персональних комп’ютерів надають перевагу електронним таблицям в оригінальній версії 1—2—3 і не цікавляться новими багатовимірними можливостями, якщо вони не повністю сумісні з їх старими таблицями. Так само концепція невеликих багатовимірних настільних електронних таблиць, пропонованих як продуктивний інструмент для персональних додатків, насправді виявилася незручною і не прижилася на практиці.

Компанія Microsoftпішла цим шляхом, додавши PivotTablesдо Excel. Хоча небагато користувачів Excelотримали вигоду від використання цієї можливості, це, ймовірно, єдиний факт широкого застосування можливостей багатовимірного аналізу просто тому, що в світі дуже багато користувачів Excel. Excel2000 містить витонченішу версію PivotTables, призначену для використання і як настільний інструмент OLAPі як клієнтська частина для взаємодії з MicrosoftOLAPServices. Проте можливості OLAPв Excel2000 не є базовими, провідними, вони вбудовані як додаткова, другорядна можливість.

Наприкінці 80-х років фірма Sinperувійшла до світу багатовимірних електронних таблиць спочатку з власного електронною таблицею для DOS, а потім приєднаною до версії 1—2—3 для DOS. Перетворений на продукт TM/1, він увійшов до ери Windowsяк сервер баз даних у форматі Excelі 1—2—3. Трохи пізніше Arborзробив аналогічну річ, хоча його новий Essbasemіг працювати тільки в режимі “клієнт — сервер”, тоді як продукт фірми Sinperміг так само працювати на локальному комп’ютері. Цей підхід привніс мультирозмірність в електронні таблиці, які є такими популярними серед користувачів. Традиційні постачальники власних інструментів подання даних кінцевому користувачеві пішли цим шляхом, і такі продукти, як DSSServer, Express, Holos, Gentia, Mineshare, Powerplay, Metacubeі Whitelightтепер з гордістю пропонують високоінтегрований

доступ до електронних таблиць у своїх серверах додатків. За іронією долі за перші шість місяців існування MicrosoftOLAPServicesбув одним з декількох OLAP-серверів, що не мають клієнтської частини у вигляді електронної таблиці. Пропозиції компанії Microsoftз'явилися тільки в червні 1999 р. в Excel2000. Проте OLAP@Work, вбудований в Excel, заповнив цю прогалину, і поки що має набагато кращі експлуатаційні характеристики, ніж власний інтерфейс Excelкомпанії Microsoft.

Деякі користувачі потребують для своїх багатовимірних додатків можливості обробки дуже великих багатовимірних баз даних. І реляційні OLAP- інструменти розвиваються в цьому напрямі, відгукуючись на ці потреби. Вони надають звичайні засоби проглядання багатовимірних даних, а іноді включають інтерфейс кінцевого користувача у вигляді електронної таблиці, навіть якщо всі дані зберігаються в реляційних СУБД. Такі засоби є дуже дорогими для користувача, вони менш продуктивні, ніж спеціалізовані інструменти багатовимірного аналізу, але вони забезпечують цю, таку популярну форму аналізу даних, навіть якщо останні зберігаються не у вигляді багатовимірних структур.

Інші постачальники розвивають те, що сьогодні називається настільним OLAP: невеликі куби, що генеруються з великих баз даних і потім завантажуються в персональний комп'ютер для обробки. Вони справді досягають великого успіху. А коли постачальник продає обидві можливості: і інструмент формування реляційних запитів, і інструмент багатовимірного аналізу й формування звітів, то досягає більшого успіху у кінцевих користувачів, ніж в інших випадках.

2.5 Основні архітектури OLAP-систем

Системи інтелектуального аналізу даних зазвичай володіють засобами надання користувачеві агрегатних даних для різних вибірок з початкового набору в зручному для сприйняття й аналізу вигляді. Як правило, такі агрегатні функції утворюють багатовимірний (і, отже, нереляційний) набір даних (часто він має назву гіперкуб або метакуб), осі якого містять параметри, а клітинки — залежні від них агрегатні дані — причому зберігатися такі дані можуть і в реляційних таблицях, але в цьому випадку йдеться про логічну організацію даних, а не про фізичну реалізацію їх зберігання. Уздовж кожної осі дані можуть бути організовані у вигляді ієрархії, що представляє різні рівні їх деталізації. Завдяки такій моделі даних користувачі можуть формулювати складні запити, генерувати звіти, отримувати підмножини даних.

Технологія комплексного багатовимірного аналізу даних отримала назву OLAP (On-LineAnalyticalProcessing). OLAP — це ключовий компонент організації сховищ даних. Концепцію OLAPописав у 1993 р. Едгар Кодд, відомий дослідник баз даних і автор реляційної моделі даних.

У 1995 р. на основі вимог, викладених Коддом, був сформульований так званий тест FASMI (FastAnalysisofSharedMultidimensionalInformation — швидкий аналіз розподіленої багатовимірної інформації), що включає такі вимоги до додатків для багатовимірного аналізу:

- Fast(швидкий). Надання користувачеві результатів аналізу за прийнятний час (зазвичай не більше 5 с), нехай навіть за рахунок менш детального аналізу;
- Analysis(аналіз). Можливість здійснення будь-якого логічного і статистичного аналізу, характерного для цього додатку, і його збереження в доступному для кінцевого користувача вигляді;
- Shared(розподілений доступ). Розрахований на багато користувачів доступ до даних із підтримкою відповідних механізмів блокувань і засобів авторизованого доступу;
- Multidimensional (багатовимірність). Багатовимірне концептуальне подання даних, включаючи повну підтримку для ієрархій і множинних ієрархій (це — ключова вимога OLAP);
- Information (інформація). Можливість звертатися до будь-якої потрібної інформації незалежно від її обсягу і місця зберігання.

Слід зазначити, що OLAP-функціональність може бути реалізована різними способами, починаючи з простих засобів аналізу даних в офісних додатках і закінчуючи розподіленими аналітичними системами, заснованими на серверних продуктах. Але перш ніж говорити про різні реалізації цієї функціональності, розглянемо, що ж є кубами OLAP з логічного погляду. Для ілюстрації принципів OLAPскористаємося базою даних Northwind, що входить у комплекти постачання MicrosoftSQLServer і є типовою базою даних, яка зберігає відомості про торгові операції компанії, що займаються оптовими постачаннями продовольства. До таких даних належать відомості про постачальників, клієнтів, компанії, що здійснюють доставку, список товарів, що постачаються та їх категорій, дані про замовлення і замовлені товари, список співробітників компанії.

Для розгляду концепції OLAPскористаємося представленням Invoicesі таблицями Productsі Categoriesз бази даних Northwind, створивши запит, у результаті якого отримаємо докладні відомості про всі замовлені товари і виписані рахунки:

- SELECT dbo.Invoices.Country,
- dbo.Invoices.City,
- dbo.Invoices.CustomerName,
- dbo.Invoices.Salesperson,
- dbo.Invoices.OrderDate,
- dbo.Categories.CategoryName,
- dbo.Invoices.ProductName,
- dbo.Invoices.ShipperName,
- dbo.Invoices.ExtendedPrice


```

— FROM dbo.Products INNER JOIN
—   dbo.Categories      ON      dbo.Products.CategoryID      =
dbo.Categories.CategoryID INNER JOIN
  dbo.Invoices ON dbo.Products.ProductID = dbo.Invoices.ProductID

```

Цей запит звертається до представлення Invoices, що містить відомості про всі виписані рахунки, а також до таблиць Categories і Products, що містить дані про категорії продуктів, які замовлялися, і про самі продукти відповідно. У результаті цього запиту отримаємо набір даних про замовлення, що включає категорію і найменування замовленого товару, дату розміщення замовлення, ім'я співробітника, що виписав рахунок, місто, країну і назву компанії-замовника, а також найменування компанії, що відповідає за доставку. Для зручності збережемо цей запит у вигляді уявлення, назвавши його Invoices!..

Яка сумарна вартість замовлень, зроблених клієнтами з Франції у 1997 р. і доставлених компанією SpeedyExpress?

Переведемо ці запитання в запити мовою SQL(табл. 2.3).

Таблиця 2.3 - Трансформація запитання у запити мовою SQL

Питання	SQL-запит
Яка сумарна вартість замовлень, зроблених клієнтами з Франції?	SELECT SUM (ExtendedPrice) FROM invoices 1 WHERE Country='France'
Яка сумарна вартість замовлень, зроблених клієнтами з Франції і доставлених компанією Speedy Express?	SELECT SUM (ExtendedPrice) FROM invoices1 WHERE Country='France' AND ShipperName='Speedy Express'
Яка сумарна вартість замовлень, зроблених клієнтами з Франції в 1996 р. і доставлених компанією Speedy Express?	SELECT SUM (ExtendedPrice) FROM Ord_ pint WHERE CompanyName='Speedy Express' AND OrderDate BETWEEN 'December 31, 1995 ' AND 'April 1, 1996 ' AND ShipperName='Speedy Express'

Результатом будь-якого з перерахованих запитів є число. Якщо в першому із запитів замінити параметр 'France' на 'Austria' або на назву іншої країни, можна знову виконати цей запит і отримати інше число. Виконавши цю процедуру зі всіма країнами, отримаємо набір даних, поданий у табл. 2.4.

Звернемося до другого із зазначених запитів, який містить дві умови в пропозиції WHERE. Якщо виконувати цей запит, підставляючи в нього всі можливі значення параметрів Country і ShipperName, отримаємо такий двомірний набір даних (табл. 2.5).

Такий набір даних називається зведеною таблицею (Pivot Table), або крос- таблицею (Cross Table). Створювати подібні таблиці дозволяють багато електронних таблиць і настільні СУБД — від Paradox для DOS до

Microsoft Excel.

Таблиця 2.4 - Результати запиту, який містить кілька умов у пропозиції

Country	ShipperName		
	Federal	Speedy	United
Argentina	1 210,30	1 816,20	5 092,60
Austria	40 870,77	41004,13	46 128,93
Belgium	11393,30	4 717,56	17 713,99
Brazil	16 514,56	35 398,14	55 013,08
Canada	19 598,78	5 440,42	25 157,08
Denmark	18 295,30	6 573,97	7 791,74
Finland	4 889,84	5 966,21	7 954,00
France	28 737,23	21140,18	31480,90
Germany	53 474,88	94 847,12	81962,58
		...	

Третій з розглянутих запитів має три параметри в умові WHERE. Варіюючи їх, отримаємо тримірний набір даних. Клітинки куба містять агрегатні дані, що відповідають значенням параметрів запиту, які знаходяться на осях куба у пропозиції WHERE. Можна отримати набір двомірних таблиць за допомогою перетину куба площинами, паралельними його граням (для їх позначення використовують терміни CrossSections і Slices).

Якщо в пропозиції WHERE міститься чотири або більше параметрів, результуючий набір значень (також званий OLAP-кубом) може бути чотиривимірним, п'ятивимірним та ін.

Таблиця 2.5 - Результати запитів у вигляді чисел

Country	SUM (ExtendedPrice)
Argentina	7327,3
Austria	110788,4
Belgium	28491,65
Brazil	97407,74
Canada	46190,1
Denmark	28392,32
Finland	15296,35
France	69185,48
Germany	209373,6

Розглянемо деякі ключові терміни і поняття, що використовуються при багатовимірному аналізі даних.

Разом із сумами в клітинках OLAP-куба можуть міститися результати виконання інших агрегатних функцій мови SQL, таких як MIN, MAX, AVG, COUNT, а в деяких випадках — й інших (дисперсії, середньоквадратичного відхилення тощо). Для опису значень даних у комірках використовується термін Summary(у загальному випадку в одному кубі їх може бути декілька), для позначення початкових даних, на основі яких вони обчислюються, — термін Measure, а для позначення параметрів запитів — термін Dimension(що перекладається як “вимірювання”, коли йдеться про OLAP-куби, і як “розмірність”, коли йдеться про сховища даних). Значення, що відкладаються на осях, називаються членами вимірювань (Members).

Говорячи про вимірювання, згадаємо, що значення, які наносяться на осі, можуть мати різні рівні деталізації. Наприклад, нас може цікавити сумарна вартість замовлень, зроблених клієнтами в різних країнах, або сумарна вартість замовлень, зроблених іногородніми клієнтами чи навіть окремими клієнтами. Природно, результуючий набір агрегатних даних в другому і третьому випадках буде детальнішим, ніж у першому. Зазначимо, що можливість одержання агрегатних даних із різним ступенем деталізації відповідає одній з вимог, що висувуються до сховищ даних, — вимозі доступності різних зрізів даних для порівняння й аналізу.

Оскільки в розглянутому прикладі в загальному випадку в кожній країні може бути декілька міст, а в місті — декілька клієнтів, можна говорити про ієрархії значень у вимірюваннях. У цьому випадку на першому рівні ієрархії розташовуються країни, на другому — міста, а на третьому — клієнти.

Відзначимо, що ієрархії можуть бути збалансованими (balanced), як, наприклад, ієрархія, а також ієрархії, засновані на даних типу “дата — час”, і незбалансованими (unbalanced). Типовий приклад незбалансованої ієрархії — ієрархія типу “начальник — підлеглий (її можна побудувати, наприклад, використовуючи значення поля Salespersonпочаткового набору даних із розглянутого вище прикладу)”, подана на рис. 2.9 Іноді для таких ієрархій використовується термін Parent-childhierarchy.

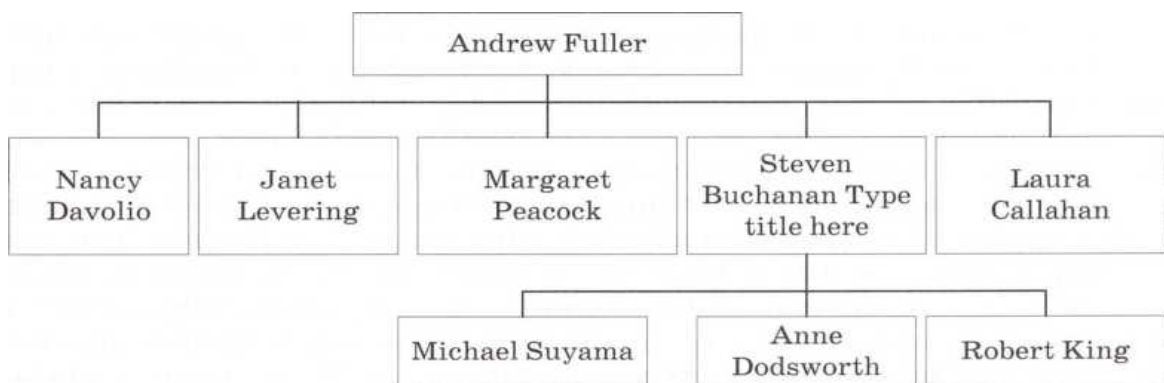


Рисунок 2.9 - Незбалансована ієрархія

Є також ієрархії, що займають проміжне становище між збалансованими і незбалансованими (вони позначаються терміном ragged— “нерівний”). Зазвичай вони містять такі члени, логічні “батьки” яких перебувають не на безпосередньо вищому рівні (наприклад, у географічній ієрархії є рівні Country, City і State, але при цьому в наборі даних є країни, що не мають штатів або регіонів між рівнями Country і City) (рис. 2.10).

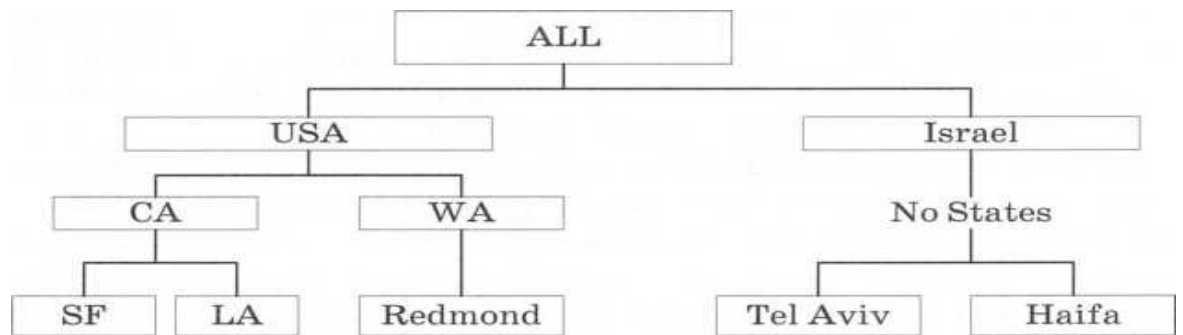


Рисунок 2.10 - “Нерівна” ієрархія

Наголосимо, що незбалансовані та “нерівні” ієрархії підтримуються не всіма OLAP-засобами. Наприклад, у MicrosoftAnalysisServices2000 підтримуються обидва типи ієрархії, а в MicrosoftOLAPServices7.0 — тільки збалансовані. Різною в OLAP-засобах може бути і кількість рівнів ієрархії, і максимально допустима кількість членів одного рівня, і максимально можлива кількість самих вимірювань.

Багатовимірний аналіз даних може бути проведений за допомогою різних засобів, які умовно можна поділити на клієнтські і серверні OLAP-засоби. Клієнтськими OLAP-засоби є додатки, що здійснюють обчислення агрегатних даних (сум, середніх величин, максимальних або мінімальних значень) і їх відображення, при цьому самі агрегатні дані містяться в кеші усередині адресного простору такого OLAP-засобу. Якщо початкові дані містяться в настільній СУБД, агрегатні дані обчислюють самим OLAP-засобом. Якщо ж джерело початкових даних — серверна СУБД, багато хто з клієнтських OLAP-засобів посилає на сервер SQL-запити і в результаті отримують агрегатні дані, обчислені на сервері.

Клієнтські OLAP-засоби використовуються, як правило, у разі не значної кількості вимірювань (зазвичай рекомендується не більше шести) і невеликій різноманітності значень цих параметрів, — адже отримані агрегатні дані повинні уміщатися в адресному просторі подібного засобу, а їх кількість росте експоненційно при збільшенні числа вимірювань. Тому навіть найпримітивніші клієнтські OLAP-засоби, як правило, дають можливість зробити попередній підрахунок обсягу необхідної оперативної пам’яті для створення в ній багатовимірного куба. Багато клієнтських OLAP-засобів дають змогу зберегти вміст кеша з агрегатними даними у

вигляді файла, що, у свою чергу, дозволяє не проводити їх повторне обчислення. Зазначимо, що нерідко така можливість використовується для відчуження агрегатних даних із метою передачі їх іншим організаціям або для публікації. Типовим прикладом таких відчужуваних агрегатних даних є статистика захворюваності в різних регіонах і за віковими групами, яка є відкритою інформацією та публікується міністерствами охорони здоров'я різних країн і Всесвітньою організацією охорони здоров'я.

Ідея збереження кеша з агрегатними даними у файлі отримала подальший розвиток у серверних OLAP-засобах, де збереження і зміна агрегатних даних, а також підтримка сховища, що містить їх, здійснюються окремим додатком або процесом, званим OLAP-сервером. Клієнтські додатки можуть запрошувати подібне багатовимірне сховище і у відповідь отримувати різні дані. Деякі клієнтські додатки можуть також створювати такі сховища або оновлювати їх відповідно до початкових даних, що змінилися.

Переваги застосування серверних OLAP-засобів порівняно з клієнтськими OLAP-засобами схожі з перевагами застосування серверних СУБД порівняно з настільними: у разі застосування серверних засобів обчислення і зберігання агрегатних даних відбуваються на сервері, а клієнтський додаток отримує лише результати запитів до них, що дає змогу в загальному випадку знизити мережевий трафік, час виконання запитів і вимоги до ресурсів, споживаних клієнтським додатком. Зазначимо, що засоби аналізу і обробки даних масштабу підприємства, як правило, базуються саме на серверних OLAP-засобах, наприклад таких як Oracle Express Server, Microsoft SQL Server 2000 Analysis Services, Hyperion Essbase, продуктах компаній Crystal Decisions, BusinessObjects, Cognos, SAS Institute. Оскільки всі провідні виробники серверних СУБД створюють різні серверні OLAP-засоби, вибір їх достатньо широкий і майже у всіх випадках можна придбати OLAP-сервер того ж виробника, що і у самого сервера баз даних.

OLAP-система складається з множини компонентів. На найвищому рівні представлення система включає джерело даних, OLAP-сервер і клієнта. Джерело даних — це засіб, з якого беруться дані для аналізу. Дані з джерела переносяться або копіюються на OLAP-сервер, де вони систематизуються і готуються для швидкого формулювання відповідей на запити. Клієнт — це призначений для користувача інтерфейс до OLAP-сервера.

Джерела даних. Джерелом в OLAP-системах є сервер, що постачає дані для аналізу. Залежно від галузі використання OLAP-продукту джерелом може слугувати сховище даних, успадкована база даних, що містить загальні дані, набір таблиць, що об'єднують фінансові дані або будь-яка комбінація з перерахованого. Здатність OLAP-продукту працювати з даними з різних джерел дуже важлива. Вимога єдиного

формату або єдиної бази, в яких би зберігалися всі початкові дані, не підходить адміністраторам баз даних. Крім того, такий підхід зменшує гнучкість і потужність OLAP-продукту. Адміністратори і користувачі вважають, що OLAP-продукти, які забезпечують витягання даних не тільки з різних, а і з безлічі джерел, виявляються гнучкішими і кориснішими, ніж ті, що мають жорсткіші вимоги.

Сервер. Прикладною частиною OLAP-системи є OLAP-сервер. Ця складова виконує всю роботу і зберігає всю інформацію, до якої забезпечується активний доступ. Архітектурою сервера управляють різні концепції. Зокрема, основною функціональною характеристикою OLAP-продукту є використання для зберігання даних багатовимірної (ММБД, MDDDB) або реляційної (РДБ, RDB) бази даних.

MOLAP. MOLAP (Multidimensional On-Line Analytical Processing — багатовимірний OLAP) означає, що сервер для зберігання даних використовує ММБД. Оскільки більшість OLAP-продуктів засновані на ММБД, під OLAP часто розуміють також і MOLAP. Сенс використання ММБД очевидний. Вона може ефективно зберігати багатовимірні за своєю природою дані, забезпечуючи засоби швидкого обслуговування запитів до бази даних. Дані передаються від джерела даних до багатовимірної бази даних, а потім база даних піддається агрегації. Попередній розрахунок — це те, що прискорює OLAP-запити, оскільки розрахунок зведених даних вже проведений. Час запиту стає функцією виключно часу, необхідного для доступу до окремого фрагмента даних і виконання розрахунку. Цей метод підтримує концепцію, згідно з якою робота проводиться одного разу, а результати потім використовуються багаторазово. Багатовимірні бази даних є відносно новою технологією. Використання ММБД має ті самі недоліки, що і більшість нових технологій: вони не такі стійкі, як РБД, і в тій же мірі не оптимізовані. Інше слабке місце ММБД полягає в неможливості використовувати більшість багатовимірних баз у процесі агрегації даних, тому потрібний час для того, щоб нова інформація стала доступна для аналізу.

Вибух бази даних є феноменом багатовимірних баз. Незважаючи на те що ця проблема досліджувалася фахівцями, важко пояснити, чому і як це відбувається. Здається, що це пов'язано з розрідженістю бази даних і попередньою агрегацією даних. Якщо багатовимірний база даних містить невелику кількість елементів даних, порівнянну з кількістю забезпечуваних нею рівнів агрегації, кожен фрагмент даних робитиме більший внесок до всіх отримуваних із нього даних. Коли база даних “вибухає”, розмір її стає істотно більшим, ніж повинен бути. Складно визначити умови “вибуху” бази даних або передбачити, чи “вибухне” якась конкретна структура. Одним із підходів, який може допомогти вирішити проблему “вибуху”, є динамічне управління розрідженими даними. Ця методика допомагає аналізувати власні моделі зберігання і

оптимізувати їх з метою запобігання “вибуху” бази даних.

(On-Line Analytical Processing), реляційний OLAP, означає, що OLAP-сервер базується на реляційній базі даних. Початкові дані вводяться в реляційну базу даних зазвичай за схемою “зірка” або “сніжинка”, що сприяє скороченню часу витягання. Сервер забезпечує багатовимірну модель даних за допомогою оптимізованих SQL-запитів. Є низка причин для вибору саме реляційної, а не багатовимірної бази даних. РБД — це добре відпрацьована технологія, що має безліч можливостей для оптимізації. До того ж РБД підтримують потужніші обсяги даних, ніж ММБД. Вони якраз і спроектовані для таких обсягів. Основним аргументом проти РБД є складність запитів, необхідних для одержання інформації з великої бази даних за допомогою SQL.

Прикладний OLAP (HOLAP). Безумовно, це найбільша галузь, і це, загалом, те, з чим зазвичай зв’язують і що зазвичай розуміють під терміном OLAP. Прикладний OLAP, як правило, складається з багатовимірних баз даних, доступ до яких відбувається через конкретний додаток, або, можливо, через безліч додатків. Постачальники в цій галузі ринку в основному пропонують клієнтів для бази даних. Клієнт може бути як простим засобом перегляду, так і потужнішим додатком.

Настільний OLAP (DOLAP). Представниками настільного OLAP є продукти, що необов’язково поєднуються з сервером. Вони можуть запускатися в основному на клієнтській частині, хоча дані у формі куба даних можуть завантажуватися із сервера. Той факт, що куб даних будується і зберігається на машині користувача, дає можливість рекомендувати їх тим, хто часто використовує портативні комп’ютери або нечасто запускає настільки складні звіти, що для їх формування необхідна вища швидкість клієнта, а отже, і потужніший сервер для її забезпечення.

Швидка реалізація запитів є імперативом для OLAP. Це один із базових принципів OLAP — здатність інтуїтивно маніпулювати даними потребує швидкого витягання інформації. В цілому, чим більше обчислень необхідно провести, щоб отримати фрагмент інформації, тим повільніше відбувається відгук. Тому, щоб зберегти короткий проміжок часу для реалізації запитів, фрагменти інформації, до яких зазвичай звертаються найчастіше, але які при цьому потребують обчислення, піддаються попередній агрегації. Тобто вони підраховуються і потім зберігаються в базі даних як нові дані. Як приклад типу даних, який допустимо розрахувати наперед, можна навести зведені дані — як-от, показники продажу за місяцями, кварталами або роками, для яких дійсно введеними даними є щоденні показники.

Різні постачальники дотримуються різних методів відбору параметрів, що потребують попередньої агрегації і кількості заздалегідь обчислюваних величин. Підхід до агрегації впливає одночасно і на базу даних, і на час реалізації запитів. Якщо обчислюється більше величин,

вірогідність того, що користувач затребує вже обчислену величину, зростає, і тому час відгуку скорочується, оскільки не доведеться запрошувати початкову величину для обчислення. Проте, якщо обчислити всі можливі величини — не найкраще рішення — у такому разі істотно зростає розмір бази даних, що робить її некерованою, і час агрегації буде дуже великим. До того ж, коли в базу даних додаються числові значення, або якщо вони змінюються, інформація ця повинна відбиватися на заздалегідь обчислених величинах, залежних від нових даних.

Клієнт. Клієнт використовується для подання і маніпуляцій з даними в базі даних. Клієнт може бути і достатньо нескладним — у вигляді таблиці, що включає такі можливості OLAP, як, наприклад, обертання даних (півотінг) і поглиблення в дані (дриллінг). Ці процеси повинні бути спеціалізованими, але мати такий же простий засіб перегляду звітів або бути таким самим потужним інструментом, як створений на замовлення додаток, спроектований для складних маніпуляцій із даними. Клієнт є настільки важливим, що безліч постачальників зосереджують свої зусилля виключно на його розробці. Все, що включається до складу цих додатків, становить стандартний погляд на інтерфейс, наперед задані функції і структуру, а також швидкі рішення для більшості стандартних ситуацій, наприклад популярні фінансові пакети. Наперед створені фінансові додатки нададуть можливість фахівцям використовувати звичні фінансові інструменти без необхідності проектувати структуру бази даних або загальноприйняті форми і звіти.

Інструмент запитів/генератор звітів. Інструмент запитів або генератор звітів пропонує простий доступ до OLAP-даних. Вони мають простий у використанні графічний інтерфейс і дають можливість користувачам створювати звіти шляхом переміщення об'єктів методом “draganddrop”. Тоді як традиційний генератор звітів надає користувачеві можливість швидко випускати формат-вані звіти, генератори звітів, що підтримують OLAP, формують актуальні звіти. Кінцевим продуктом є звіт, що має можливості заглиблення в дані до рівня подробиць, обертання (півотінг) звітів, підтримки ієрархій та ін.

Сьогодні в багатьох напрямках бізнесу за допомогою електронних таблиць проводяться різні форми аналізу корпоративних даних. У певному розумінні це ідеальний засіб створення звітів і перегляду даних. Аналітик може створювати макроси, що працюють з даними у вибраному напрямі, а шаблон може бути спроектований так, що коли відбувається введення даних, формули розраховують правильні величини, виключаючи необхідність неодноразового введення простих розрахунків. Проте все це дає в результаті “плоский” звіт, що означає: як тільки він створений, важко розглядати його в різних аспектах. Наприклад, діаграма відображає інформацію за певний часовий період, скажімо, за місяць. І якщо хтось бажає побачити показники за день (на противагу даним за місяць),

необхідно створити абсолютно нову діаграму: визначити нові набори даних, додати в діаграму нові позначки і внести безліч інших простих, але трудомістких змін. Крім того є низка галузей, в яких можуть бути допущені помилки, що в цілому зменшує надійність. Коли до таблиці додається OLAP, з'являється можливість створювати єдину діаграму, а потім піддавати її різним маніпуляціям із метою надання користувачеві необхідної інформації, не обтяжуючи себе створенням всіх можливих представлень.

Інтернет у ролі клієнта. Новим членом сімейства OLAP-клієнтів є Інтернет. Існує маса переваг у формуванні OLAP-звітів через Інтернет. Найбільш істотною є відсутність необхідності в спеціалізованому програмному забезпеченні для доступу до інформації.

Кожен інтернет-продукт специфічний. Деякі спрощують створення web- сторінок, але мають меншу гнучкість. Інші дозволяють створювати представлення даних, а потім зберігати їх як статичні HTML-файли. Все це дає можливість проглядати дані через Інтернет, але не більше того. Активно маніпулювати даними з їх допомогою неможливо. Є й інший тип продуктів — інтерактивний і динамічний, такий, що перетворює ці продукти на повнофункціональні інструменти. Користувачі можуть заглиблюватися в дані, півотінг, обмеження вимірювань тощо. Перш ніж вибрати засіб реалізації Інтернет, важливо зрозуміти, які функціональні можливості потрібні від web-рішення, а потім визначити, який продукт найкраще втілить цю функціональність.

Додатки. Додатки — це тип клієнта, що використовує бази даних OLAP. Вони ідентичні інструментам запитів і генераторам звітів, описаним вище, але, крім того вносять до продукту ширші функціональні можливості. Додаток, як правило, володіє більшою потужністю, ніж інструмент запити.

Середовище. Зазвичай постачальники OLAP забезпечують середовище розробки для створення користувачами власних настроєних додатків. Середовище розробки в цілому є графічним інтерфейсом, що підтримує об'єктно-орієнтовану розробку додатків. До того ж, більшість постачальників забезпечують API, який може використовуватися для інтеграції баз даних OLAP з іншими додатками.

Розглянемо деякі напрями діяльності основних виробників програмних засобів підтримки OLAP-технологій з урахуванням наведеної вище класифікації архітектур.

Різні постачальники реалізують OLAP на основі власних корпоративних уявлень про те, що повинно входити до ідеального OLAP-продукту. Постачальники, розглянуті нижче, дотримувалися різних підходів, включаючи й ті, що ґрунтуються на багатовимірній базі даних, реляційній базі даних і додатках, які реалізують можливості OLAP на різних рівнях. Процес їх реального функціонування може слугувати

кількісним параметром при розгляді різних підходів до OLAP. З цією метою OLAP Council розробив атестаційне завдання APB-1 для кількісного порівняння роботи різних OLAP-продуктів. OLAP Council є консорціумом, утвореним декількома постачальниками OLAP для підтримки ключових принципів OLAP. Вони визнають, що OLAP є найважливішою технологією для забезпечення корпоративних аналітиків інструментами, потрібними для виконання необхідного їм аналізу. У квітні 1996 р. було випущено перше атестаційне завдання в галузі OLAP— APB-1. Контрольне завдання визначає параметри бази даних і встановлює набір з 10 запитів, що відображають нормальне використання. Постачальник OLAP створює відповідну базу даних і потім запускає запити. Окреме вимірювання — AQT (Average Query Time— середній час запиту), генерується на основі часу, який витрачається на завантаження бази даних, агрегацію даних і подальший запуск запитів. Правила визначають, що легально і що нелегально, наприклад, чи потрібно розраховувати наперед значення даних, які з розрахованих величин можуть зберігатися та ін.

У галузі MOLAP-архітектури лідером є компанія Oracle. Лінійка продуктів Express та Oracle OLAP Services є корпоративним підходом Oracle до OLAP, включаючи сервер, клієнтську частину, можливість ROLAP та інтернет-рішення. Express Server був головною OLAP-машиною для Oracle та інших OLAP-продуктів фірми і надавав багатовимірну базу даних. До того ж Oracle пропонує Personal Express— сервер Express, що працює локально. Він надає користувачам доступ і можливість працювати з базою даних в автономному режимі. Все це ідеально підходить для мобільних комп'ютерів. Головною особливістю Express Server є здатність використовувати різноманітні джерела даних. Дані для багатовимірної бази даних можуть збиратися з реляційної бази (за допомогою Express Relational Access Manager), багатовимірної бази, табличного або плоского файлу. Користувачі можуть розробляти власні OLAP- додатки для Express за допомогою Express Objects, який забезпечує розробника графічним інтерфейсом і об'єктно-орієнтованим підходом для створення додатків. Oracle також забезпечує безліч шляхів доступу до бази даних. Express Analyzer містить графічний інтерфейс до бази даних і дає змогу користувачеві легко формувати звіти. Крім того Analyzer може мати загальні об'єкти з Express Objects, а також випускати додатки, розроблені за їх допомогою. Discoverer найточніше можна описати як інструмент запитів до даних. Він простіший, ніж Analyzer, проте найбільш популярний серед засобів запитів до даних. Крім цього, Express містить додаткові таблиці, які можна використовувати спільно з Excel.

В Oracle є два готові додатки: Financial Analyzer та Sales Analyzer. Вони використовують машину і кеш даних Express Server, містять конфігурації звітів, розроблені для потреб фінансових аналітиків і аналітиків продажу, корисні і користувачам, оскільки визначають важливі

функції, що зазвичай беруть участь в обох видах аналізу, який реалізовано в Express Server.

WebAgent і WebPublisher — це засоби створення інтернет-ресурсів, що надаються Express. Використовуючи будь-який з цих засобів, користувач може створити динамічний та інтерактивний сайт, що забезпечується застосуванням різних можливостей OLAP. WebAgent більшою мірою інструмент розробника, він є набором заздалегідь певних процедур, включених в ExpressServerSPL. Сайти можуть будуватися так само, як будуються призначені для користувача інтерфейси до бази даних. Для кінцевого користувача є WebPublisher. WebPublisher зв'язаний з ExpressAnalyzer і має можливість створення власних інтерактивних сайтів тими, хто не має серйозного досвіду програміста. WebPublisher в основному є “майстром”, який веде користувача через всі етапи побудови сайта і забезпечує графічний інтерфейс для підтримки його створення.

Arbor Software Corporation — це головний суперник Oracle. Його продуктом є Essbase, а останнім релізом став EssbaseServer5. Дуже популярний сервер Arbor для різних OLAP-продуктів, це свідчить про те, що багато постачальників OLAP не обов'язково випускають повні додатки, а можуть використовувати як базу даних Arbor, і потім створювати інтерфейси до бази даних. Як приклад можна навести Comshare і web-компоненту для Arbor-Crystalinfo, SeagateSoftware, що випускається. Останнім часом Arbor уклав партнерську угоду з IBM. Багатовимірне зберігання даних в ArborEssbaseServer буде замінено на DB2 від IBM. Передбачається, що це буде ROLAP-система, але фактично це не так. Це просто OLAP-система без однієї з кращих властивостей багатовимірних баз даних, але з перевагами системи РБД.

Додатково до таблиць EssbaseSpreadsheetAdd-in, що забезпечують користувачів можливостями OLAP, Arbor пропонує WIREDforOLAP (засіб аналізу і презентацій), CrystalInfoforEssbase (генератор звітів і розкладів) і SQLDrill-Through, що дає змогу користувачам проглядати подробиці даних у початкових реляційних базах. Arbor також випустив ArborEssbaseAdjustmentModule. Цей додаток допомагає користувачам у підготовці звітів, що регулярно випускаються. Він сприяє автоматизації форматування звітів і процесів розрахунку. Крім того, є ще ArborCurrencyConversionModule, здатний конвертувати різні валюти в національну на основі моделі для відстежування обмінних курсів.

У галузі ROLAP визнаним лідером є компанія MicroStrategy. Її філософією є відсутність обмежень на розмір сховища даних, отже немає жодних проблем з його збільшенням. Оскільки вони є виробниками реляційного OLAP, рівень їх аналітичних систем достатньо високий. У MicroStrategy немає OLAP-машини, яка могла б працювати локально, що незручно для користувачів, які часто працюють з ноутбуками або для тих,

хто просто вважає за краще працювати автономно. Проте у них існує продукт, DSSBroadcaster, що дає можливість посилати дані на різні вихідні пристрої. DSSBroadcaster надсилає дані за запитом або коли відбувається певна подія. Наприклад, менеджером може відсилатися щоденне оновлення з сумами прибутку за попередній день. Ця інформація може надходити електронною поштою, на пейджер або мобільний телефон, а також факсом.

DSSServer є центральним продуктом у лінійці продуктів MicroStrategy. Це потужна машина, що дає змогу іншим агентам отримувати доступ до реляційної бази даних у багатовимірному режимі. DSSServer містить різноманітні драйвери баз даних для оптимізації їх під необхідну реляційну базу даних (вони підтримують Oracle, DB2, Sybase, RedBrick, Informix та інші реляційні бази). До того ж акцент робиться на здатності їх зростання і включає драйвер для адаптації до дуже великих баз даних (VeryLargeDatabases, VLDBs), розмір яких перевищує терабайти. Природа реляційного OLAP-продукту обмежує MicroStrategy в можливості надання справді індивідуального сервера для автономної роботи, проте за допомогою DSSAgent набір даних може завантажуватися і аналіз може виконуватися навіть в автономному режимі. DSSAgent є клієнтом або клієнтським інструментом до DSSServer. Однією з переваг DSSAgent є застосування інтелектуальних агентів для автоматизації бізнес-процесів. Наприклад, за допомогою DSSAgent можна створити агента, що знає, коли і де необхідно шукати дані, що з ними робити (тобто, як проводити їх очищення і куди їх помістити). Використовуючи агентів, можна автоматизувати безліч звичайних, але часто повторюваних завдань. DSSExecutive використовує можливості DSSAgent для реалізації високоякісного генератора звітів і засобу аналізу. Він використовує об'єктно-орієнтований підхід та інтерфейс, що працює за технологією "drag-and-drop" для швидкого створення додатків управлінських інформаційних систем силами самих користувачів. І, нарешті, MicroStrategy пропонує доповнення ExcelAdd-in, яке можна застосовувати для додавання таблицям функціональних можливостей OLAP.

Нове покоління інтернет-орієнтованого OLAP від MicroStrategy представлено DSSWeb5.0. Однією з особливостей DSSWeb5.0 є підтримка Microsoftwebcastingstandart. Це дає можливість автоматично передавати web-сторінки на комп'ютер користувача. Серед найважливіших можливостей DSSWeb можна назвати здатність зберігати карти або діаграми, отримані з Інтернету, майстри звітів і пакети звітів, що настроюються.

Як зразок прикладного OLAP-продукту можна узяти Comshare. Незважаючи на те, що це додаток, він доповнює продукт функціональними можливостями OLAP. ComshareDecision проявляє гнучкість щодо використовуваного спільно з ним сервера. ArborEssbase і OracleExpress—

всі ці багатовимірні сервери баз даних можуть використовуватися спільно з Decision.

HyperionSoftware— також виробник прикладного OLAP-продукту, що випускає виключно OLAP-клієнтів. Останнім продуктом був HyperionMBA, або MultidimensionalBusinessAnalyst, що замінив HyperionOLAP. Згідно з останніми даними, Hyperion посів другий за розміром сегмент ринку. Природа продукту гарантує, що велика частина функціональних можливостей ґрунтується на серверній базі даних. Hyperion популярний завдяки своїм додатковим аналітичним можливостям, що реалізуються у формі складних вимірювань, заздалегідь певних функцій і звітності. Метою його є формування могутнього фінансового пакета, що включає OLAP.

Hyperion пропонує два клієнтські OLAP-рішення. Перше, HyperionMBA, використовує OLAP для бізнес-аналізу. Як це зазвичай буває в OLAP-додатках, Hyperion застосовує у своїх рішеннях складні вимірювання і заздалегідь певні функції для розрахунків і маніпуляцій з валютами. Програма HyperionAnalyticAccounting включає властивості OLAP в розрахунковий пакет.

Cognos слугує непоганим прикладом настільного OLAP-продукту. Це означає, що більша частина обробки проводиться не на сервері, а локально. Impromptu є інструментом запитів, використовуваним для витягання даних з багатовимірної бази даних. Дані потім поміщаються в Powerplay, яка зберігає куб даних на робочому столі комп'ютера користувача.

2.6 OLAP-системи та інтернет-технології

Щоб підтвердити необхідність і значне поширення OLAP, розглянемо декілька прикладів розв'язання задач за допомогою OLAP-інструментів великими зарубіжними підприємствами.

Наприклад, DukeEnergy— одна з найбільших американських енергетичних компаній, активи якої перевищують 20 млрд дол. США, а кількість співробітників становить 22 000 осіб, які працюють у різних підрозділах компанії по всьому світу. Сферою діяльності компанії є енергетичне забезпечення, обслуговування трубопроводів і постачання електричної енергії, природного газу і зрідженого природного газу. На момент впровадження OLAP компанія планувала вийти на нерегульований урядом енергетичний ринок і відвоювати собі певний шматок світового енергетичного бізнесу. У процесі підготовки до цього нового етапу відділ управління інформацією DukeEnergy опрацьовував різні можливості застосування технологій для управління змінними інформаційними потребами підприємства і забезпечення йому гідного місця на ринку та відповідного розвитку. Питання полягало в забезпеченні кінцевих

користувачів доступом до інформації, що зберігається в корпоративних базах даних. Фахівці з управління інформацією з'ясували, що в процесі ухвалення рішень всім підрозділам, починаючи з фінансових відділів і закінчуючи кадровими, важливо мати доступ до такої інформації тоді і там, де вона їм необхідна.

Система, яку DukeEnergy застосувала до цього моменту, була складною у користуванні і без втручання фахівців з управління інформацією не забезпечувала необхідний підприємству рівень точності та своєчасності даних. Плануючи апгрейд систем підтримки прийняття рішень так, щоб співробітники фінансових та інших нетехнічних підрозділів компанії могли якнайповніше використовувати корпоративні дані, DukeEnergy ухвалила рішення про впровадження OLAP-системи. Була вибрана система підтримки прийняття рішень від BusinessObjects. У результаті її впровадження дані, що містяться в оперативній базі PeopleSoft і вітрині даних IBMDB2 HRMS, стали доступні для OLAP-аналізу, запитів і звітів різних підрозділів компанії. Рішення використовує RDTs (RapidDeploymentTemplates — шаблони швидкого розгортання), що містять зразки звітів, які користувачі без спеціальних технічних знань можуть налаштовувати на власний розсуд для генерації незапланованих запитів до даних і звітів.

У цьому випадку компанія визначила, що пропозиція BusinessObjects як найкраще відповідає її потребам. Але це не означає, що інші рішення гірші. Кожне підприємство, що ухвалило рішення про впровадження OLAP, в першу чергу розглядає оптимальність пропозицій різних постачальників стосовно власних проблем і потреб. Наприклад, фірма Cognos успішно постачає свої продукти таким клієнтам, як Міністерство оборони США і компанія Boeing. Співпрацюючи з Cognos з 1996 р., Міністерство оборони постійно оновлює наявні системи, доповнюючи їх новими можливостями і розробками. Нещодавно фінансовим підрозділом Міністерства (DFAS) була придбана система бізнес-репортування і аналізу з новими можливостями візуалізації, вартістю 1 млн дол. Нове програмне рішення об'єднує дані з безлічі розрізаних систем, роблячи їх доступними для тисяч співробітників DFAS та інших агентств Міністерства через захищений інтернет-портал. "Візуалізатор" перетворює складні дані на зрозумілу зручну для використання інформацію, відображаючи її з використанням багатой графіки, різноманітних презентаційних можливостей і функцій оцінки даних. Роберт Еш, перший віце-президент Cognos, називає це рішення "ядром електронного уряду (e-government)", що забезпечує федеральним агентствам можливість використовувати Інтернет для спрощення і прискорення доступу до необхідної їм інформації. Впровадження OLAP-технології допомогло створити таку картину для кінцевих користувачів, що отримали у результаті можливість сформулювати поглиблене уявлення

про роботу Агентства, що у свою чергу, сприятиме поліпшенню управління і ухваленню рішень відповідними співробітниками.

BoeingCompany мало дещо інше завдання. Оскільки всі літаки зібрані на замовлення і містять до декількох мільйонів різних компонентів, співробітникам було потрібне серйозне аналітичне рішення для обробки даних таких істотних обсягів. У межах спеціальної програми зі створення інфраструктури для репортінгу у сфері прийняття рішень в підрозділі комерційної авіації почато впровадження продукту CognosImpromptuWebReports, за допомогою якого 11 500 співробітників підрозділу зможуть генерувати через Інтернет звіти до бази даних за накладними і специфікаціями на компоненти літальних апаратів. Керівники BoeingCompany вважають, що нове рішення заощадить значну кількість людино-годин, які раніше витрачалися на створення таких звітів уручну, що закономірно сприятиме поліпшенню процесу прийняття рішень.

Компанія MicroStrategy, як і більшість лідерів цього ринку, також пропонує спеціалізовані рішення для окремих областей бізнесу — фінансів, страхування, охорони здоров'я, урядових організацій та ін. — на базі MicroStrategyBusinessIntelligencePlatform. Одне з таких рішень отримало застосування в галузі енергетики. Компанія KeySpan, найбільший на північному сході США постачальник природного газу, що має підрозділи в Брукліні, Бостоні і на Лонг-Айленді, керівник низки сервісних енергетичних компаній і більш ніж 13 000 співробітників, використовує MicroStrategyNarrowcastServer для забезпечення Нью-Йоркських домовласників і підприємств достатньою кількістю енергії для опалювання та підігріву води. NarrowcastServer, що є ключовим компонентом MicroStrategyBusinessIntelligencePlatform, надає користувачам значні аналітичні можливості та інтелектуальну систему попередження. Зокрема, система надсилає попередження про зниження або підвищення витрати газу щодо проектного на відповідні адреси електронної пошти і пейджери. Клієнти KeySpan, “продавці”, що постачають газ кінцевим користувачам, можуть підключитися до інформаційних ресурсів компанії для перегляду і аналізу проектного і реального використання газу та ухвалення обґрунтованих рішень щодо обсягів газу, необхідних для передачі мережою трубопроводів у кожний конкретний день. KeySpan використовує технології MicroStrategy для аналізу таких чинників, як історичні дані і погодні умови, щоб спланувати обсяги постачання газу. NarrowcastServer чотири рази на день порівнює фактичний потік газу із запланованим раніше того ж дня і посилає звіт про непланові ситуації електронною поштою або пейджером відповідному клієнтові KeySpan та оперативним підрозділам, що контролюють вентилі трубопроводу. Таким чином компанії-постачальники можуть оптимізувати обсяги газу, що постачається споживачам залежно від низки різних

чинників.

Одним із сучасних напрямів розвитку систем інтелектуального аналізу даних є об'єднання OLAP із технологією DataMiningі сховищами даних. Ці всі три технології розвиваються у міру того, як компанії починають усвідомлювати цінність даних. Реляційні бази даних свого часу були революційним рішенням, яке дало можливість підприємствам збирати дані зі щоденних транзакцій у великомасштабні засоби зберігання. За допомогою SQL було можливо виконувати елементарний аналіз цих даних. Коли ж був потрібен складніший аналіз, з'ясувалося, що SQLі РБД зовсім не ідеальне рішення. Таблиці могли забезпечувати гнучкіший аналіз, але мали низку істотних недоліків. Дані, що підлягають імпорту в таблицю з бази даних і сама таблиця були не в змозі ефективно оперувати великими обсягами даних. З часом все більше і більше компаній почали реалізовувати сховища даних і застосовувати до своїх даних засоби DataMining. Сховище даних забезпечує зберігання очищених корпоративних даних. Дані за транзакціями перевіряються на коректність, категоризуються і потім розміщуються у сховищі. Інструменти DataMining дають можливість аналітикам підприємств виявити приховані тенденції даних. Інструмент OLAP допомагає виконувати швидкий і простий аналіз даних. У цілому користувач - аналітик має уявлення про те, що він збирається знайти в деякому представленні даних. Він просто хоче мати засіб маніпулювання даними, щоб більш наочно відобразити деякі їх аспекти.

Іншим напрямом розвитку таких систем є застосування OLAP у корпоративних порталах. Корпоративний портал — це “точка доступу”, яка забезпечує зовнішніх і внутрішніх користувачів єдиним, безпечним мережевим інтерфейсом із персоніфікованим контентом. Контент слід розуміти в найширшому значенні, як все, що об'єднує портал в одному призначеному для користувача інтерфейсі: додатки, інформація та інструменти спільної діяльності. До додатків належать внутрішні або зовнішні операційні або аналітичні програмно- апаратні засоби. Інформація є структурованими даними, результатами аналізу (звіти, куби, графіки, таблиці тощо), внутрішнім або зовнішнім неструктурованим контентом (документи, вміст цифрових накопичувачів, статистика відвідання сайта). Інструменти спільної діяльності — це web-чати, мережеві конференції, електронна пошта, служби миттєвих повідомлень тощо.

Кажучи про місце OLAP-технологій у порталі, буде доречним навести позицію фахівців із консалтингової компанії ObjectSystemsGroup. На їх думку, сфера застосування OLAP-інструментів у загальному розумінні — це “просунутий” бізнес-аналіз (BusinessIntelligence). Ці інструменти можуть звертатися до даних різних об'єктів, зіставляти їх, а також здійснюють пошук інформації (з тематики) в масштабах,

недосяжних для більшості користувачів, на відміну від ручного способу. Нині більшість компаній вже мають у своєму використанні багато компонентів, необхідних для впровадження OLAP-порталів. Безперечно, можливість отримати будь-яку бізнес-інформацію, взявши за мету питання “як це стосується того, чим я займаюся?” — потужний інструмент підтримки прийняття рішень, про який користувачі можуть тільки мріяти.

Важливим напрямом розвитку систем інтелектуального аналізу даних, який набуває популярності, є поєднання OLAP-засобів та web-технологій. Динамічні технології, поява яких стала можлива в результаті розвитку WorldWideWeb, є прекрасною альтернативою традиційним клієнт-серверним OLAP-методам. Останнім часом з’явилася низка OLAP-засобів (їх називають Web-OLAP, або WOLAP), устаткованих web-можливостями. Вони виконують аналітичні функції, такі як агрегація і деталізація (drill-up і drill-down), а також забезпечують високу продуктивність у поєднанні з всіма перевагами, які дає web-додаток. Поява web-OLAP-засобів стирає межі, що відокремлюють OLAP-ринок від суміжних категорій програмного забезпечення. Web-платформи інтерактивної звітності за своєю функціональністю все більше і більше схожі на стандартні web-OLAP-продукти. Більшість web-OLAP-додатків використовують загальну архітектуру, в якій клієнтський браузер взаємодіє з HTTP-сервером, що пересилає HTML-сторінки. Але крім цього надається ще і проміжне ПО, що зберігається на сервері. Такий компонент може безпосередньо зв’язуватися з web-браузером або взаємодіяти з HTTP-сервером, який потім повертає браузеру HTML-сторінки з додатковими даними.

Web-OLAP-компонент проміжного рівня виконує набір функцій, які не може забезпечити HTML, а саме:

- взаємодія з базою даних, де знаходиться сховище;
- зберігання станів (попередніх транзакцій бази даних);
- обчислення і буферизація даних, що повертаються до клієнта.

На сьогодні реалізовано декілька різних рішень Web-OLAP, у тому числі на основі технологій HTML (DHTML), Java, ActiveX, а також їх комбінацій. Розглянемо основні типи таких програмних продуктів:

- HTML (DHTML)-рішення;
- HTML з розширеннями — CGI;
- HTML з використанням Java-апплетів;
- Java або ActiveX-компоненти.

HTML-рішення. Для реалізації OLAP-функціональності в web-браузері використовується лише HTML. Простим прикладом такого рішення є OLAP-інструмент, що дає змогу користувачеві виконувати обумовлені OLAP-запити а отримувати звіти з браузера. В цьому випадку для здобуття даних з автономних OLAP-інструмент використовують планувальники, що формують статичні HTML-звіти за HTML-шаблонами.

Шаблони створюються так, щоб всі звіти мали узгоджений вигляд. Звіти обробляються і передаються в браузер за допомогою web-сервера. Статичні звіти швидко доставляються в браузер, при цьому взаємодії користувача з браузером практично не відбувається. У деяких випадках імітується перехід між вимірами шляхом навігації по звіту. Планувальник може створити набір звітів, пов'язаних між собою гіперпосиланнями. Наприклад, користувач, клацнувши на посилання під назвою “Третій квартал”, перейде до іншого звіту, що містить дані за липень, серпень і вересень.

Існує и інші підхід, коли OLAP-сервер наповнює HTML-шаблон даними в оперативному режимі, тобто в міру появи запиту користувача через браузер. У цьому випадку на web-сервері зберігаються лише шаблони звітів і метадані. Цю дію, необхідну web-серверу для передачі даних у HTML файл перед тим, як відправити його до браузера.

За будь-якого способу зберігання метаданих шаблонів і звітів інформація збирається web-сервером згідно з кодом звіту, що надсилається з браузера. Програмне забезпечення web-сервера використовує метадані звіту, або витягує відповідні дані з бази. База може зберігатися як на тому ж комп'ютері, що і додаток web-сервера, так і на іншому. Отримані з бази дані об'єднуються на основі шаблону в звіт і передаються в браузер. Як правило, звіт вже містить за умовчанням певну OLAP-функціональність. При взаємодії з звітом призначені для користувача код посилається разом з іншою інформацією на web-сервер, що використовує цей код для відстежування інформації, яку користувач бачить в браузері. Повний набір OLAP-функцій (обертання, агрегація і заглиблення в дані) за такого підходу не передбачений.

HTML з розширеннями — CGI. Одне з найслабших місць HTML — неможливість збереження стану. Пропонований варіант вирішення проблеми — використання CGI (Common Gateway Interface — загальний шлюзовий інтерфейс) або інших WebAPI (Application Programming Interface — інтерфейс прикладного програмування) для реалізації єдиного ПО (middleware). За допомогою цього методу можна забезпечити зберігання станів, а також буферизацію рядків і виконання деяких обчислень переданих даних. У разі використання такої архітектури переносимість для клієнтських платформ забезпечується за рахунок використання HTML як інтерфейсу. Проте і в цьому випадку застосування лише HTML накладає обмеження на інтерфейси. Правильне відображення графіків і звітів буде ускладнено. Ця проблема вирішується за допомогою Java-апплетів, які мають значні можливості з управління виведенням інформації на монітор.

HTML з використанням Java-апплетів. Спільне використання HTML і Java-апплетів допомагає створити потрібне рішення Web-OLAP. У цьому випадку HTML застосовується для відображення меню і виконання простих інтерфейсних функцій, а за допомогою Java-апплетів

забезпечуються складніші компоненти інтерфейсу програми, а також погоджене відображення діаграм, графіків і таблиць. У таких застосуваннях найчастіше реалізовані функції агрегації і деталізації, а також обертання даних. За рахунок широких графічних можливостей Javaпорівняно з HTMLможна подавати дані у вигляді діаграм і змінювати їх в інтерактивному режимі.

Javaабо ActiveX-компоненти. Наступний підхід передбачає використання Javaабо ActiveX-компонентів для мінімізації взаємодії між браузером і web-сервером, а також розширення можливостей користувача у роботі з даними за рахунок якіснішого інтерфейсу. Існує два способи використання цих компонентів.

У першому випадку web-сервер заповнює файл даними для звітів, і інтерфейси компоненти посилаються в браузер разом із відповідним HTML-файлом. Браузер викачує цей файл, а компонент завантажує дані. Таким чином, компоненти забезпечують виконання таких OLAP-функцій, як агрегація, деталізація і обертання за допомогою зручного інтерфейсу, без звернення до сервера.

У другому випадку, інтерфейсний компонент безпосередньо з'єднується з сервером, який в інтерактивному режимі передає дані про клієнта в міру появи запитів користувачів. У цьому випадку інтерфейсний компонент запрошує дані з web-сервера, відкриваючи HTTP-потік. Потім, отримавши результати з сервера, він аналізує отриману інформацію і передає дані на об'єкт для відображення.

Історично склалося так, що інтелектуальний аналіз даних (DataMining) і онлайн аналітична обробка інформації (OLAP) були виключною прерогативою людини, тобто саме люди визначали і створювали аналітичні моделі, а потім використовували отримані з їх допомогою результати. Але з появою обчислювальної моделі web-служб, які розглядаються як універсальний засіб об'єднання різнорідних систем, аналітику тепер можна легко пов'язати з іншими обчислювальними завданнями. Іншими словами, люди перестали бути єдиними творцями або споживачами аналітичних сервісів. Як ви розумієте, це відкриває захоплюючі можливості.

Що ж таке аналітична взаємодія “без людей”? Візьмемо, наприклад, “інтелектуального агента”, що обслуговує B2B-обмін у процесі пошуку партнерів з операції. Окрім іншого, пошук передбачає кількісне визначення ризиків — процес, аналогічний оцінці ризиків під час видачі кредиту. Кінцеве завдання агента відповісти на питання, чи підходить партнер X для включення в операцію. Агент визначає важливі критерії (змінні і алгоритми) в запиті на оцінювання ризику. Цей запит може включати обчислення тимчасових рядів, ранжирування і схожу на DataMining процедуру оцінки невідомих параметрів партнерів, але автором і споживачем запиту є агент, а не людина. Більше того, інтелектуальний

агент повинен володіти реальною аналітичною “майстерністю на підставі результатів одного запиту створити другий набір запитів, на підставі другого третій тощо. На перший погляд API нагадує інтерактивний сеанс, а з позицій користувача — систему з власним “інтелектом”. Донедавна можливості організації взаємодії аналітичних “двигунців” були обмежені. Розробники мали в розпорядженні низку створених виробниками API, а також технологію MicrosoftOLEDBforOLAP (або ODBO), яку окрім Microsoft підтримували й інші сервери. Але протокол ODBO доступний лише в Win32, що навряд чи могло зробити його універсальним засобом доступу для web- служб або корпоративних додатків середнього рівня. Java-версію створеного Радою з OLAP (OLAP Council) інтерфейсу MD-API не реалізував жоден постачальник серверних або клієнтських продуктів, тому для серверів додатків на основі відмінних від Win32 технологій мультивендорні API взагалі відсутні. Проте з’являються нові можливості. Перший інтерфейс називається XMLforAnalysis (XML/A), інший — JavaOLAPInterface (JOLAP). Перша перевага цих інтерфейсів — підтримка додатків середнього рівня, створених не на Win32. Зрозуміло, що обидва інтерфейси корисні для інших цілей. Так, для клієнтських застосувань ці API як мінімум забезпечують доступ до раніше недоступної підмножини серверів.

XML/A — це API для доступу до інформації. Як і в OLEDB для OLAP, клієнт може давати запит на багатовимірну інформацію у вигляді набору записів реляційних даних, а також багатовимірних наборів комірок. Проте специфікації досить відкриті і забезпечують підключення інших типів провайдерів і доступ до них без порушення нормативів.

JOLAP оснований на Java, і тому всі аспекти цього API об’єктно-орієнтовані. Важливо, що не визначено жодної текстової мови запитів — замість цього в об’єктній моделі наявні класи, які можна комбінувати для визначення вибірки і розміщення результатів.

Отже, закономірною є думка багатьох дослідників та розробників OLAP- систем, що нове покоління засобів інтелектуального аналізу даних розробляється саме для мережевих користувачів і оптимізоване для застосування в Інтернеті (у зв’язку з чим йдеться про інтернет-аналітику). Ці засоби дають можливість забезпечувати даними й інструментами їх аналізу користувачів мережі, для чого застосовуються особливі технології зберігання багатовимірних даних. До сьогодні практично всі провідні постачальники засобів аналізу даних пропонують системи з тонким клієнтом, орієнтовані на роботу в Інтернеті. У деяких з цих продуктів тонкий клієнт використовується для доступу до даних “традиційних” OLAP-серверів, в інших — для обробки багатовимірних кубів, розміщених на web-сайтах. Такі куби можна скачати з сайта або переслати електронною поштою і помістити на локальний або мережевий диск. Обсяг даних, вміщених у мікрокуб, стискається приблизно в 40 разів. Ефект

досягається за рахунок нетрадиційних для OLAP-підходів до зберігання даних. По-перше, у мікрокубах, не розміщуються заздалегідь підраховані агрегати. Вони замінюються спеціальними механізмами індексування, які дають змогу обчислювати агрегати “на льоту”. По-друге, значення всіх вимірювань зберігаються в мікрокубах в єдиних екземплярах. Нарешті, весь їх вміст архівується.

Одним з найвідоміших прикладів реалізації технологій інтелектуального аналізу даних в Інтернеті є проект провідного американського виробника засобів аналізу компанії Cognos для технологічної біржі Nasdaq. Російський аналог цього рішення — сервісний аналітичний центр для комерційних банків Banklist.ru, що використовує ПО компанії IntersoftLabs.

Рішення Nasdaq&AmexCompaniesEdition надає аналітичний сервіс учасникам біржі Nasdaq. Передплатники Nasdaq регулярно електронною поштою отримують локальні куби з біржовими даними. У цих кубах у розрізі індексу Nasdaq, ринкової вартості компаній, річних доходів та інших вимірювань зберігаються показники ліквідності, прибутку та інші дані про діяльність понад 5000 компаній зі списку Nasdaqі Amex, акції яких є предметом біржового торгу. Біржові дані поміщаються в куби за допомогою генератора CognosPowerplay. Для OLAP-аналізу кубів і випуску звітів передплатникам надається спеціально конфігурований для роботи з кубами OLAP-клієнт Cognosі пакет OLAP-звітів для перегляду і аналізу кубів.

Мета проекту Banklist.ru — забезпечити російським банкам, підприємствам та іншим зацікавленим особам та організаціям швидкий доступ до достовірної фінансової звітності комерційних банків і можливість її аналізу. Це необхідно для підвищення довіри до російських банків як їх клієнтів, так і діючих та потенційних партнерів. Проект створений за ініціативою Асоціації російських банків і реалізований під патронатом ЦБ РФ. Готове рішення організоване як web-сайт, база даних якого містить мікрокуби аналітичної платформи “Контур”. У цих мікрокубах зберігається фінансова інформація: баланси, звіти про прибутки і збитки, розрахунок обов’язкових економічних нормативів діяльності кредитних організацій. Вимірюваннями кубів є регіони, банки, балансові рахунки 1-го і 2-го порядку, статті доходів і витрат, банківські нормативи, а показниками — розгорнуті залишки за рахунками, суми доходів, витрат і нормативів. Достовірна звітність 733 російських банків надходить з офіційного сайту ЦБ РФ.

Мікрокуби створюються генератором “Контур CubeMaker”. Передплатникам сервісу Banklist.ru надається аналітична складова — OLAP-клієнт “Контур OlapBrowser”. Це спеціальний web-браузер для OLAP-аналізу, який забезпечує доступ до мікрокубів через Інтернет і в локальній мережі, перегляд і аналіз даних OLAP-кубів, випуск звітів. З

його допомогою можна знайти і відкрити для аналізу будь-який мікрокуб безпосередньо з web-сайта або з диска.

Обидва розглянуті рішення належать до класу OLAP-клієнтів із локальним кубом, тобто забезпечують відкладений (off-line) аналіз одного разу підготовлених файлів із даними на ПК користувача. Куби для аналізу можуть зберігатися на web-сайті (варіант Banklist.ru) або на ПК користувача, як у рішенні для Nasdaq.

Принципова відмінність у технологічній реалізації цих проектів — розміщення метаданих. У випадку Nasdaq користувачі отримують персональну OLAP-систему, жорстко конфігуровану для роботи з PowerCubeNasdaq. Метадані зберігаються у файлах OLAP-звітів, налаштованих на куби Nasdaq. Звіти встановлюються разом зі спеціальною запускаячою програмою CognosSpecialEditionLauncherі OLAP-клієнтом CognosPowerplaySpecialEdition. Це так званий “спеціальний випуск” продуктів Cognos, розроблений для проекту Nasdaq. Передвстановлена конфігурація робочого місця користувача системи аналізу біржових даних Nasdaq є істотним плюсом цього підходу.

Контур-мікрокуб містить метадані (опис вимірювань і показників, механізму агрегації і сортування даних, аналітичних інтерфейсів і т. ін.) в самому собі. Такий підхід універсальний: зміна складу і даних мікрокубів не потребує оновлення клієнтського ПО. Один і той самий OLAP-клієнт може застосовуватися для аналізу OLAP-кубів різних прикладних сервісів. Це створює передумови для тиражування технології в різних галузях і знижує вартість володіння нею для користувачів.

3 НЕЙРОКОМП'ЮТЕРНІ ТЕХНОЛОГІЇ ТА МЕРЕЖІ

3.1. Поняття та можливості нейрокомп'ютерних технологій

Впровадження нових наукомістких технологій у комерційній сфері — досить непроста справа, що потребує, окрім грошей і часу, ще й деякої зміни психології. Практика свідчить, що ці вкладення окупаються і виводять компанію на якісно новий рівень. У такій діяльності часто доводиться розв'язувати різні задачі, причому їх формулювання неформальне, а розв'язок неоднозначний. Відомо, що дохід компанії певним чином залежить від якості вирішення цих задач. Навіть якщо чіткий алгоритмічний підхід застосувати не вдається, а отримати точний розв'язок принципово неможливо, існують інші ефективні способи розв'язання. Важливе місце серед них належить нейрокомп'ютерним технологіям та нейронним мережам.

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 некласифікованої 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 всюди, де потрібно розв'язувати задачі прогнозування, класифікації, управління — іншими словами, в галузі людської діяльності, де є задачі, що погано алгоритмізуються, для розв'язання яких необхідні або постійна робота групи кваліфікованих експертів, або адаптивні системи автоматизації, якими і є нейронні мережі.

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

Для неформалізованих задач нейромережеві моделі здатні на

порядок перевершувати традиційні методи розв'язання. Але використання нейронних мереж доцільне, якщо:

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

Нейронні мережі найкраще виявляють себе там, де є велика кількість вхідних даних, між якими існують неявні взаємозв'язки і закономірності. У 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 під час попереднього аналізу або відбору, виявленні “випадних фактів” або грубих помилок людини, що приймає рішення. Доцільно використовувати нейромережеві методи в задачах з неповною або “зашумленою” інформацією, особливо в задачах, які можна розв'язати інтуїтивно, і при цьому традиційні математичні моделі не дають бажаного результату.

Методи нейронних мереж можуть використовуватися незалежно або слугувати доповненням до традиційних методів статистичного аналізу, більшість з яких пов'язані з побудовою моделей, основаних на тих або інших припущеннях і теоретичних висновках (наприклад, що досліджувана залежність є лінійною або деяка змінна має нормальний розподіл). Нейромережевий підхід не пов'язаний з такими припущеннями: він однаково придатний для лінійних і складних нелінійних залежностей, особливо ефективний у 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 не виявляється традиційними кореляційними методами, то нейронна мережа здатна автоматично налаштуватися на нього із заданою мірою точності. Крім того, сучасні нейронні мережі володіють додатковими можливостями: вони дають змогу оцінювати порівняльну важливість різних видів вхідної інформації, зменшувати її обсяг без втрати істотних даних, розпізнавати симптоми наближення критичних ситуацій і так далі.

Штучні нейронні мережі породжені біологією, оскільки складаються з елементів, функціональні можливості яких аналогічні більшості елементарних функцій біологічного нейрона. Ці елементи організуються за способом, який може відповідати (або не відповідати) анатомії мозку. Незважаючи на таку поверхневу схожість, штучні нейронні мережі демонструють певні властивості мозку. Наприклад, вони вчаться на основі досвіду, узагальнюють попередні прецеденти та нові випадки і витягують

істотні властивості з інформації, що надходить та містить зайві дані. Але однаковою мірою було б невірно ігнорувати дивну схожість у функціонуванні деяких нейронних мереж із 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 не за горами.

Для повного розуміння сутності нейрокомп'ютерних технологій розглянемо як мозок обробляє інформацію. Мозок складається з клітин двох типів: гліальних і нейронів. І хоча роль її 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. Але є й інші погляди. Зокрема, такі учені як Р. Пенроуз і С. Хамерофф вважають, що головні події відбуваються не в нейронній мережі, а в клітинах, а саме в їх цитоскелетоні, в так званих мікротрубочках. На їхню думку, і пам'ять, і навіть свідомість визначаються конформаційними змінами білків у внутрішньоклітинних структурах і пов'язаних із ними квантовими ефектами.

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

Внутрішньоклітинний простір нейрона має негативний електричний потенціал відносно позаклітинного (-70 mV), тобто клітина в цілому поляризована. Поляризація виникає за рахунок вибіркової проникності клітинної мембрани для іонів натрію і калію, що призводить до різної їхньої концентрації усередині і поза клітиною. Проте, якщо зовні досить сильно змінити потенціал мембрани одного нейрона (передавача) поблизу виходу аксона з його клітинного тіла, то проникність мембрани змінюється і перерозподіл іонів у внутрішньоклітинному і позаклітинному просторі аксона приводить до поширення по ньому хвилі короткочасної депольризації.

Електричний імпульс, поширюючись всіма розгалуженнями закінчення аксона зі швидкістю близько 100 м/с, досягає синапсів, розташованих у місцях його контакту з дендритом або сомою інших клітин. Під впливом цього імпульсу в синапсах виділяються спеціальні хімічні речовини — нейромедіатори, які, перетинаючи синаптичну щілину, взаємодіють з мембраною нейроприймача і змінюють її потенціал. Таким чином дія передається від одного нейрона до інших. Зазначимо, що ця дія може бути як збуджувальною — такою, що сприяє подальшій генерації хвилі депольризації в нейроні-приймачі, так і такою, що інгібує

— 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 зі 10^4 іншими нейронами. Тому кількість зв'язків у мозку оцінюється в 10^{14} — 10^{15} . До нині 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 мозку 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 клітину (тобто не перевищують порогу активації). Проте при будь-якому припущенні модель мережі нейронів, що взаємодіють, виявляється винятково багатомірною і такою, що має властивості, які можна зіставити з реальними можливостями мозку.

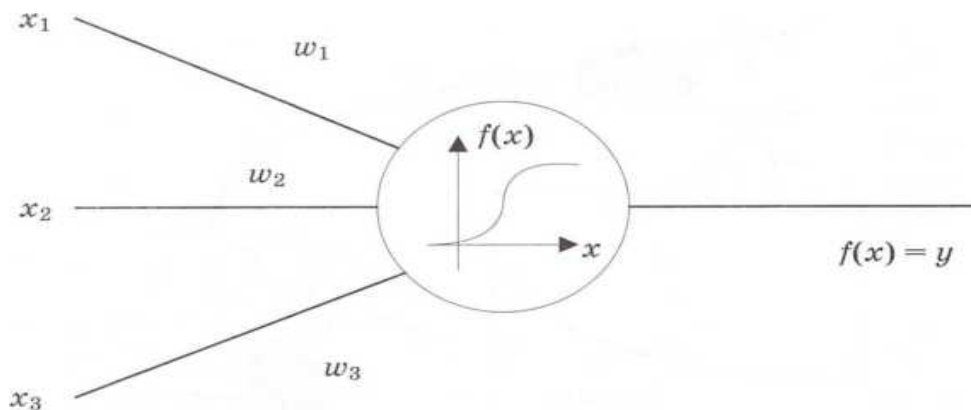


Рисунок 3.1 - Модель нейрона

На рисунку зображена модель нейрона з трьома входами (дендритами), причому синапси цих дендритів мають ваги w_1 , w_2 , w_3 . Нехай до синапсів надходять імпульси сили x_1, x_2, x_3 відповідно, тоді після проходження синапсів і дендритів до нейрона надходять 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. Однією з найбільш поширених є нелінійна функція з насиченням, так звана логістична функція або сигмоїд (функція S-

подібного вигляду).

При зменшенні a сигмоїд стає пологішим і при $a = 0$ перетворюється на горизонтальну лінію на рівні 0,5, при збільшенні a сигмоїд наближається за зовнішнім виглядом до функції одиничного стрибка з порогом T у 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 наявні традиційні обчислювальні машини. Вони лише заповнюють ті можливості, для яких не вдається побудувати формальних алгоритмічних схем. Подібно до того, як у людському мозку ліва і права півкулі працюють спільно, сучасні інформаційні системи повинні використовувати симбіоз традиційних комп'ютерів і нейротехнологій для повноцінного та продуктивного інтелектуального аналізу інформації.

Як відомо, є дві парадигми обробки інформації — “логічна” і “образна”. Перша домінує в комп'ютерах, друга — закладена в основі роботи мозку, хоча людський мозок відрізняється від мозку тварин наявністю обох компонентів мислення. Для повноцінного існування в навколишньому світі цінні обидва способи обробки інформації. І вони рано чи пізно виникають — як у ході біологічної еволюції, так і в процесі еволюції комп'ютерів, але, що характерно, в різній послідовності. Біологічна еволюція відбувалася “від образів — до логіки”. Комп'ютери ж, навпаки, почавши з логіки, лише через декілька десятиліть починають опановувати розпізнавання образів.

Знижує ефективність людино-машинного симбіозу нездатність сучасних ЕОМ оперувати сенсорною або більш загальною образною інформацією. Людина, мозок якої орієнтований саме на таку 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 дає змогу їм звільнити людей від їх теперішніх обов'язків “універсальних маніпуляторів” і контролерів при машинному виробництві. Відсутність сенсорного сприйняття світу комп'ютерами робить доступні їм моделі світу безпорадними. А втім уже зараз вартість комп'ютерної обробки інформації і людського мислення майже зрівнялися. У мозку людини близько 10^{10} нейронів, з яких одночасно активізовані приблизно 10^8 , працюючих із характерною частотою 10^2 Гц. Взяти зарплату “білого комірця” 30 000 дол. США на рік, отримаємо оцінку вартості обробки інформації людиною:

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

Обчислення на сучасному персональному комп'ютері продуктивністю

10' оп/с при амортизації близько 300 дол. на рік коштують лише на порядок менше:

$$\cdot \text{Універсальний процесор: } (10^7 \text{ оп/с} \cdot 3 \cdot 10^7 \text{ с/рік}) : (300 \text{ дол./рік})$$

$= = 10^{12}$ оп/дол.

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

• *Спеціалізований процесор:* $(10^8 \text{ оп/с} \blacksquare 3 \cdot 10^1 \text{ с/рік}) : (300 \text{ дол./рік})$
 $= = 10^{18}$ оп/дол., приблизно порівнюється 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 у розв'язанні не алгоритмізованих задач — обробці образів. Це і є те проблемне місце, яке звужує можливі галузі застосування комп'ютерів і, відповідно, — інтелектуальних аналітичних систем. Штучні нейромережі покликані “розшити” його, забезпечивши комп'ютерам здатність оперувати образною інформацією.

Нейрокомп'ютинг є новою парадигмою обчислювальних систем. Основне завдання нейрокомп'ютерів — обробка образів, заснована на навчанні, — таке саме, що і у біологічних нейросистем. Подібно до біологічних, штучні нейромережі спрямовані на паралельну обробку образів. У новій схемотехніці, як і в мозку, відсутні загальні шини, немає поділу на активний процесор і пасивну пам'ять. Обчислення, як і навчання, розподілені між всіма активними елементами — нейронами, кожний з яких є елементарним процесором образів, оскільки виконує хоч і просту операцію, але відразу над великою кількістю входів. Як обчислення, так і навчання повністю паралельні. У цьому перевага природних нейрокомп'ютерів. Це дає можливість розв'язувати задачі, непосильні навіть найпотужніших суперкомп'ютерів, не зважаючи на багатократну різницю в швидкодії елементної бази.

На думку Аніл К. Джейн із Мічиганського університету і фахівців Дослідницького центру ІВМДжанчанг Мао і К.М. Мохіддін, список таких задач можна класифікувати так.

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

Кластеризація/категоризація. При розв'язанні задачі кластеризації, відомої також як класифікація образів “без вчителя”, в якій немає навчальної вибірки з мітками класів. Алгоритм кластеризації ґрунтується на подібності образів і розміщує схожі 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, спотвореного шумом. Завдання апроксимації полягає в пошуку оцінки невідомої функції. Апроксимація функцій необхідна при розв'язанні численних інженерних і наукових задач моделювання. Типовим прикладом є шумозаглушення при прийомі сигналу різної природи, незалежно від переданої інформації.

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

Оптимізація. Численні проблеми в математиці, статистиці, техніці, науці, медицині й економіці можуть розглядатися як проблеми оптимізації. Завданням алгоритму оптимізації є пошук розв'язку, що задовольняє систему 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 вмісту пам'яті. Більше того, якщо допущена помилка в обчисленні адреси, то може бути знайдена абсолютно інша інформація. Асоціативна пам'ять доступна за вказівкою заданого змісту. Вміст пам'яті може бути викликаний навіть за частковим входом або спотвореним змістом. Асоціативна пам'ять може отримати застосування при створенні мультимедійних інформаційних баз даних.

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

Наголосимо, що в загальному випадку найчастіше трапляються ситуації, що є комплексом визначених раніше.

Розглянемо парадигми нейрокомп'ютерів, тобто родові риси, що об'єднують 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.

Відмітною 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. 3.2), що ілюструє найбільш поширену в сучасних додатках архітектуру багат шарового перцептрона. У цьому основна відмінність формальних нейронів від базових елементів послідовних ЕОМ — логічних вентилів, що мають лише два входи.

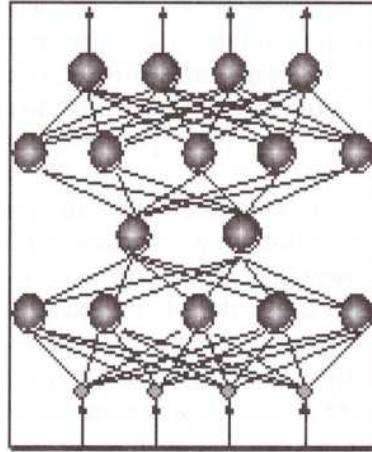


Рисунок 3.2 - Глобальний зв'язок у штучних нейромережах

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 із своїми ж вагами, що локально зберігаються, і проводить з їх сумою нелінійне перетворення.

Нелінійність вихідної функції активації принципова. Якби нейрони були лінійними елементами, то будь-яка послідовність нейронів також виконувала 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 нейрообчислень, необхідний для ефективної обробки образів, забезпечується локальністю обробки інформації в нейромережах. Кожен нейрон реагує лише на локальну інформацію, що надходить до нього у певний момент від пов'язаних із ним таких самих нейронів, без апеляції до загального плану обчислень, звичайною для універсальних ЕОМ. Таким чином, нейромережеві алгоритми локальні, і нейрони здатні функціонувати паралельно.

- Програмування: навчання, що ґрунтується на даних.

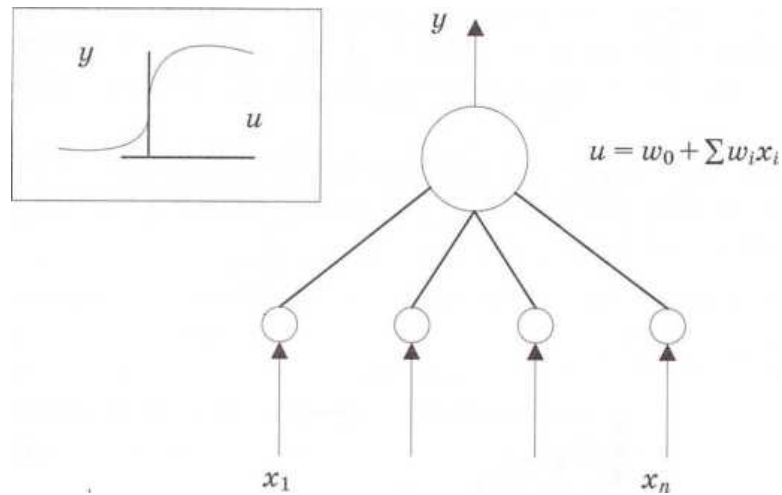


Рисунок 3.3 - Нелінійна операція над лінійною комбінацією входів

Відсутність глобального плану обчислень в нейромережах припускає і особливий характер їх програмування. Воно також має локальний характер: кожен нейрон змінює свої параметри — синаптичні ваги — відповідно до локальної інформації, що надходить до нього, про ефективність роботи всієї мережі як цілого. Режим поширення такої інформації мережею і відповідної до неї адаптації нейронів має навчальний характер. Такий спосіб програмування дає можливість ефективно враховувати специфіку потрібного від мережі способу обробки даних, бо алгоритм не задається наперед, а породжується самими даними — прикладами, на яких мережа навчається. Саме так у процесі самонавчання біологічні нейромережі виробили такі ефективні алгоритми обробки сенсорної інформації.

Характерною особливістю нейромереж є їх здатність до узагальнення, що дає змогу навчати мережу на мізерній частці всіх можливих у процесі функціонування ситуацій. У цьому їх різюча відмінність від звичайних ЕОМ, програма яких повинна наперед передбачати їх поведінку в усіх можливих ситуаціях. Ця ж їх здатність дає можливість кардинально здешевити процес розробки додатків.

- Універсальність навчальних алгоритмів.

Привабливою рисою нейрокомп'ютингу є єдиний принцип навчання нейромереж мінімізація емпіричної помилки. Функція помилки, що оцінює цю конфігурацію мережі, задається ззовні — залежно від того, якою є мета 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 виконує покладене на неї завдання.

Не вдаючись до математичних тонкощів, образно цей процес можна уявити як пошук мінімуму функції помилки $E(u)$, залежної від набору всіх синаптичних ваг мережі w (рис. 3.4).

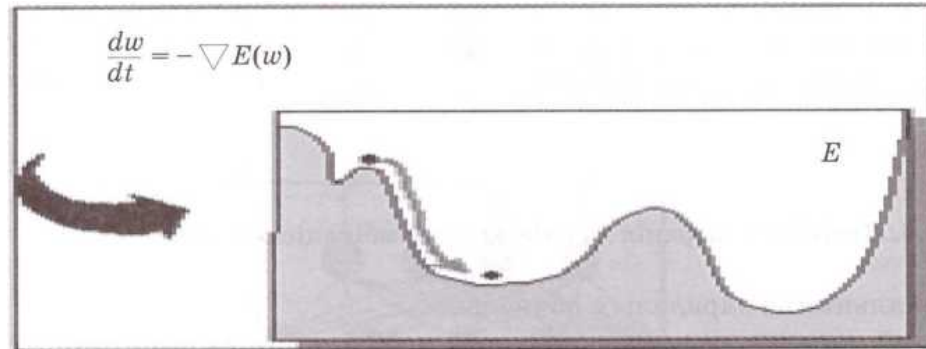


Рисунок 3.4 - Навчання мережі як задача оптимізації

Базовою ідеєю всіх алгоритмів навчання є облік локального градієнта у просторі конфігурацій для вибору траєкторії якнайшвидшого спуску по функції помилки. Функція помилки, проте, може мати безліч локальних мінімумів, що представляють субоптимальні розв'язки. 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'ютингу в загальний процес еволюції комп'ютерів, отримуємо можливість зазирнути у найближче майбутнє, екстраполюючи сьогоденні тенденції.

1. “Розумні” нейрочипи. 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'ютингу почнеться з виходом на ринок широкого асортименту апаратних засобів — спеціалізованих нейрочипів для обробки зображень, мови, аналітики та сенсорної інформації. Вже сьогодні є, наприклад, дверні системи, що розпізнають власника за зовнішнім виглядом, голосом, і мабуть, запахом у сукупності. Системи життєзабезпечення житла стануть адаптивними і зможуть навчатися. Всі побутові прилади порозумнішають і набудуть здатності “здогадуватися”, що від них потребується саме в цей момент. Сенсорні датчики почнуть реагувати, а регулюючі системи — відчувати. Розумні контролери, що розпізнають потенційно небезпечні ситуації і уміють приймати адекватні превентивні рішення, набудуть поширення в складних електричних і теплових мережах. На них ґрунтуватимуться системи регулювання транспортними потоками і потоками даних у комп'ютерних мережах і стільниковому зв'язку.

2. Операційні системи нової архітектури. Проте найбільші зміни стосуватимуться самих комп'ютерів. На думку Білла Гейтса, глави відомої компанії “Microsoft”, висловлену ним на зборах ради директорів у 2005 р., через 10 років 90 % операційних систем буде зайнято вирішенням завдань

розпізнавання образів. Отже, при проектуванні майбутніх поколінь комп'ютерів нейрокомп'ютинг виходить на перший план. Можна навіть уявити зразковий сценарій проникнення нейросистем у комп'ютери майбутнього, пов'язаний з розвитком глобальної мережі Інтернет. Зараз саме вона спрямовує розвиток комп'ютерних систем, поступово перетворюючи розрізнену мережу персональних комп'ютерів, робочих станцій і мейнфреймів у єдиний світовий мережевий комп'ютер із необмеженими інформаційними ресурсами. І подібно до того, як епоха Великих відкриттів XV ст. стимулювала розвиток астрономії і точної механіки для вдосконалення навігаційних приладів, опанування нового інформаційного океану потребує 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 чільне місце у 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, нині найкоротша дорога до створення автономних електронних секретарів, що легко навчаються. Природно передбачити, що саме цей напрям через його стратегічну важливість у найближчому майбутньому буде найбільш прогресивним.

Розглянемо сучасний стан нейрокомп'ютингу та нейротехнологій.

- Елементна база нейрокомп'ютерів. Практично всі нейрокомп'ютери, що діють, використовують традиційну елементну базу: мікроелектронні СБІС — (*рос.* — насверхбольшие интегральные схемы; надвеликі інтегральні схеми НВІС). Сотні мільярдів доларів, які вже вкладені в розвиток цієї технології, надають їй вирішальну перевагу над іншими альтернативами, такими як, наприклад, оптичні обчислення. Сучасна електроніка спирається, в основному, на цифрову обробку сигналів, стійку до перешкод і технологічних відхилень у параметрах базових елементів. Цифрова схемотехніка надає нейроконструкторам багатий інструментарій. Тому не дивно, що найбільш поширеними є саме цифрові нейрокомп'ютери. Це, по суті, спеціалізовані матричні прискорювачі, що використовують матричний, пошаровий характер обробки сигналів у нейромережах. Широко використовуються стандартні процесори обробки сигналів (DigitalSignalProcessors, DSP), оптимізовані під такі операції. Прикладом сучасного DSP-процесора, пристосованого для прискорення нейрообчислень, є продукт TexasInstrumentsTMS320C80, продуктивністю 2 млрд операцій на секунду. Цей кристал включає п'ять процесорів і реалізує відразу дві технології — DSPi RISC (рис. 3.5).

Проте сама природа нейромережевої обробки інформації — аналогова, і додаткового виграшу в швидкості обчислень (за деякими оцінками 10^3 — 10^4) та щільності обчислювальних елементів можна

домогтися, використовуючи спеціалізовану аналогову елементну базу. Найбільш перспективні аналогові мікросхеми з локальними зв'язками між елементами (т. зв. клітинні нейромережі (CellularNeuralNetworks, CNN)), наприклад силіконова ретина фірми Synaptics. З іншого боку, розробка аналогових чипів з використанням нетрадиційних рішень схемотехнік потребує додаткових і чималих витрат. Нині ці роботи розгорнуті широким фронтом, наприклад у рамках проекту SCX-1 (SiliconCortex— кремнієва кора) (рис. 3.6).

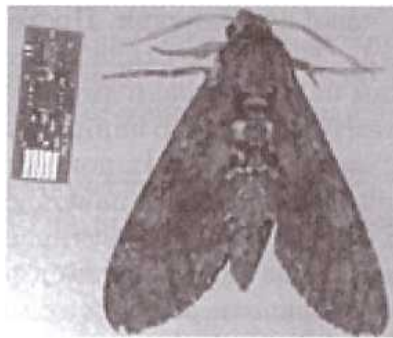


Рисунок 3.5 - Сучасні цифрові нейрокомп'ютерні елементи

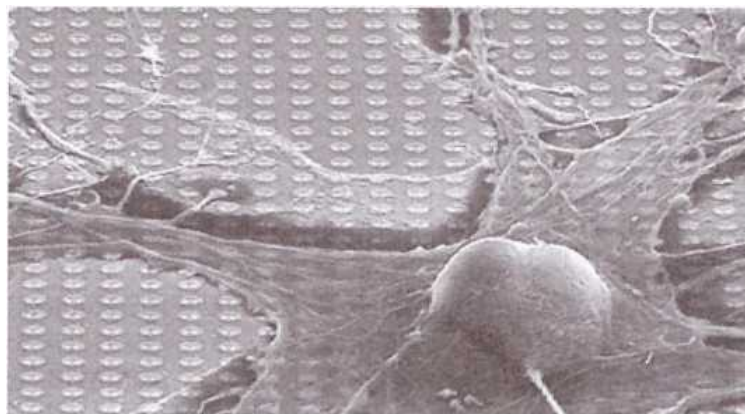


Рисунок 3.6 - Сучасні аналогові нейрокомп'ютерні елементи

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'ютингу полягає в можливості організовувати масові паралельні обчислення. Тому базові процесорні елементи зазвичай поєднуються в обчислювальні комплекси: якомога більше — на одному чипі, а що не вмістилося — в мультипроцесорні плати. Ці плати потім або вмонтовують у персональні комп'ютери і робочі станції як нейроприскорювачі, або збирають у повномасштабні нейрокомп'ютери.

- Порівняння вартості звичайних і нейрообчислень. Продуктивність сучасних персональних комп'ютерів становить приблизно 10' операцій. Отже, при вартості всього на порядок більше звичайних ПК, нейроприскорювач у декілька сотень разів перевершує їх у швидкості. Тобто питома вартість сучасних нейрообчислень приблизно на порядок нижча, ніж у традиційних комп'ютерів. Це лише наслідок спеціалізації матричних процесорів (DSP), що мають ту саму елементну базу, що й універсальні мікропроцесори. Проте вираш на один порядок у вартості обчислень нечасто стає вирішальним аргументом для використання спеціалізованої апаратури, пов'язаної з додатковими витратами, у тому числі на навчання персоналу. Тому реально нейрокомп'ютери використовуються у спеціалізованих системах, коли потрібно навчати і постійно перенавчати сотні нейромереж, об'єднаних у єдині інформаційні комплекси, або в системах реального часу, де швидкість обробки даних є критичною.

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

Переваги таких “віртуальних” нейрокомп'ютерів для відносно невеликих завдань очевидні. По-перше, не треба витрачатися на нову апаратуру, якщо можна завантажити вже наявні комп'ютери загального призначення. По-друге, користувач не повинен опановувати 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 мають обмежений набір “захитих” у 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 нейропакетів, що є на ринку, мають дружній інтерфейс користувача, не потребують знайомства з мовами програмування. Безліч нейроемуляторів початкового рівня можна знайти в Інтернеті як *shareware* або *freeware*. Це, зазвичай, багатосарові персептрони з одним або декількома правилами навчання. Винятком є повністю професійний Штутгартський симулятор із великим набором можливостей, працює він лише на UNIX- машинах. Комерційні пакети відрізняються від поширюваних набором засобів імпорту і

передоброби даних, додатковими можливостями з аналізу значущості входів і оптимізації структури мережі. Як правило, такі пакети (BrainMakerProfessional, NeuroForecaster, Лора-IQ300) мають власний вбудований блок передоброби даних, хоча інколи з цією метою зручніше використовувати стандартні електронні таблиці. Так, нейропродукти групи нейрокомп'ютингу ФІАН вбудовується безпосередньо в MicrosoftExcel 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 й експорту даних. Такі пакети націлені на розв'язання інформаційних задач у діалоговому режимі — за безпосередньою участю користувача.

Інструменти розробки нейрододатків. Головне, що відрізняє цей клас програмного забезпечення — здатність генерувати “відчужувані” нейромереві продукти, тобто генерувати програмний код, що використовує навчені нейромереві для аналізу даних. Такий код може бути вбудований як підсистема в будь-які складні інформаційні комплекси. Прикладами подібних систем, здатних генерувати вихідні тексти програм, є NeuralWorksProfessionalPlus фірми NeuralWareі NeuralBench. Остання цікава, окрім іншого, тим, що може генерувати коди багатьма мовами, включаючи Java. Такі Java-аплети можуть використовуватися для організації різних сервісів у глобальних і локальних мережах. Зручним інструментом розробки складних нейросистем є MATLAB з нейромеревим інструментарієм, що додається до нього. MATLAB надає зручне середовище для синтезу нейромеревих методик з іншими методами обробки даних (wavelet-аналіз, статистика, фінансовий аналіз.

Готові рішення на основі нейромерев. У таких системах нейромереві приховані від користувача в надрах готових автоматизованих комплексів, призначених для вирішення конкретних аналітичних завдань. Наприклад, уже згадуваний продукт Falcon вбудовується в банківську автоматизовану систему обслуговування платежів за пластиковими картками. В 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 мають унікальні можливості і забезпечують реальні конкурентні переваги, їх вартість може бути досить висока. Іспанська компанія SEMP займається підвищенням ефективності обслуговування кредитних карт VISA, що емітуються іспанськими банками. Кількість подібних транзакцій — від 500 000 до 1 000 000 на день. Нейромерева система, розроблена для неї вченими з Мадридського інституту інженерії знань (InstitutodeIngenieriadelConocimiento), зменшила вірогідність несанкціонованого використання карт на 30—40 % для основних каналів шахрайства.

• Нейромережевий консалтинг. Опис ринку нейропродуктів буде не повним без згадки про нейроконсалтинг. 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. У разі великої кількості клієнтів ціна таких передбачень може бути помірною. Наприклад, PredictionCompany, заснована в 1991 р. фізиками Дойном Фармером і Норманом Паккардом, — фахівцями в галузі динамічного хаосу. Продукція цієї компанії користується великим успіхом серед швейцарських банків, що скуповують прогнози для гри на фондових і валютних ринках.

• Ринок нейропродукції. Обсяг ринку нейропродукції, його структура зростає стрімкими темпами: за різними оцінками — від 30 до 50% на рік, перевищивши нещодавно декілька мільярдів доларів. Таке ж зростання спостерігалось на початку 80-х років на ринку персональних комп'ютерів. В результаті виниклої конкуренції за гроші кінцевого користувача змінився весь комп'ютерний світ. Нейрокомп'ютери та їх програмні емулятори цікаві не самі по собі, а як інструмент вирішення практичних завдань. Лише в цьому випадку нейропродукція володітиме споживчою вартістю і матиме відповідний обсяг ринку.

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

Контроль операцій з кредитними картками. Здатність нейромереж до класифікації застосовують для відстежування операцій з краденими кредитними картками і підробленими чеками. Спеціалізована система Falcon фірми HNC дає змогу за частотою операцій і характером покупок виокремити підозрілі операції і сигналізувати про це. Ця система використовує запатентовану уніфіковану технологію підтримки ухвалення рішення (UnifiedDecisionsTechnology), 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 дають можливість з високою мірою достовірності визначати потенційні випадки шахрайства з банківськими картками за географічним розташуванням або поштовим індексом отримувача картки. 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. Нині Falcon контролює більше 260 мільйонів рахунків 16 найбільших емітентів кредитних карток.

Аналогічна система, розроблена фірмою ITC, використовується для обробки операцій з кредитними картками Visa. У 2005 р. за допомогою цієї системи запобігли нелегальним операціям на суму понад 100 млн дол.

Родина систем PRISM, розроблена компанією Nestor, основана на використанні нейромереж, експертних систем і статистичних методів для виявлення в реальному режимі часу шахрайства з кредитними і дебетовими картками, а також інших типів шахрайства при здійсненні фінансових або торговельних операцій. Нейромережа, використовувана в системах родини PRISM, була навчена на основі більш ніж півмільйона транзакцій з різними типами карт. За деякими оцінками, використання систем родини PRISM дає змогу зменшити кількість шахрайств на 50 %.

Нейромережі на фінансових ринках. Американський Citibank використовує нейромережеві передбачення з 1990 р. У 2002 р., за даними журналу The Economist, автоматичний дилінг показував прибутковість на рівні 25 % річних, що набагато перевищує показники більшості брокерів. ChemicalBank використовує нейросистему фірми NeuralData для попередньої обробки транзакції на валютних біржах 23 країн, фільтруючи “підозрілі” операції. Fidelity of Boston використовується нейромережами в управлінні портфелями з сумарним обсягом 3 млрд дол. Повністю автоматизовані системи ведення портфелів із використанням нейромереж застосовують, наприклад Deere & Co — на суму 100 млн дол. і LBSCapital на суму 400 млн дол. В останньому випадку експертна система об'єднується з приблизно 900 нейромережами.

Компанія Alela Corp. займається прогнозуванням зміни біржових індексів. Для передбачення ознак зміни індексів застосовується нейронна мережа, що використовує РБФ (радіальні базисні функції). На сайті компанії можна безкоштовно отримати прогнози зміни індексів Dow Jones, S&P500 і Merval, а також переконатися, що частка правильних передбачень становить не менше 80 %.

Аналіз страхових позовів. Фірмою NeuralInnovation Ltd. створена нейромережева система ClaimFraudAnalyser, що дає змогу миттєво виявляти підозрілі страхові позови, що стосуються пошкоджених автомобілів. На входи системи подаються такі параметри, як вік і досвід водія, вартість автомобіля, наявність схожих випадків у минулому та ін. У результаті обробки видається число — вірогідність того, що цей позов пов'язаний з шахрайством. Така система дає можливість не лише заощадити за рахунок виявлення фальсифікацій, а 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 продуктів харчування, що має величезні ринки збуту. Одним з основних маркетингових механізмів замовника є поширення купонів, що дають право придбання певного товару з знижкою. Оскільки витрати на розсилання купонів досить великі, вирішальним чинником є ефективність розсилки, тобто частка клієнтів, що скористалися знижкою. Для підвищення ефективності купонної системи важливо провести попередню сегментацію ринку, а потім адресувати клієнтам кожного сегмента саме ті купони, якими вони з більшою вірогідністю скористаються. У термінах інтелектуального аналізу даних потрібно було розв'язати 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 обсягу продажу для кожного сегмента.

Інший варіант розв'язку цієї задачі обрала компанія GoalAssistCorporation, виконуючи замовлення великої маркетингової фірми. Потрібно було дослідити стратегію заохочувальних товарів (коли, наприклад, надіславши 5 етикеток від чипсів, клієнт отримує безкоштовно футболку) для певної компанії, яка торгувала харчовими продуктами. Звичайні методи прогнозування реакції споживачів виявилися в цьому випадку недостатньо точними. У результаті попит на футболки був дуже великий і багатьом покупцям довелося довго чекати здобутого призу, тоді як інші подарунки залишилися незатребуваними. Аби підвищити точність прогнозування, було вирішено використовувати історичні дані і нейронні мережі. Компанія GoalAssistCorporation побудувала дві нейромережі для вирішення цього завдання. Перша з них — це мережа з адаптивною архітектурою пакета NeuroShellClassifier компанії WardSystemsGroup, на входи якої подавалися різні параметри товарів і рекламної політики. За допомогою цієї мережі, призначеної спеціально для класифікації, було отримано поділ входів на 4 класи, що характеризують відгук споживачів. Ті ж входи разом з відповіддю першої мережі подавалися далі на вхід пакета NeuroShellPredictor, який також містить складну мережу, пристосовану для задач кількісного прогнозування. Середня помилка передбачень становила всього близько 4 %. Побудова цієї моделі зайняла близько 120 годин, також був потрібен час на передобробку вхідних даних. Експерти GoalAssistCorp. вважають, що ця модель і далі успішно застосовуватиметься для розв'язання маркетингових задач.

Компанія NeuralInnovationLtd. використовує при роботі з маркетинговими компаніями конкретну стратегію прямої розсилки. Спочатку розсилається 25 % від загальної кількості пропозицій і збирається інформація про відгуки споживачів. Потім ця інформація надходить на вхід нейрокомп'ютера, який здійснює пошук оптимального

сегмента споживчого ринку для цього товару. Останні 75 % пропозицій розсилаються у зазначений сегмент. При цьому ефективність розсилки істотно зростає.

Доктор Аль Бехренс, співробітник NorthernNaturalGas в штаті Небраска, навчив нейронну мережу передбачати зміну цін на газ у наступному місяці з середньою точністю 97 % . Щомісячна ціна інколи прив'язується до індексації цін у друкованих виданнях (InsideFERCі NaturalGasWeek), включаючи нещодавню ринкову діяльність компанії, сезонні чинники, погоду тощо.

Відома корпорація Microsoftтакож використовує програмні продукти на основі нейромереж у своїй маркетинговій політиці. Щороку Microsoft надсилає більше 40 млн рекламних пропозицій придбати свої продукти більш ніж 8,5 млн зареєстрованих користувачів. Велика частина спрямована на модернізацію ПО або купівлю додаткових пакетів. Мета використання ПО на основі нейронних мереж очевидна — збільшення обсягу продажу. Такими засобами було досягнуто збільшення попиту з 4,9 до 8,2 %, а витрати на рекламу знизилися на 35 %.

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

При проведенні випробувань якості бетону використовують багато методів. Одним із 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 подання звукових хвиль і прийманням відбитого сигналу, а потім обробкою нейронною мережею, фахівці з NationalInstituteofStandardsandTechnology (NIST) здатні перевірити якість бетону за товщини матеріалу до півметра.

FordMotorsCompanувпровадила 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 нейросистеми подаються дані від 31 датчика. Нейронна мережа вивчала різні види несправностей за 868 прикладами. Після повного циклу навчання якість діагностування несправностей мережею досягла рівня кращих експертів, і значно перевершувала їх у швидкості.

Медична діагностика. Компанія “НейроПроект” створила систему

об'єктивної діагностики слуху у немовлят. 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 детальної 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 у програмах медичної діагностики слугує пакет кардіодіагностики, розроблений фірмою RESInformatіcаспільно з Центром кардіологічних досліджень у Мілані. Причому для таких хвороб, як ішемія міокарда і артеріальна гіпертензія, досягається точність постановки діагнозу більш ніж 95 % . Окрім цього, одним з перспективних напрямів у 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 допоможе збільшити продуктивність більш ніж у 1000 разів, здешевити дослідження не менше ніж у 500 разів, обробити дані від практично необмеженої кількості хворих при збільшенні діагностичної точності.

Системи безпеки в аеропортах. Американська фірма SAIC (ScienceApplicationInternationalCorporation) використовувала нейронні мережі у своєму проекті TNA. TNA є ящиком вартістю 750 тис. дол., 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 на хвилину.

Активна реклама в Інтернеті. Нейромережевий продукт SelectCastфірми AptexSoftwareInc. виявляє профілі інтересів користувачів Інтернету і пропонує їм відповідним чином відфільтровану рекламу. Після установки на серверах Exciteі Infoseek, нейромережева реклама охопила близько третини всіх користувачів Інтернету. Згідно з проведеними дослідженнями, встановлено, що відгук на таку активну рекламу в середньому вдвічі вищий, ніж на звичайну, що розміщується в мережі. А на окремі види реклами відгук зріс в п'ятеро. Відзначимо, що рекламний сектор мережі переживає зараз період бурхливого розвитку. Результати першого півріччя 2007 р. свідчать про річний темп зростання 250 %, що в грошовому вираженні становить 400 млн дол.

Політичні технології. Компанія “НейроПроект” досить упевнено передбачає результати президентських виборів у США на підставі анкети з

12 питань. Причому аналіз навченої нейромережі дав змогу виявити п'ять ключових питань, відповіді на які формують два головні чинники, що визначають успіх президентської кампанії.

Наукові дослідження. Доктор Генрік Лундштедт з LundObservatory (Швеція), навчив нейронні мережі прогнозувати ефекти від сонячних спалахів, такі як збурення магнітних полів Землі. Як відомо, т. з. магнітні бурі чинять вплив не лише на стан і здоров'я людини, а і на роботу електростанцій, стаються збої в передаванні радіо і телепередач, в роботі геологічного устаткування, супутників тощо. Нейронна мережа, що враховує 37 відомих впливових чинників раз у чотири дні і аналізує їх зміни, здатна з високою мірою точності визначити “космічну погоду”. Відомий випадок, коли стандартний метод прогнозування не визначив сильних магнітних бур взагалі, а метод на основі нейронних мереж точно спрогнозував дві з трьох бур.

Автопілотований гіперзвуковий літак-розвідник. Названий LoFLYTE (Low-ObservableFlightTestExperiment) реактивний безпілотний літак завдовжки 2,5 м був розроблений для NASA і AirForce фірмою AccurateAutomationCorp. у межах програми підтримки малого інноваційного бізнесу. Це експериментальна розробка для дослідження нових принципів пілотування, включаючи нейронні мережі, що дають можливість автопілоту навчатися, копіюючи прийоми пілотування льотчика. З часом нейромережі переймають досвід управління, а швидкість обробки інформації допомагає 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'ютингу з'явилися практично одночасно з зародженням послідовних ЕОМ. Основні роботи з нейрообчислень з'явилися на два роки раніше знаменитої доповідної записки фон Неймана про принципи організації обчислень у послідовних універсальних ЕОМ. Вияв цікавості до штучних нейронних мереж був зумовлений роботами піонерів у цій справі — У. Маккалоха; У. Піттса. У 1943 р. увагу громадськості привернула робота під назвою “Логічне обчислення ідей, що належать до нервової діяльності”, в якій вони запропонували математичну модель нейрона і сформулювали принципи побудови штучних нейронних мереж, згідно з розробленою ними моделлю функціонування головного мозку. Багато учених з ентузіазмом почали пропонувати свої рішення і нову архітектуру нейронних мереж.

Паралельно з прогресом у нейроанатомії та нейрофізіології психологами були створені моделі людського навчання. Найбільш плідною

була модель Д. Хебба, який у 1949 р. запропонував закон навчання, що став початком для алгоритмів навчання штучних нейронних мереж та продемонстрував вченим того часу, як мережа нейронів може навчатися.

Перший експериментальний нейрокомп'ютер Snark був створений Мар-віном Мінським у 1951 р., проте він не був пристосований до розв'язання практично цікавих задач. Перший успіх нейрокомп'ютингу пов'язують з розробкою іншого американця — Френка Розенблатта — перцептроном (англ. *perception* — сприйняття). Перцептрон був вперше змодельований на універсальній ЕОМ IBM-704 у 1958 р., причому його навчання вимагало близько півгодини машинного часу. Апаратний варіант — Mark I Perceptron, був побудований у 1960 р. і призначався для розпізнавання зорових образів. Його рецепторне поле складалося з 400 пікселів (матриця фотоприймачів 20x20), і він успішно розв'язував низку задач, зокрема міг розрізняти деякі букви (рис. 3.7).



Рисунок 3.7 - Френк Розенблатт із Mark I Perceptron

М. Мінський, Ф. Розенблатт та інші вчені 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 швидко розвіялася. Мережі не могли вирішувати завдання, ззовні вельми схожі на ті, які вони успішно виконували. З цих нез'ясовних невдач почався період інтенсивного аналізу.

М. Мінський, використовуючи точні математичні методи, довів низку теорем, що стосуються функціонування мереж. Його дослідження були використані у книзі, в якій він разом з С. Пайпертом довів, що застосовувані у той час одношарові мережі теоретично неспроможні

вирішити багато простих завдань. М. Мінський також не був оптимістичним стосовно потенційно можливого прогресу.

Їх праця загальмувала науковий запал багатьох учених-ентузіастів практично на два десятиліття. Почався період, пов'язаний із розчаруванням у можливостях нейронних мереж, викликаний здебільшого психологічним чинником, що виявлявся в нездатності людини описати словами процес мислення. І хоча у цей час було запропоновано багато розробок і досліджень, що заслуговують на увагу, науковий світ ставився до них скептично. Дослідження в цьому напрямі були згорнуті аж до 1983 р., коли, нарешті, отримали фінансування від Агентства перспективних військових досліджень США (DARPA), що стало сигналом до початку нового нейромережевого буму.

Інтерес наукової громадськості до нейромереж прокинувся після теоретичної розвідки фізика Джона Хопфілда (1982 р.), що запропонував модель асоціативної пам'яті у нейронних ансамблях. Дж. Хопфілд і його численні послідовники збагатили теорію нейромереж великою кількістю ідей з арсеналу фізики, зокрема такими як колективні взаємодії нейронів, енергія мережі, температура навчання тощо. Проте справжній бум практичного вживання нейромереж почався після публікації в 1986 р. Девідом Румельхартом зі співавторами методу навчання багатощарового перцептрона, названого ними методом зворотного поширення помилки (*errorback-propagation*). Тоді ж з'явилися перші комерційні проекти нейрокомп'ютерів (MarkIII фірми TRW, США). Обмеження перцептронів, про які писали М. Мінський і С. Пейперт, виявилися подоланими, а можливості обчислювальної техніки — достатніми для вирішення широкого кола прикладних завдань. У 90-х роках продуктивність послідовних комп'ютерів зросла настільки, що це дало змогу моделювати за їх допомогою роботу паралельних нейронних мереж із кількістю нейронів від декількох сотень до десятків тисяч. Такі емулятори нейромереж здатні вирішувати багато важливих у практичному аспекті завдань. У свою чергу, нейромережеві програмні комплекси стануть тим носієм, який введе на технологічну зрбіту сьогодення паралельне нейромережеве *hardware*.

Подальший розвиток нейромережевих технологій підтримувався створенням спеціалізованих центрів нейрокомп'ютерів, проведенням міжнародних конференцій і форумів. Обсяг продажу комерційних нейромережевих продуктів із 1991 по 2007 р. виріс більш ніж у чотирнадцять разів. Нині, з урахуванням 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. В основі їх побудови лежить ідея про те, що нейрони можна моделювати досить простими автоматами (штучними нейронами), а вся складність мозку, гнучкість його функціонування й інші найважливіші якості визначаються зв'язками між нейронами.

Відомі такі типи нейронних мереж:

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

Як вже було зазначено, світовий ринок пропонує більше кількох тисяч нейромережових пакетів. Загальний обсяг ринку нейронних мереж до 2008 р. перевищив 10 млрд дол., 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 великого банку. Наприклад, кількість продажів одного тільки нейромережового пакета “BrainMakerPro” дорівнює обсягам продажу найпопулярнішого пакета технічного аналізу MetaStock.

У зв'язку з великою різноманітністю штучних нейронних мереж і специфікою їх реалізації важливою є класифікація їхніх архітектур для подальшого застосування в інтелектуальних системах аналізу даних.

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

У синхронних нейронних мережах у кожен момент часу свій стан змінює лише один нейрон. У асинхронних — стан змінюється відразу у цілої групи нейронів, як правило, у всього шару.

- За типом базової архітектури — *шаровані* і *повнозв'язні* мережі.

Ключовим у шарованих мережах є поняття прошарка. Шар — один або декілька нейронів, на входи яких подається один і той самий 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 у кожному шарі не пов'язане з кількістю нейронів в інших шарах і може бути довільним.

У межах одного шару дані обробляються паралельно, а в масштабах всієї мережі обробка ведеться послідовно — від шару до шару. До шарованих нейронних мереж належать, наприклад, багатошарові перцептрони, мережі радіальних базисних функцій, когнітрон, неокогнітрон, 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 складається з одного шару, а багатошарова має декілька шарів. У багатошаровій мережі перший шар називається вхідним, подальші — внутрішніми або прихованими, останній шар — вихідним. Отже, проміжні шари — це всі шари у багатошаровій нейронній мережі, окрім вхідного і вихідного. Вхідний шар мережі реалізує зв'язок із вхідними даними, вихідний — з вихідними.

Нейрони можуть бути вхідними, вихідними і прихованими. Вхідний шар організований із вхідних нейронів (*inputneuron*), які отримують дані і поширюють їх на входи нейронів прихованого шару мережі. Прихований нейрон (*hiddenneuron*) міститься у прихованому шарі нейронної мережі. Вихідні нейрони (*outputneuron*), з яких організований вихідний шар мережі, видають результати роботи нейронної мережі.

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

- За спрямованістю зв'язків нейронні мережі бувають зі *зворотними зв'язками* і *без зворотних зв'язків*.

Серед мереж без зворотних зв'язків розрізняють *мережі зі зворотним поширенням помилки* та *інші мережі*. Мережі першої групи характеризуються фіксованою структурою, ітераційним навчанням, коригуванням вагів за помилками. Перевагами мереж без зворотних зв'язків є простота їх реалізації і гарантоване одержання відповіді після проходження даних шарами. Недоліком цього виду мереж вважається мінімізація розмірів мережі — нейрони багато разів беруть участь у обробці даних. Менший обсяг мережі полегшує процес навчання.

До мереж зі зворотними зв'язками відносять мережі Хопфілда та мережі Кохонена. Перевагами мереж зі зворотними зв'язками є складність навчання, викликана великою кількістю нейронів для алгоритмів одного і того ж рівня складності. Недоліки цього виду мереж — потрібні спеціальні умови, що гарантують збіжність обчислень.

- Мережі прямого поширення і рекурентні мережі.

До мереж *прямого поширення* зазвичай відносять перцептрони,

мережу BackPropagation, мережу зустрічного поширення та карти Кохонена. Усі зв'язки спрямовані строго від вхідних нейронів до вихідних.

Характерна особливість *рекурентних мереж* — наявність блоків динамічної затримки і зворотних зв'язків, що дає їм змогу обробляти динамічні моделі, наприклад, мережа Хопфілда або мережа Елмана — мережа, що складається з двох шарів, у якій прихований шар охоплений динамічним зворотним зв'язком, що дає можливість 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 у напрямі від вхідного шару до вихідного, так і у зворотному. Класичним прикладом є нейронна мережа Косько.

- Навчання з учителем або без нього. Перед використанням нейронної мережі її необхідно навчити. Процес навчання нейронної мережі полягає у підлаштуванні її внутрішніх параметрів під конкретне завдання. Алгоритм роботи нейронної мережі є ітеративним, його кроки називають епохами або циклами.

Епоха одна ітерація у процесі навчання, що включає пред'явлення всіх прикладів із навчальної множини і, можливо, перевірку якості навчання на контрольній множині. Процес навчання здійснюється на навчальній вибірці. Вона включає вхідні і відповідні їм вихідні значення набору даних. У процесі навчання нейронна мережа знаходить деякі залежності вихідних полів від вхідних. Таким чином, постає питання — які вхідні поля (ознаки) необхідно використовувати. Спочатку вибір здійснюється евристично, потім кількість входів може бути змінена. Ускладнення може викликати питання про кількість спостережень у наборі даних. І хоч існують деякі правила, що описують зв'язок між необхідною кількістю спостережень і розміром мережі, їх правильність не доведена. Кількість необхідних спостережень залежить від рівня складності вирішуваного завдання. При збільшенні кількості ознак кількість спостережень зростає нелінійно, ця проблема має назву “Прокльон розмірності”. У разі недостатньої кількості даних рекомендується використовувати лінійну модель.

Аналітик повинен визначити кількість шарів у мережі і кількість нейронів у кожному шарі. Далі необхідно призначити такі значення вагів і зміщень, які зможуть мінімізувати помилку рішення. Ваги і зміщення автоматично підлаштовуються 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 потребує 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 нейронної 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 задаються значення вхідних і вихідних сигналів, а вона за певним алгоритмом підлаштує ваги синаптичних зв'язків. У процесі навчання проводиться коригування вагів мережі за наслідками порівняння фактичних вихідних значень з вхідними, відомими наперед.

При *навчанні без учителя* розкривається внутрішня структура даних або кореляції між зразками у наборі даних. Виходи нейронної мережі формуються самостійно, а ваги змінюються за алгоритмом, що враховує тільки вхідні і похідні від них сигнали. Це навчання називають також *некерованим*. У результаті такого навчання об'єкти або приклади розподіляються за категоріями, самі категорії та 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, за якого мережа втрачає здібність до узагальнення.

Перенавчання виникає в разі дуже тривалого навчання, недостатньої кількості навчальних прикладів або переускладненої структури нейронної мережі. Перенавчання пов'язане з тим, що вибір навчальної (тренувальної) множини є випадковим. З перших кроків навчання помилка зменшується. На подальших кроках із метою зменшення помилки (цільової функції) параметри підлаштовуються під особливості навчальної множини. Проте при цьому відбувається “підлаштування” не під загальні закономірності ряду, а під особливості його частини — навчальної підмножини. При цьому точність прогнозу зменшується. Один з варіантів боротьби з перенавчанням мережі — поділ навчальної вибірки на дві множини (навчальну і тестову). На навчальній множині відбувається навчання нейронній мережі, на тестовій 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 можна спостерігати зміну помилки прогнозу на тестовій множині паралельно зі 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.

- Одно- та 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, що утворюють шар (рис. 3.11). Зазначимо, що вершини (кола ліворуч) слугують лише для розподілу вхідних сигналів. Вони не виконують будь-яких обчислень і тому не вважаються шаром. Для більшої наочності позначимо їх колами, щоб відрізнити від обчислювальних нейронів, позначених квадратами. Кожен елемент із множини входів X окремою вагою поєднаний із кожним штучним нейроном. А кожен нейрон видає зважену суму входів у мережу. У штучних і біологічних мережах багато з'єднань можуть бути відкритими, але тут вони показані всі для демонстрації загальної картини. Можуть існувати також з'єднання між виходами і входами елементів у шарі.

Зручно вважати ваги елементами матриці W . Матриця має m рядків і n стовпців, де m — число входів, а n — число нейронів. Наприклад, w_{23} — що пов'язує другий вхід із третім нейроном. Тобто обчислення вихідного вектора N , компонентами якого є виходи ОІУТ-нейронів, зводиться до матричного множення $N = XW$, де N і X — вектори-рядки.

Більші і складніші багатошарові нейронні мережі мають, як правило, і з тликі обчислювальні можливості. Хоча створені мережі всіх конфігурацій, не тільки можна собі уявити, пошарова організація нейронів копіює шару ваті 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 наступного (рис. 3.8).

Багатошарові мережі не можуть призвести до збільшення обчислювальної потужності порівняно з одношаровою, якщо активаційна функція між шарами лінійна. Обчислення виходу шару полягає у 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, що дорівнює добутку двох вагових матриць. Отже, будь-яка багатошарова лінійна мережа може бути замінена еквівалентною одношаровою

мережею. Проте одношарові мережі дуже обмежені за своїми обчислювальними можливостями. Тобто для розширення можливостей мереж порівняно з одношаровою мережею потрібна нелінійна активаційна функція.

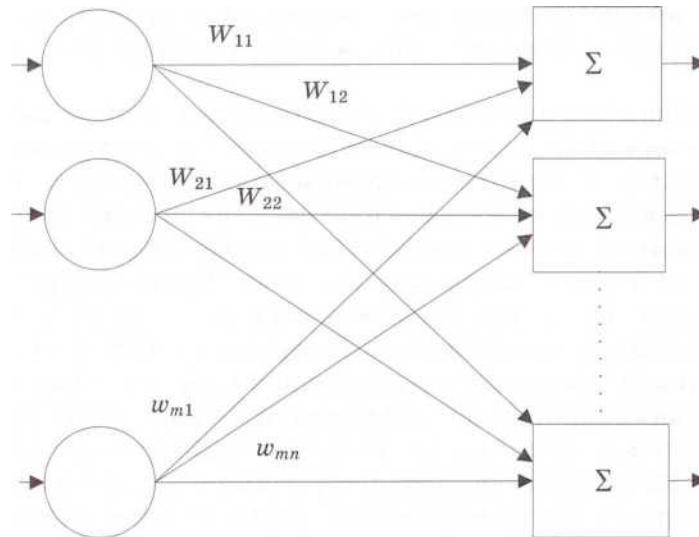


Рисунок 3.8 - Одношарова нейронна мережа

На жаль, немає загальноприйнятого способу підрахунку кількості шарів у мережі. Багатошарова мережа складається, як показано нарис. 3.9, з множини нейронів і вагів, що чергуються.

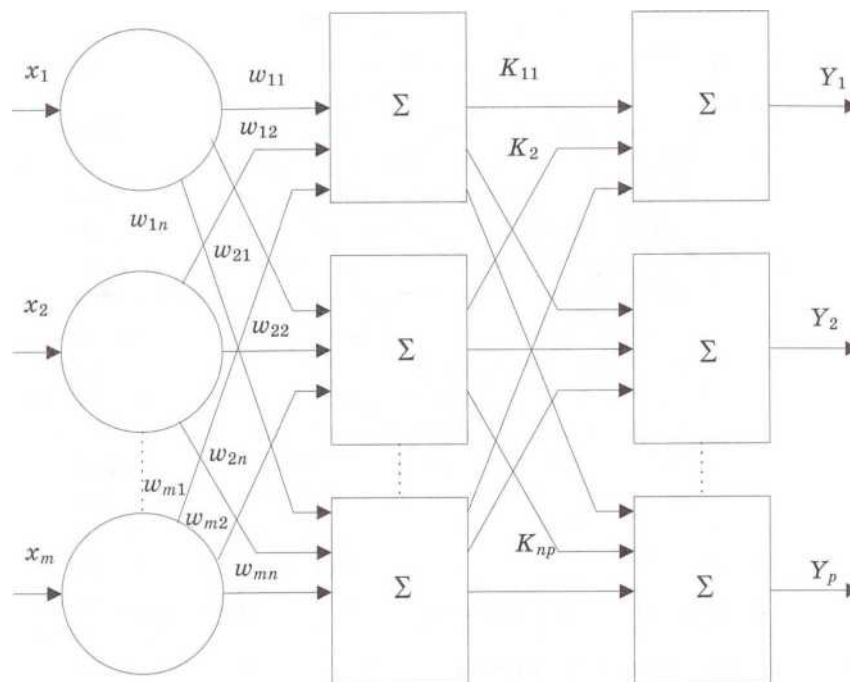


Рисунок 3.9 - Багатошарова нейронна мережа

Перший шар не береться до уваги при підрахунку шарів, і мережа, схожа до зображеної на рис. 3.9, вважається двошаровою, оскільки тільки два шари виконують обчислення. Ваги шару вважаються пов'язаними з наступними за ними нейронами. Отже, шар складається з множини вагів із наступними за ними нейронами, що підсумовують зважені сигнали.

Розглянемо найбільш поширені архітектури штучних нейронних мереж.

Персептрон. Персептрон (*perceptron*) — математична і комп'ютерна модель сприйняття інформації мозком (кібернетична модель мозку), запропонована Френком Розенблаттом і реалізована у вигляді електронної машини “Марк-1” у 1960 р. Персептрон став однією з перших моделей нейромереж, а “Марк-1” — першим у світі нейрокомп'ютером. Незважаючи на простоту, персептрон здатний навчатися і розв'язувати досить складні задачі.

Персептрон складається з трьох типів елементів, а саме: сигнали, що надходять від сенсорів (S-елементи), передаються асоціативним елементам (А-елементи), а потім реагуючим елементам (ІТ-елементи). Кожен сенсор може перебувати в одному з двох станів — *спокою* або *збудження*, і лише в останньому випадку він передає одиничний сигнал до наступного шару, 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 0.

Сигнали від збуджених А-елементів, у свою чергу, передаються в суматор R, причому сигнал від і-го асоціативного елемента передається з коефіцієнтом w_i . Цей коефіцієнт називається *вагою А-R* зв'язку. Так само як А-елементи, ІТ-елемент підраховує суму значень вхідних сигналів, помножених на ваги.

А-елемент, а разом з ним і елементарний персептрон, видає “1”, якщо лінійна форма перевищує поріг i , інакше на виході буде “-1”. Математично функцію, що реалізовується її-елементом, можна записати так:

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

Отже, персептрони дають можливість створити набір “асоціацій” між вхідними стимулами і необхідною реакцією на виході. У біологічному плані це відповідає перетворенню, наприклад зорової інформації на фізіологічну відповідь від рухових нейронів. Згідно з сучасною термінологією персептрони можуть бути класифіковані як штучні нейронні мережі:

- 1) з одним прихованим шаром;
- 2) з пороговою передавальною функцією;

з) з прямим поширенням сигналу.

Навчання елементарного перцептрона полягає у зміні вагових коефіцієнтів ш. зв'язків $A-R$. Ваги зв'язків $S-A$ (які можуть набувати значення $\{-1; 0; +1\}$) і значення порогів A -елементів обираються випадково 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 такому: при пред'явленні об'єкта збуджені A -елементи передають сигнал IT -елементу, що дорівнює сумі відповідних коефіцієнтів w_i . Якщо ця сума додатна, то ухвалюється рішення, що цей об'єкт належить до першого класу, а якщо від'ємна — до другого.

Важливою властивістю будь-якої нейронної мережі є здатність до навчання. Ф. Розенблатт намагався класифікувати різні алгоритми навчання перцептрона, називаючи їх системами підкріплення. При цьому слід розрізняти здатність до представлення і навчання. Здатність до представлення реалізується у можливості перцептрона (або іншої мережі) моделювати певну функцію, тоді як навчання потребує 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, за якого вага зв'язку не змінюється доти, доки поточна реакція перцептрона залишається правильною. За появи неправильної реакції вага змінюється на одиницю, а знак (+/-) визначається протилежним від знака помилки.

Припустимо, ми хочемо навчити перцептрон розділяти два класи об'єктів так, щоб при пред'явленні об'єктів першого класу вихід перцептрона був додатним (+1), а при пред'явленні об'єктів другого класу — від'ємним (-1). Для цього виконаємо такий алгоритм.

1. Випадково обираємо пороги для A -елементів і встановлюємо зв'язки $S-A$ (далі вони змінюватимуться не будуть).

2. Початкові коефіцієнти w_i вважаємо такими, що дорівнюють нулю.

3. Пред'являємо навчальну вибірку із зазначенням класу, до якого вона належить.

• Показуємо перцептроні об'єкт першого класу. При цьому деякі A -елементи збуджуються. Коефіцієнти w_i , що відповідають цим збудженим елементам, збільшуємо на 1.

• Пред'являємо об'єкт другого класу і коефіцієнти тих A -елементів, які збуджуються при цьому показі, зменшуємо на 1.

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

Теорема збіжності персептрона показує, що елементарний персептрон, який навчається за таким алгоритмом, 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 персептрона Ф. Розенблатт також увів поняття навчання без вчителя, запропонувавши спосіб навчання, названий “альфа-система підкріплення”. Це система, за якої ваги всіх активних зв’язків c_{ij} , що ведуть до елемента U_j , змінюються на однакову величину α , а ваги неактивних зв’язків за цей час не змінюються. З часом, після впровадження поняття багат шарового персептрона, альфа-система була модифікована, і її почали 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 більше одного шару.

Метод зворотного поширення помилки. Для навчання багат шарових мереж був запропонований градієнтний алгоритм навчання з учителем, який проводить сигнал помилки, обчислений виходами персептрона, до його входів, шар за шаром. Зараз це найпопулярніший метод навчання багат шарових персептронів. Його перевага в тому, що він може навчити всі шари нейронної мережі, і його легко прорахувати локально. Проте цей метод є 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 методу зворотного поширення помилки потребує виконання таких 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 доти, доки помилка на всій множині не досягне прийнятного рівня.

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

Персептрони демонструють великі можливості для свого

застосування, зокрема в задачах класифікації, кластеризації, прогнозування та апроксимації, стиснення даних та асоціативної пам'яті.

Задачі класифікації. Образами можуть бути різні за природою об'єкти: економічні показники, зображення, зразки звуків та ін. При навчанні мережі пропонуються різні зразки образів із вказівкою того, до якого класу вони належать. Зразок, як правило, подається як вектор значень ознак. При цьому сукупність всіх ознак повинна однозначно визначати клас, до якого належить зразок. У випадку, якщо ознак недостатньо, мережа може співвіднести один і той же зразок із декількома класами, що неправильно. Після закінчення навчання мережі їй можна пред'являти невідомі раніше образи і отримувати відповідь про належність до певного класу.

Топологія такої мережі характеризується тим, що кількість нейронів у вихідному шарі, як правило, дорівнює кількості визначуваних класів. При цьому встановлюється відповідність між виходом нейронної мережі і класом, який він представляє. Коли мережі пропонується якийсь образ, на одному з її виходів повинна з'явитися ознака того, що образ належить цьому класу, водночас на інших виходах повинна бути ознака того, що образ цьому класу не належить. Якщо на двох або більше виходах є ознака належності до класу, вважається що мережа “не впевнена” у своїй відповіді.

Задачі кластеризації. Кластеризація це 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 належить вхідний сигнал. Мережа також може сигналізувати про те, що вхідний сигнал не належить до жодного з виділених класів — це є ознакою нових, відсутніх у навчальній вибірці, даних. Отже, така мережа може виявляти нові, не відомі раніше класи сигналів. Відповідність між класами, виділеними мережею, і класами, що існують у праметній галузі, встановлюється людиною.

Прогнозування і апроксимація. Здатності нейронної мережі до прогнозування безпосередньо виходять з її спроможності 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 наявних чинників. Слід зазначити, що прогнозування можливе тільки тоді, коли попередні зміни справді 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 жодних 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 взаємопов'язані один з одним. Зворотний процес — відновлення початкового набору даних з частини інформації — називається асоціативною пам'яттю. Асоціативна пам'ять дає змогу також відновлювати початковий образ із зашумлених вхідних даних. Рішення задачі гетероасоціативної пам'яті допомагає реалізувати пам'ять, що адресується за змістом.

Багатошаровий перцептрон. Багатошаровий перцептрон Розенблатта — перцептрон з додатковими шарами А-елементів, розташованими між Si Д-елементами. Оскільки елементарний перцептрон вже має два шари зв'язків - три шари елементів (нейронів), то такий перцептрон не вважається багатошаровим, і багатошаровість зумовлювалась тільки за наявності мінімум чотирьох шарів елементів. Інша важлива відмінність полягає в тому, що не обов'язково всі зв'язки мають бути навчальні, частина з них може бути випадково вибрана . фіксована.

Нейронна мережа Ворда. Топологія штучної нейронної мережі характеризується тим, що внутрішні (приховані) шари нейронів розбиті на блоки (рис. 3.11): 1) нейрони вхідного шару, 2) нейрони блоку прихованого шару та 3) нейрони вихідного шару.

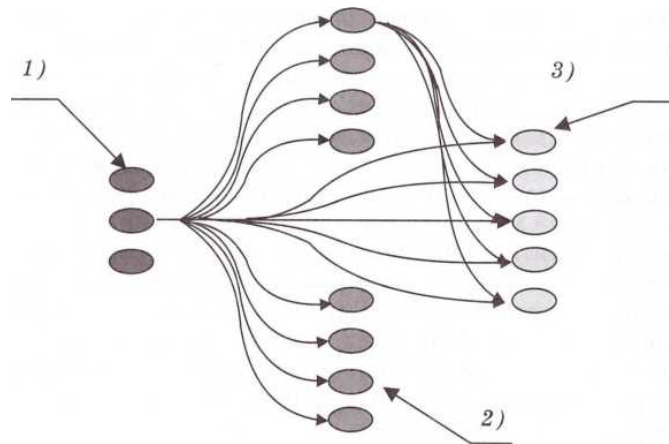


Рисунок 3.11 - Нейронна мережа Ворда з двома прихованими шарами: 1) нейрони вхідного шару; 2) нейрони блоку прихованого шару; 3) нейрони вихідного шару

Нейронні мережі Ворда розрізняються за кількістю блоків прихованого шару і наявністю або відсутністю обхідних з'єднань. Розбиття прихованих шарів на блоки дає змогу використовувати різні передавальні функції для різних полтоків прихованого шару. Отже, одні й ті самі 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 й того самого набору даних дає змогу стверджувати, що нейронна мережа аналізує дані з різних аспектів. На практиці мережа показує дуже гарні результати при розв'язанні задач прогнозування і розпізнавання образів. Для нейронів вхідного шару, як правило, істановлюється лінійна функція активації. Функція активації для нейронів з токів прихованого і вихідного шару визначається експериментально.

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

Нейронна мережа Джордана. Мережа Джордана — це вид мереж, який виходить з багатошарового персептрона, якщо на його вхід подати крім вхідного вектора вихідний із затримкою на один або декілька тактів. У перших рекурентних мережах головною ідеєю було дати мережі бачити свій вихідний образ на попередньому кроці. У такій мережі тільки частина сенсорів приймає сигнали з навколишнього світу, на інші сенсори надходить вихідний образ з попереднього моменту часу. Розглянемо проходження послідовності сигналів через мережу. Сигнал надходить на групу сенсорів, сполучених із зовнішнім світом (INPUT) і проходить у прихований шар (HIDDEN). Перетворений прихованим шаром сигнал піде на вихідний шар (OUTPUT) і вийде з мережі, а його копія потрапить на затримку. Потім у мережу, на сенсори, що сприймають зовнішні сигнали, надходить інший образ, а на контекстну групу сенсорів (CONTEXT) — вихідний образ з попереднього кроку з затримки. Далі з всіх сенсорів сигнал йде до прихованого шару, потім — до вихідного.

Нейронна мережа Елмана. Нейронна мережа Елмана — один із видів рекурентної мережі, яка, так само як і мережа Джордана, виходить з багатошарового персептрона введенням зворотних зв'язків, тільки зв'язки йдуть не від виходу мережі, а від виходів внутрішніх нейронів. Це дає можливість враховувати передісторію спостережуваних процесів і накопичувати 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 багатошарову 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 галузях візуальної області, тоді як реакція інших шарів є складнішою, абстрактнішою і незалежною від позиції образу.

Когнітрон конструюється у вигляді шарів нейронів, сполучених синапсами. Передсинаптичний нейрон в одному шарі пов'язаний з постсинаптичним нейроном у наступному. Є два типи нейронів:

збуджувальні вузли, які прагнуть викликати збудження постсинаптичного вузла, і гальмівні вузли, які гальмують це збудження. Збудження нейрона визначається зваженою сумою його збуджувальних і гальмівних входів, проте насправді механізм є складнішим, ніж просте підсумовування. Ця 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, дає змогу 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 неокогнітроні немає нічого, що обмежує його використання тільки для обробки візуальних даних, він достатньо універсальний і може отримати широкое застосування як узагальнена система розпізнавання образів.

Хаотичні нейронні мережі. За структурою такі мережі є одношаровими рекурентними мережами, елементи яких пов'язані "кожен із кожним", без утворення зв'язку "сам на себе". Структуру мережі утворюють N-нейрони, 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 виокремленні 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 поле, що діє на них.

У хаотичній нейромережі на вхід у 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. Подання на вхід мережі набору об'єктів може бути розглянуте як вплив навколишнього середовища на спочатку неврегульовану поведінку нейронів і формування під її дією нейронних ансамблів. При цьому сумісне функціонування нейронів є виявом інтелекту на рівні малого колективу, вирішального завдання кластеризації.

Розглядаючи сумісну динаміку поведінки нейронів після проходження перехідного процесу, виокремлюють пари нейронів, які змінюють 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 належать до одного кластера. Первинний хаос перехідного процесу з часом зменшується, оскільки за рахунок виявлення взаємозв'язку значення 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 становлять науковий інтерес, оскільки істотну роль у розумових процесах людини відіграють коливання.

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 ця бажана властивість досягається не безкоштовно, мережі без зворотних зв'язків володіють більш обмеженими можливостями порівняно з мережами із зворотними зв'язками.

Оскільки мережі зі 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 мережею зворотного зв'язку, модифікує вхід. Потім вихід повторно обчислюється, і процес повторюється знову і знову. Для стійкої мережі послідовні ітерації приводять до все менших змін на виході, 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 міг передбачити, які з мереж будуть стійкими, а які перебуватимуть у постійній зміні. Більше того, проблема уявлялася настільки важкою, що багато дослідників були налаштовані песимістично щодо можливості застосування нейронних мереж. На щастя, була отримана теорема, що описала підмножину мереж зі 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 внесок як у теорію, так і у використання систем зі зворотними зв'язками. Тому деякі з конфігурацій відомі як мережі Хопфілда.

Розглянемо мережу зі зворотними зв'язками, що складається з двох шарів (рис. 3.12). 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 перетворюється на сигнал OUT . Ці операції схожі з нейронами інших мереж.

У 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.

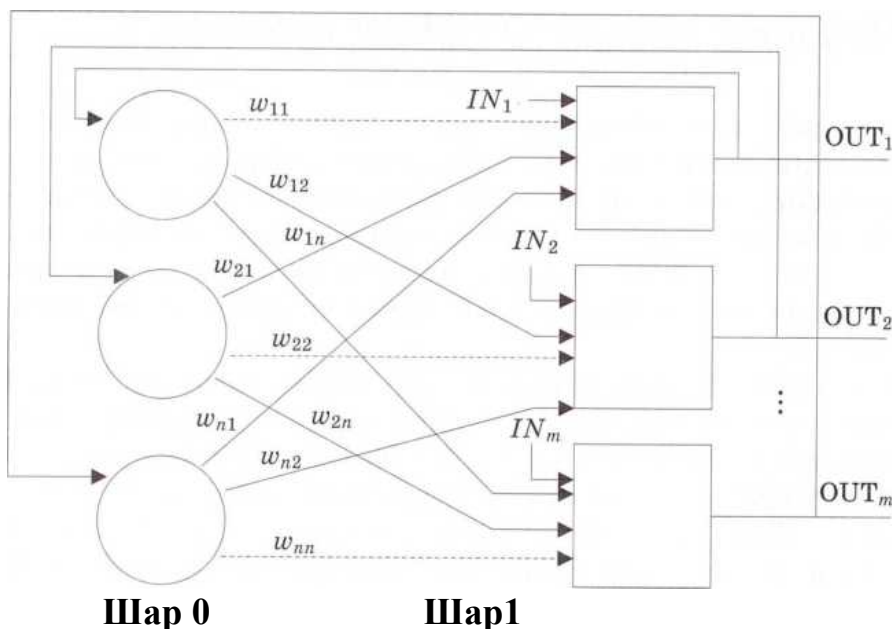


Рисунок - 3.12 Одношарова мережа зі зворотними зв'язками

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

нейрона.

Функціонування мережі легко візуалізується геометрично. На рис. 3.13 показана тринейронна система, представлена кубом, що має вісім вершин, кожна з яких помічена трибітовим бінарним числом. У загальному випадку система з n нейронами має 2^n різних станів і представляється n -мірним гіперкубом.

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

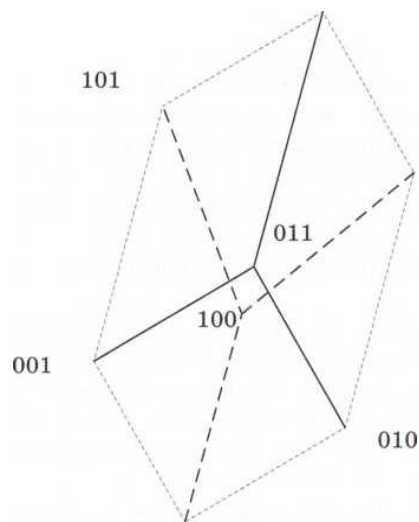


Рисунок 3.13 Три нейрони породжують систему з вісьмома станами

Людська пам'ять асоціативна, тобто певний 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 адресованою, подається адреса і витягується інформація за цією адресою. Мережа з зворотним зв'язком формує асоціативну пам'ять. Подібно до людської пам'яті за заданою частиною потрібної інформації вся інформація витягується з "пам'яті". Аби організувати асоціативну пам'ять за допомогою мережі зі зворотними зв'язками, ваги повинні збиратися так, щоб утворювати енергетичні мінімуми у потрібних вершинах одиничного гіперкуба. Дж. Хопфілд розробив асоціативну пам'ять з безперервними виходами, що змінюються в межах від +1 до -1, відповідних двійковим значенням 0 і 1. Інформація, що запам'ятовується, кодується двійковими векторами і зберігається у вагах за такою формулою.

Цей вираз може стати яснішим, якщо зазначити, що ваговий масив W може бути знайдений обчисленням зовнішнього множення кожного

вектора, що запам'ятовується, з самим собою і підсумовуванням матриць, отриманих таким чином.

Як тільки ваги задані, мережа може бути використана для одержання вихідного вектора, що запам'ятовується, за даним вхідним вектором, 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 у найближчий глибокий мінімум. Мережа йде за локальним нахилом функції енергії і може бути захоплена локальним мінімумом, не досягнувши найкращого в глобальному розумінні рішення.

Недоліком мереж Хопфілда є їх тенденція стабілізуватися в локальному, а не глобальному мінімумі функції енергії. Ця проблема долається в основному за допомогою класу мереж, відомих під назвою машин Больцмана, в яких зміни станів нейронів обумовлені статистичними, а не детермінованими закономірностями. Є 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, що часто виникає на практиці. Вона може бути сформульована так: для деякої групи міст з заданими відстанями між ними потрібно знайти найкоротший маршрут із відвіданням 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 немає кращого методу розв'язання, ніж повний спектр усіх можливих варіантів. Оскільки такий повний пошук практично нездійснений для великої кількості міст, то евристичні методи використовуються для пошуку прийнятних, хоча і неоптимальних рішень.

Запропоноване рішення, основане на мережах з зворотними зв'язками, є типовим у цьому сенсі. Відповідь знаходиться так швидко, що в певних випадках метод може виявитися корисним. Припустимо, що міста, які необхідно відвідати, позначені буквами *A*, *B*, *C* і *D*, а відстані між парами міст це d_{ab} , d_{bc} та ін. 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, у якому це місто відвідується при обході. На рис. 3.14 показаний випадок, коли місто *C* відвідується першим, місто *A* — другим, місто *D* — третім і місто *B* — четвертим.

Для такого варіанта потрібно n^2 нейронів — число, яке швидко зростає зі збільшенням кількості міст. Довжина такого маршруту дорівнювала б $d_{ca} + d_{ad} + d_{ab} + d_{bc}$. Оскільки кожне місто відвідується лише

один раз і у кожен момент відвідується лише одне місто, то в кожному стовпці є по одній одиниці.

Місто	Порядок дотримування			
	1	2	3	4
<i>A</i>	0	1	0	0
<i>B</i>	0	0	0	1
<i>C</i>	1	0	0	0
<i>D</i>	0	0	1	0

Рисунок 3.14 - Варіант коли місто відвідується при обході

Для задачі з n містами всього є ni різних маршрутів обходу. Якщо $n = 60$, то існує $6\,934\,155 \cdot 10^{78}$ можливих маршрутів. Якщо взяти 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 у світі комп'ютері займе час, порівнянний зі 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 у маршруті. Наприклад, $OUT_x = 1$ показує, що місто x було j -м за порядком містом маршруту.

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

1. Перша потрібна сума дорівнює нулю в тому (і лише в тому випадку), якщо кожен рядок (місто) має не більше, ніж одну одиницю.

Друга потрібна сума дорівнює нулю в тому (і лише в тому випадку), якщо кожен стовпець (порядковий номер відвідання) містить не більше, ніж одну одиницю

3. Третя сума дорівнює нулю в тому (і лише в тому випадку), якщо матриця містить рівно n одиниць.

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

Зазначимо, що цей член є довжиною будь-якого допустимого маршруту. Для зручності індекси визначаються за модулем n , тобто $OUT_{n+j} = OUT_j$, aD — деяка константа. За достатньо великих значень A , B і C низькоенергетичні стани стануть допустимі маршрути, а великі значення D гарантують, що буде знайдено короткий маршрут.

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

$W = \sum_{xy} A_{xy}^8 (X - 5y)$ (не допускає більше, ніж одну одиницю в рядку),
 $-B_{8y} (1 - 5)$ (не допускає більше, ніж одну одиницю в стовпці),
 $-C$ (глобальне обмеження),
 $-D d_{xy} (5_{j+i} + 5_{-j})$ (член, що відповідає за довжину циклу),
де $5_{i-} = 1$, якщо $i = j$, інакше $5_{i-} = 0$.

Результати експерименту, в якому задача комівояжера була розв'язана для 10 міст, продемонстрували, що 16 з 20 прогонів зійшлися до допустимого маршруту і близько 50 % розв'язків виявилися найкоротшими маршрутами, як це було встановлено за допомогою повного перебору. Цей результат стане більш вражаючим, якщо усвідомити, що є 181 440 допустимих маршрутів.

Розглянемо практичне застосування нейромереж на прикладі розв'язання задачі “Видача кредиту клієнтові” в аналітичному пакеті Deductor. Навчальним набором даних є 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, що хоче отримати кредит, до групи ризику неповернення кредиту, тобто користувач повинен отримати відповідь на питання чи “Видавати кредит?”. Завдання належить до групи завдань класифікації, тобто навчання з учителем.

Дані для аналізу перебувають у заздалегідь підготовленому файлі. Імпортуємо дані з файла за допомогою майстра імпорту. Запускаємо майстер обробки і обираємо метод обробки даних — нейронна мережа. Задаємо призначення початкових стовпців даних. Вихідний стовпець у нашому завданні — “Давати кредит”, всі інші — вхідні (рис. 3.20).

На наступному кроці розбиваємо початкову множину даних на навчальну і тестову та визначаємо структуру нейронної мережі, тобто зазначаємо кількість нейронів у вхідному шарі — 33 (кількість вхідних змінних), у прихованому — 1, у вихідному шарі — 1 (кількість вихідних змінних). Активаційна функція — Сигмоїда, і її крутизна дорівнює.

Далі обираємо алгоритм і параметри навчання нейронної мережі і налаштовуємо 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 у навчальній і тестовій множинах. У нашому випадку бачимо, що на епісі № 4536 у навчальній множині розпізнано 83,10 % прикладів, а на тестовому — 85,71 % прикладів.

Після закінчення процесу навчання для інтерпретації отриманих результатів маємо можливість вибрати візуалізатори зі списку запропонованих, зокрема таблицю зв'язаності, граф нейромережі, аналіз

“що, якщо”, і за допомогою їх проаналізувати отримані дані.

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 % .

Нейронна мережа підготовлена для практичного використання.

Важливим напрямом у розвитку нейронних мереж є самоорганізуючі карти (Self-Organizing Maps, SOM) або мережі (карти) Кохонена.

Мережі, названі картами Кохонена, — це один з різновидів нейронних мереж, проте вони принципово відрізняються від розглянутих вище, оскільки використовують неконтрольоване навчання. Нагадаємо, що за такого навчання навчальна множина складається лише зі значень вхідних змінних, у 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 р.). Основний принцип роботи мереж — введення в правило навчання нейрона інформації про його розташування. В основі ідеї мережі Кохонена — 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 відбивається на цій двовимірній поверхні).

Карти, що самоорганізуються, можуть використовуватися для вирішення таких завдань, як моделювання, прогнозування, пошук закономірностей у великих масивах даних, виявлення наборів незалежних ознак і стиснення інформації. Найбільш поширене застосування мереж Кохонена — розв’язання задачі класифікації без вчителя, тобто кластеризації. Нагадаємо, що за такої постановки задачі є набір об’єктів, кожному з яких відповідає рядок таблиці (вектор значень ознак). Потрібно розбити початкову множину на класи, тобто для кожного об’єкта знайти клас, до якого він належить. У результаті одержання нової інформації про класи можлива корекція наявних правил класифікації об’єктів.

Найбільше застосовують карти 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 не-йромережеву 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 у навчальних даних і зараховує всі дані до тих або інших кластерів. Якщо після цього мережа зустрінеться з набором даних, не схожим ні на один з відомих зразків, то вона не зможе класифікувати такий набір і тим самим виявить його новизну.

Мережа Кохонена, на відміну від багатошарової нейронної мережі, дуже проста, у ній є два шари: вхідний і вихідний. Її також називають картою, що самоорганізується. Елементи карти розташовуються в деякому просторі, як правило, двовимірному.

Мережа Кохонена (3.15) навчається методом послідовних наближень. У процесі навчання таких мереж на входи подаються дані, але мережа при цьому підлаштовується не під еталонне значення виходу, а під закономірності у вхідних даних. Починається навчання з обраного випадково вихідного розташування центрів. У процесі послідовного подання на вхід мережі навчальних прикладів визначається найбільш схожий нейрон (той, у якого скалярне множення вагів і поданого на вхід вектора мінімальне). Цей нейрон оголошується переможцем і є центром при підлаштуванні 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 підлаштуванні вагів (внутрішніх параметрів нейронної мережі) для найбільшого збігу з вхідними даними.

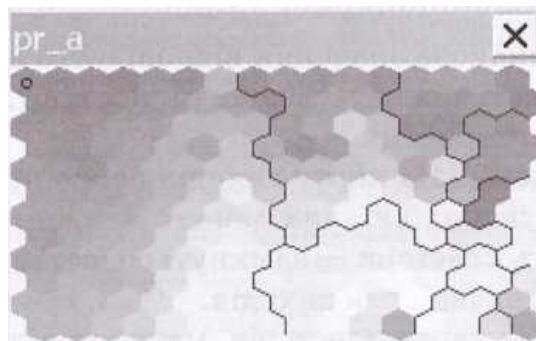


Рисунок 3.15 - Приклад карти Кохонена

Основний ітераційний алгоритм Кохонена послідовно проходить низку епох, на кожній з яких обробляється один приклад із навчальної вибірки. Вхідні сигнали послідовно пред'являються мережі, при цьому бажані вихідні сигнали не визначаються. Після пред'явлення достатньої кількості вхідних векторів синаптичні ваги мережі можуть визначити кластери. Ваги організуються так, що топологічно близькі вузли чутливі

до схожих вхідних сигналів. У результаті роботи алгоритму центр кластера встановлюється в певній позиції, що задовільно кластеризує приклади, для яких цей нейрон є “переможцем”. У результаті навчання мережі необхідно визначити міру сусідства нейронів, тобто окіл нейрона-переможця. Околом є декілька нейронів, що оточують нейрон-переможця. Спочатку до околу належить багато нейронів, потім його розмір поступово зменшується. Мережа формує топологічну структуру, в якій схожі приклади утворюють групи прикладів, що близько перебувають на топологічній карті. Отриману карту можна використовувати як засіб візуалізації при аналізі даних. У 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 помірною простору на двомірний. Застосування двовимірних сіток пов’язане з тим, що існує проблема відображення просторових структур більшої розмірності. Маючи таке представлення даних, можна візуально визначити наявність або відсутності взаємозв’язку у вхідних даних. Нейрони карти Кохонена розташовують у вигляді двомірної матриці, розфарбовують цю матрицю залежно від аналізованих параметрів нейронів (рис. 3.15).

Що ж означає її забарвлення? На рис. 3.16 подано зразок забарвлення карти, тобто її деякої ознаки у тривимірному представленні: темносині ділянки на карті відповідають найменшим значенням показника, червоні — найвищим.

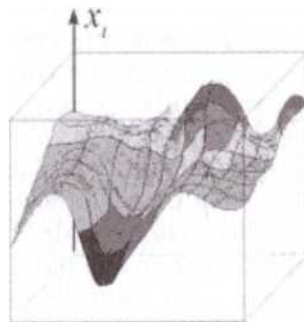


Рисунок 3.16 - Розфарбування ознаки у тривимірному просторі

Тепер можемо сказати, які об’єкти мають найбільші значення цього показника (група об’єктів, позначена червоним кольором), а які — найменші значення (група об’єктів, позначена синім кольором). Отже, карти Кохонена (як і географічні карти) можна відображати:

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

У 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 у матриці нейронів. Розглянемо карти детальніше.

Карта входів нейронів. Ваги нейронів підлаштовуються під значення вхідних змінних і відображають їх внутрішню структуру. Для кожного входу малюється своя карта, забарвлена 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 із 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 й інші карти, які характеризують кластери, отримані в результаті навчання мережі Кохонена.

Важливо розуміти, що між усіма розглянутими картами є взаємозв'язок

всі вони є різними забарвленнями одних і тих самих нейронів. Кожен приклад із навчальної вибірки має одне і те саме розташування на всіх картах.

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

До інструментарію, що включає реалізацію методу карт Кохонена, належать SoMine, Statistica, NeuroShell, NeuroScalp, Deductori багато інших. Для розв'язання задач використовуватимемо аналітичний пакет Deductor. Нехай є база даних комерційних банків із показниками діяльності за поточний період. Необхідно провести їх кластеризацію, тобто виокремити однорідні групи банків на основі показників із бази даних, всього показників — 21. Початкова таблиця міститься у файлі “banks.xls”. Вона має показники діяльності комерційних банків за звітний період. Спочатку імпортуємо дані з xls-файла до аналітичного пакета.

На першому кроці запускаємо майстер обробки і вибираємо зі списку метод обробки “Карта Кохонена”. Далі визначаємо призначення стовпців, тобто для кожного стовпця обираємо одне з призначень: вхідне, вихідне, не використовується й 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 навчальну — 95 % і тестову — 5 % . Ці кроки аналогічні до кроків у майстрові обробки для нейронних мереж. Далі задаємо параметри карти: кількість комірок по X і по Y, їх форму (шестикутну або чотирикутну), встановлюємо параметри зупинки навчання і встановлюємо епоху, за досягнення якої навчання буде припинено.

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

На восьмому кроці запускаємо процес навчання мережі. Під час навчання можемо спостерігати зміну кількості розпізнаних прикладів і поточні значення помилок. Цей процес аналогічний до того, який 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.

На одній з карт виокремлюємо область з найбільшими значеннями показника. Шукаємо ці нейрони і на інших картах. На першій карті найбільші значення мають об’єкти, розташовані в правому нижньому куті. Розглядаючи одночасно три карти, можна сказати, що ці об’єкти мають найбільші значення показника, зображеного на третій карті. Із забарвлення першої і третьої карти можна зробити висновок, що існує взаємозв’язок між цими показниками. Також можемо визначити, наприклад таку характеристику: кластер, розташований у правому верхньому куті, характеризується низькими значеннями показників “депозити юридичних осіб” і “активи банку” та високими значеннями показника “прибутковість активів”. Ця інформація дає змогу так охарактеризувати кластер, що знаходиться в правому верхньому куті: 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 рис. 3.17 подано ілюстрацію карт входів і виходів, остання — це карта кластерів. Ми бачимо декілька карт входів (показників діяльності банків) : сформовані кластери, кожен з яких виділений окремим кольором.

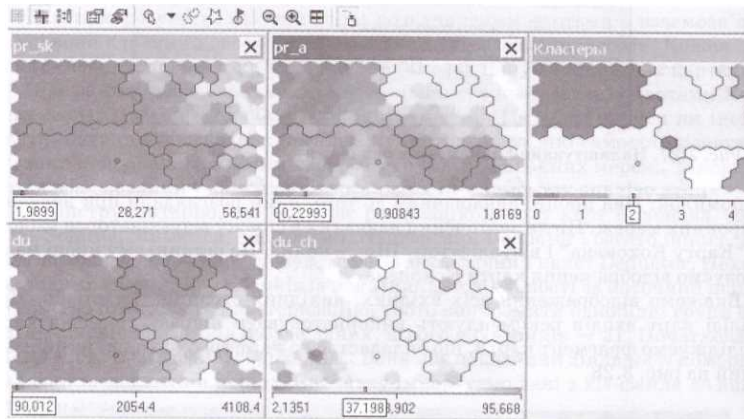


Рисунок 3.17 - Карти входів та виходів

Для пошуку конкретного об'єкта на карті необхідно клацнути правою кнопкою миші на досліджуваному об'єкті і обирати пункт “Знайти ячейку на карті”. У результаті можемо бачити як сам об'єкт, так і значення того виміру, який переглядаємо. Отже, можна оцінити положення аналізованого об'єкта, а також порівняти його з іншими об'єктами.

У результаті застосування карт, що самоорганізуються, багатовимірний простір вхідних чинників був представлений у двовимірному вигляді, в якому його досить зручно аналізувати. Зокрема, банки були класифіковані на 7 груп, для кожної з яких можливе визначення конкретних характеристик, виходячи із забарвлення відповідних показників.

Отже, мережі Кохонена дають змогу спростити багатовимірну структуру, їх можна розглядати як один із методів проектування багатовимірного простору у простір із нижчою розмірністю. Інтенсивність кольору в певній точці карти визначається даними, які туди потрапили: комірка з мінімальними значеннями зображається темно-синім кольором, комірка з максимальними значеннями — червоним.

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

3.3 Програмні засоби реалізації нейромережових технологій

Нейронні мережі активно застосовуються в задачах аналізу вже більше 10 років. На фоні постійно змінюваної динаміки ринку інтерес до нейромережових технологій не лише не слабшає, а зростає з кожним роком. Накопичений багатий досвід успішного використання нейронних мереж у практичних застосуваннях. За кількістю реальних впроваджень лідирують системи інтелектуального аналізу даних у бізнесі і в управлінні процесами.

Більшість програмних комплексів, оснований на нейромережових технологіях, включають таку 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 такі: простота вживання, наочність інформації, що представляється, можливість використовувати різні структури, швидкість роботи, наявність документації. Вибір визначається кваліфікацією і вимогами користувача.

Для застосування методів нейронних мереж у процесі інтелектуального аналізу даних у бізнес-додатках розроблено низку інструментальних засобів високого рівня. До них належать насамперед системи MathLab, STATISTICA Neural Networks, NeuroSolutions, BrainMakerPro, NeuroShell 2,4Thought, SENN Sales та ін.

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

Пакет MathLab. Пакет MathLab надає користувачам можливість роботи з нейронними мережами. Стандартне постачання MathLab “NeuralNetworkToolbox” надає широкі можливості для роботи з нейронними мережами всіх типів. Перевага цього пакета полягає в тому, що при його використанні користувач не обмежений моделями нейронних мереж і їх параметрами, жорстко закладеними в нейросимуляторі, а має можливість самостійно конструювати ту мережу, яку вважає оптимальною для вирішення поставленого завдання. Пакет містить функції командного рядка і графічний інтерактивний майстер для швидкого покрокового створення нейромереж. Окрім цього NeuralNetworkToolbox забезпечує підтримку Simulink, що дозволяє моделювати нейро- мережі і створювати блоки на основі розроблених нейромережових структур. Ключовими можливостями Neural Network Toolbox є:

- графічний інтерфейс користувача і майстер покрокового створення нейронних мереж;
- підтримка найбільш поширених мережових парадигм;

- повний набір засобів для тренування нейромереж з учителем і без;
- динамічно навчальні 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 дає змогу створювати необмежену кількість шарів і міжмережевих зв'язків;
- візуалізація топології нейронної мережі.

Розглянемо приклад конструювання нейронної мережі в пакеті Matlab. Нехай маємо 15 незалежних змінних — показників діяльності фірми і одна залежна змінна — обсяг продажу, використаємо базу даних за минулий рік. Необхідно побудувати потижневий прогноз обсягів продажу на місяць. Для вирішення завдання пропонується використовувати тришарову мережу зворотного поширення.

Сформуємо таку мережу, яка включає 15 нейронів у вхідному шарі (за кількістю вхідних змінних), 8 нейронів у другому шарі і 1 нейрон у вихідному шарі (за кількістю вихідних змінних). Для кожного шару оберемо передавальну функцію: перший шар — *logsig*, другий — *logsig*, третій — *purelin*. У середовищі MatLab синтаксис такої нейронної мережі виглядає так:

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

де PR — масив мінімальних і максимальних значень для D-векторів входу;

Si — кількість нейронів у i-му шарі;

TFi — функція активації шару *i*;

btf — навчальна функція, що реалізовує метод зворотного поширення;

blf — функція настройки, що реалізовує метод зворотного поширення;

pf — критерій якості навчання.

Активаційною функцією може бути будь-яка функція, що диференціюється, наприклад *tansig*, *logsig*, *purelin*.

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

де P — множина вхідних векторів;

n — кількість входів нейромережі;

m — кількість нейронів у прихованому шарі;

I — кількість виходів нейромережі.

Необхідно також встановити метод розрахунку значення помилки. Наприклад, якщо обраний метод найменших квадратів, то ця функція виглядатиме так: `Net.performFcn='SSE'`. Для встановлення максимальної кількості епох, що дорівнює 10 000, скористаємося функцією: `net.trainParam.epochs = 10000`. Запустити процес навчання можна так:

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

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

Отже, у 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 `NeuralNetworkToolbox`.

Розглянемо інший, більш складний, приклад застосування нейронних мереж у пакеті `MathLab`.

Прогноз фінансових результатів — необхідний елемент будь-якої інвестиційної діяльності. Сама ідея інвестицій — вкладання грошей зараз із 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. В останнє десятиліття спостерігалось стійке зростання популярності технічного аналізу — набору емпіричних правил, заснованих на різного роду індикаторах поведінки ринку. Технічний аналіз зосереджується на індивідуальній поведінці певного фінансового інструменту, поза його зв'язком із рештою цінних паперів. Але технічний аналіз дуже суб'єктивний і погано працює на правому краю графіка — саме там, де потрібно прогнозувати напрям ціни. Тому дедалі більшої популярності набуває нейромережевий аналіз, оскільки на відміну від технічного, не передбачає жодних обмежень на характер вхідної інформації. Це можуть бути як індикатори певного тимчасового ряду, так і відомості про поведінку інших ринкових інструментів. Недарма нейромережі активно використовують саме інституційні інвестори (наприклад, потужні пенсійні фонди), що працюють з великими портфелями, для яких особливо важливі кореляції між різними ринками. Як відомо, для якісного прогнозу потрібно користуватися, по-перше, професійно підготовленими даними, а по-друге, нейропакетами з великою функціональністю. Дані можна отримати стандартними засобами `MetaTrader` (рис. 3.18). Для одержання прогнозу на основі нейронної мережі застосуємо пакет `MathLab`. З командного рядка за командою `anfised.it` запускаємо пакет `ANFIS`. Редактор складається з чотирьох

панелей — для даних (Loaddata), генерації мережі (GenerateFIS), тренування (TrainFIS) і її тестування (TestFIS). Верхня панель призначена для переглядання структури отриманої нейромережі (ANFISInfo).

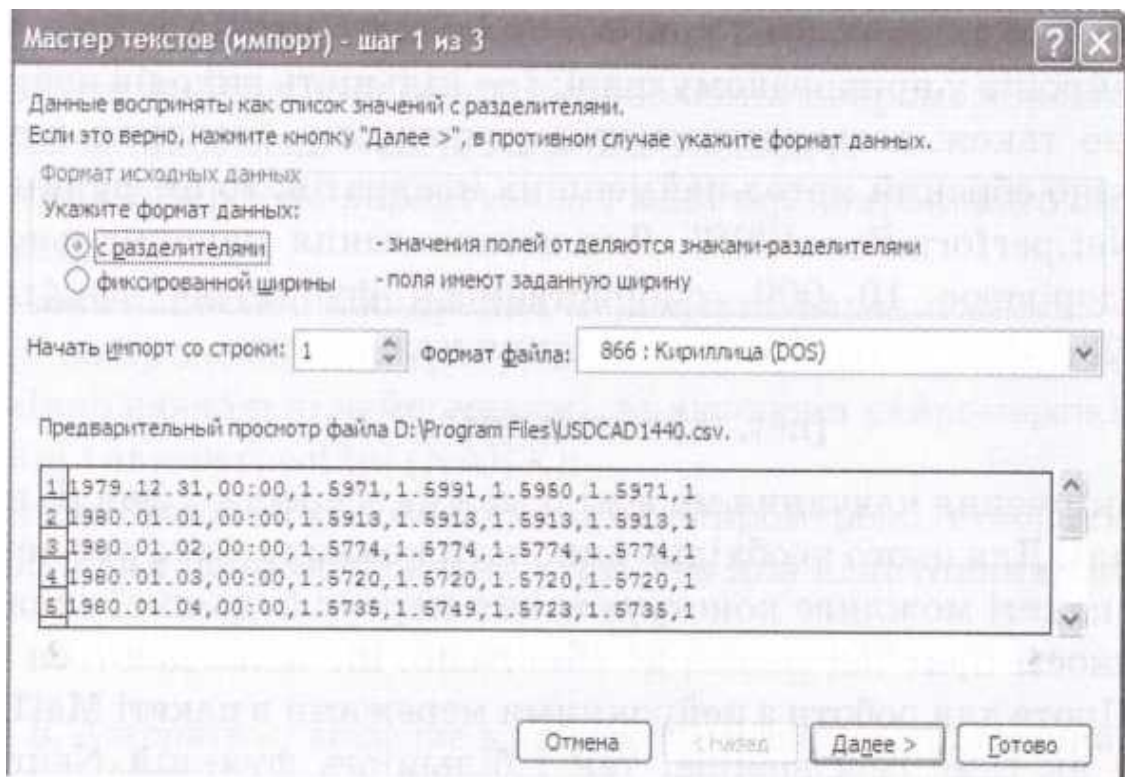


Рисунок 3.18 - Підготовка даних засобами пакета MetaTrader

Для початку роботи завантажуюмо дані, підготовлені на попередніх етапах. Для цього натискаємо кнопку LoadData і вказуємо файл із даними вибірки. Після цього створюємо нейромережу натисненням кнопки GenerateEIS.

Для кожної з вхідних змінних задамо по 3 лінгвістичних змінних із трикутною функцією належності. Як функцію належності вихідної функції задамо лінійну функцію.

Для навчання нейромереж у пакеті передбачено 2 алгоритми навчання — зворотного поширення і гібридний. За гібридного способу навчання мережа навчається буквально за кілька проходів. На тренувальній вибірці (60 значень) після навчання прогноз мережі відрізняється від реального на декілька пунктів.

Результатом роботи є багатошарова гібридна нейронна мережа, яка здатна прогнозувати абсолютні значення цін на невелике майбутнє. Отриману нейромережу можна побачити, натиснувши кнопку Structure (рис. 3.19).

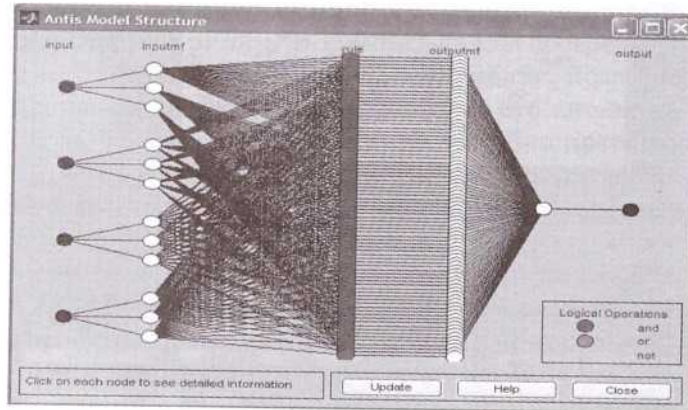


Рисунок 3.19 - Архітектура нейронної мережі

Пакет STATISTICANeuralNetworks. Ця система є потужним і надзвичайно швидким середовищем аналізу нейромережових моделей, що надає такі можливості:

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

- потужні методи розвідувальних і аналітичних технологій, зокрема Аналіз *головних компонент* і *пониження розмірності* для вибору потрібних вхідних змінних у розвідувальному (нейромережевому) аналізі даних;

- найсучасніші, оптимізовані і потужні 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 у *Наборах мереж NetworkSets*, вибіркоче навчання нейромережових 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 потужних графічних і аналітичних інструментів STATISTICA (рис. 3.20).

У пакеті STATISTICANeuralNetworks реалізовані обернений і прямий алгоритми покрокового вибору. Крім того, нейронгенетичний алгоритм відбору вхідних даних поєднує 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 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. 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 “одинична сума”, “переможець отримує все” і “вектор одиничної довжини”. Є засоби підготовки й 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 зі спеціальною реалізованою в STATISTICA Neural Networks функцією активації.

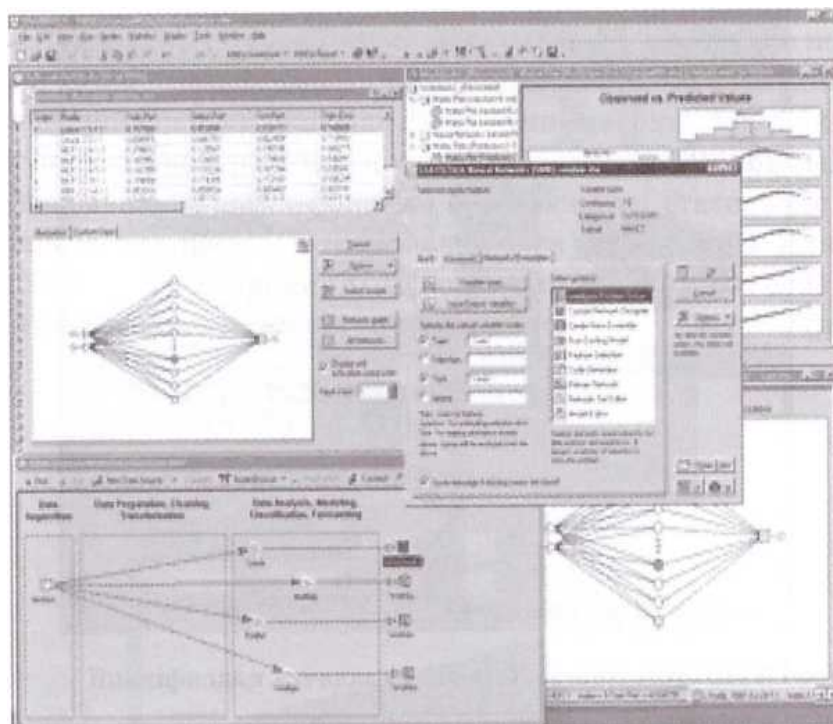


Рисунок 3.20 - Вид меню пакета STATISTICA Neural Networks

Софтмакс і кросс-ентропійними функціями помилок це дає принциповий теоретико-імо-зірнісний підхід до задач класифікації (рис. 3.21).

Різноманіття моделей нейронних мереж і безліч параметрів, які необхідно встановити (розміри мережі, параметри алгоритму навчання та ін.), може побавити іншого користувача у безвихідне становище. 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 гри розв’язанні практичних задач, зокрема

багатошарові перцептрони, мережі на радіальних базисних функціях, карти Кохонена, імовірнісні нейронні мережі; мережі для кластеризації, лінійні мережі та ін. (рис. 3.22).

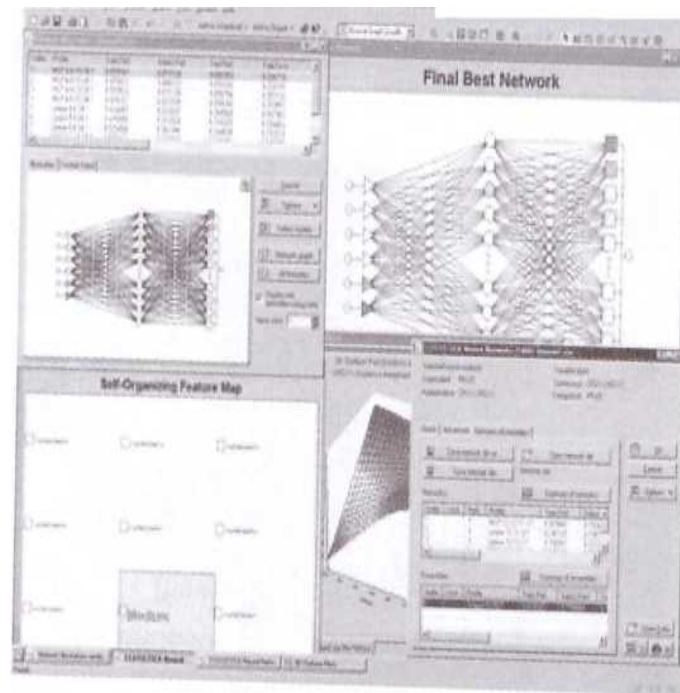


Рисунок 3.21 - STATISTICA Neural Networks у задачах класифікації

У системі реалізовані мережеві ансамблі, що формуються з випадкових та значущих) комбінацій зазначених вище мереж. Є ще один зручний засіб який можна узгодити мережі, щоб вони запускалися послідовно. Це корисно при ігнорірованні для пошуку рішень із мінімальною вартістю (рис. 3.22).

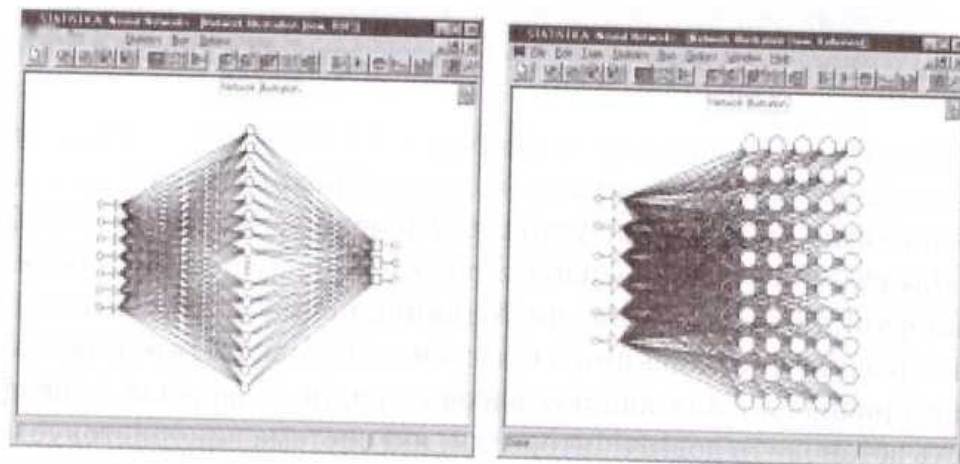


Рисунок 3.22 - Типи нейромереж у пакеті STATISTICA Neural Networks

У пакеті STATISTICANeuralNetworkse численні засоби, що полегшують користувачеві вибір відповідної архітектури мережі. Статистичний і графічний інструментарій системи включає гістограми, матриці й графіки помилок для всієї сукупності і за окремими спостереженнями, підсумкові дані про правильну або неправильну класифікацію, а всі важливі статистики — наприклад, пояснена частка дисперсії — обчислюються автоматично. Для візуалізації даних у пакеті реалізовані діаграми розсіяння і тривимірні поверхні відгуку, що допомагають користувачеві зрозуміти “поведінку” мережі (рис. 3.23).

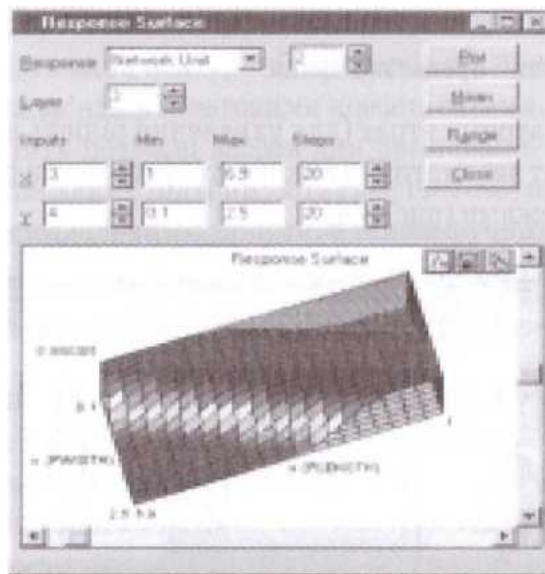


Рисунок 3.23 - Візуалізація даних у пакеті STATISTICANeuralNetworks

Для навчання багатошарових перцептронів у системі STATISTICANeuralNetworks реалізований, перш за все, метод зворотного поширення — зі змінною в часі швидкістю навчання і коефіцієнтом інерції, перемішуванням спостережень перед черговим кроком алгоритму і додаванням адитивного шуму z -ія робастного узагальнення. Крім цього, у системі реалізовано два швидкі алгоритми другого порядку — методи зв’язаних градієнтів і Левенберга- Маркара. Останній є незвичайно потужним сучасним алгоритмом нелінійної оптимізації. Ітеративний процес навчання супроводжується автоматичним відображенням поточної помилки навчання і обчислюваної незалежно від неї помилки на перевірочній множині, при цьому показується і графік сумарної помилки. Якщо перенавчання є, це не повинно турбувати користувача: пакет істотоматично запам’ятовує екземпляр якнайкращої мережі, отриманої у процесі навчання, і до цього варіанта мережі завжди можна звернутися, натиснувши відповідну кнопку. Після того, як навчання мережі завершено,

можна перевірити якість її роботи на окремій тестовій множині.

У системі STATISTICANeuralNetworkse інтелектуальні засоби, що дають імогу відрізати шматки від вже наявних мереж і сполучати декілька мережі видно. Так, можна видаляти або додавати окремі нейрони, видаляти з мережі цілком певний шар, а мережі, узгоджені за числом входів/виходів, послі- : зно сполучати одна з одною. Завдяки цим можливостям пакет дає можливість використовувати такі засоби, як зниження 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. 3.24).

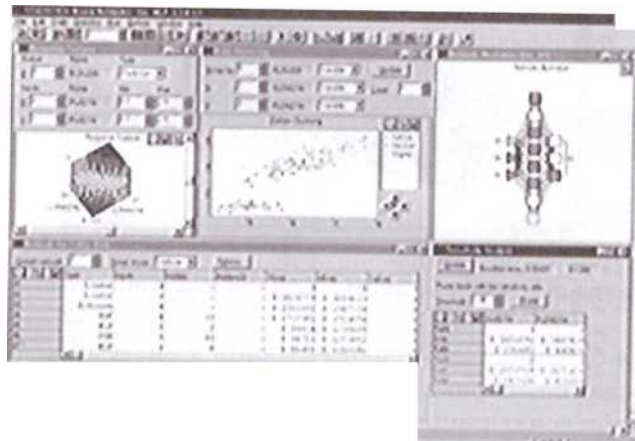


Рисунок 3.24 Переформатування нейронних мереж

Нейропакет NeuroSolutions. Програмний комплекс призначений для моделювання великого набору нейронних мереж. Основна його перевага полягає в гнучкості: крім традиційних нейромережевих парадигм (повнозв'язних і багатошарових нейронних мереж, карт Кохонена, що самоорганізуються) нейропакет включає потужний редактор візуального проектування нейронних мереж; мереж, що дає змогу створювати будь-які нейронні структури і алгоритми навчання, а також вводити власні критерії навчання. Пакет має хороші засоби візуалізації структур, процесів і результатів навчання і функціонування нейронних мереж. Ці переваги ставлять 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 дає можливість використовувати зовнішні модулі при створенні й навчанні нейронної мережі. Пакет підтримує програми, написані мовою C++ для компіляторів Microsoft Visual C++— Borland C++, а також у вигляді DLL-коду. Таким чином, NeuroSolutions є гнучкою відкритою системою, яку можна за необхідності доповнювати і модифікувати. Він містить вбудовану макромову, що дає змогу 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 обрана 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 із ваговими коефіцієнтами.

Нейропакет містить потужні 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 дає можливість використовувати будь-які зовнішні конвертори даних.

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

Нейропакет NeuralWorks Professional II/Plus. NeuralWorks Professional є потужним засобом для моделювання нейронних мереж. У ньому реалізовано 28 нейронних парадигм, а також велику кількість алгоритмів навчання. Додатковий модуль UDND (UserDefineNeuralDynamics) дає можливість створювати власні нейронні структури. Як і NeuroSolutions, NeuralWorksProfessional має хорошу систему візуалізації даних: структури нейронної мережі, зміни помилки навчання, зміни вагів і їх кореляції у процесі навчання, що є унікальною властивістю пакета і корисна при аналізі поведінки мережі. У NeuralWorksProfessional можна інтегрувати зовнішні програмні модулі. Він має вбудований генератор коду, що підтримує компілятор Microsoft Visual.

Нейропакет Brain Maker Pro. Одним з поширених зарубіжних нейросистем є пакет Brain Maker. Наприклад, необхідно розв'язати задачу прогнозу ціни закриття на сьогоднішніх торгах за валютними тисячодоларовими тримісячними ф'ючерсними контрактами. Нехай нас влаштовує точність прогнозу, якої правильно вказується ціновий тренд і

зміна ціни з точністю не нижче 90% від останнього стрибка. Застосування нейронної мережі починається з вхідних даних: курс долара, індекс інфляції, ставка міжбанка, біржові індекси, обсяг торгів, кількість операцій, максимальні і мінімальні ціни та ін. Після попереднього налаштування мережі починається ітераційний процес навчання, в результаті якого нейромережа налаштовує свою логічну структуру для точної реакції ринку на ті або інші дії. З дією метою в пакеті BraizMaker передбачений потужний аналітичний блок, який дає можливість побачити, які параметри чинять позитивний вплив на ситуацію, а які — негативний. Потім мережа знову навчається, тестується на якість і адекватність, а після вдалого тестування використовується для прогнозів. За десять біржових днів мережа жодного разу не помилилася в знаку відхилення ф'ючерсних котирувань, а дев'ять днів із десяти відхилення прогнозу від реальної ціни становило менше 10 доларів. BrainMaker це програма, з якої почалася історія застосування нейронних мереж у 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 та алгоритмів навчання. Останнім часом BrainMaker найчастіше використовується не як самостійна програма, а як надбудова до програми TradeStation для аналізу в режимі реального часу.

Нейропакет NeuroShell2. Пакет є однією з трьох програм, що входять до складу пакета TheAllTrilogy і є універсальним нейропакетом для моделювання декількох найбільш відомих нейронних парадигм: багат шарових мереж, мереж Кохонена та ін. З його допомогою можна вирішувати широкий спектр завдань, починаючи з поширених задач прогнозування курсів акцій і закінчуючи менш поширеними зворотними задачами у геофізиці. Він містить системи для початківця, професіонала, засоби автономного використання. Крім того є пакети доповнень до NeuroShell2, які допомагають істотно спрощувати розв'язки окремих задач. Зокрема такі.

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

Прогноз результатів перегонів. Виконує передобробку статистичних даних про результати перегонів для того, щоб зробити прогнози точнішими. Програма бере дані про результати всіх коней в одному забігу і перетворює їх на файл з великою кількістю рядів, у кожному з яких приведені результати тільки для 2 коней. За такої форми подання даних точність прогнозів нейромережі значно підвищується. Це доповнення застосовне не тільки для прогнозу результатів перегонів, а й для розв'язання інших задач, пов'язаних з ранжуванням даних.

Прогнозування результатів виборів (ранжирування) дає змогу успішно застосовувати нейронні мережі для прогнозування результатів

виборів за наслідками ряду регулярних соціологічних опитувань, що передували їм. Вхідні дані можуть включати дані про результати опитувань, однакові для всіх кандидатів, а також дані для кожного кандидата. Кожен тренувальний приклад може містити дані одного опитування для декількох кандидатів, що беруть участь у виборах. Кількість кандидатів у різних дослідженнях може бути різною. Це доповнення здійснює передобробку даних так, щоб мережа могла порівнювати за один раз тільки по два кандидати. Це спрощує тренування мережі і збільшує точність прогнозів.

Складовою нейропакета є програмний комплекс **NeuroShellTrader**— система, призначена для прогнозування і пошуку ефективних торгових стратегій на фінансових ринках. NeuroShellTrader— це родина продуктів, розроблена спеціально для трейдерів і покликана допомогти їм в ухваленні рішень при торгівлі. У них реалізовані технології штучного інтелекту, що дають можливість 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 зрозуміла навіть для користувача, що не є професіоналом у цій галузі. Крім того, будучи спеціалізованим інструментом для трейдерів, NeuroShellTrader має дружній графічний інтерфейс, багаті можливості для імпорту даних і могутню бібліотеку індикаторів. Він дає можливість відображати біржові дані на робочих листах у вигляді графіків, діаграм, японських свічок, може 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. NeuroShellTrader дає змогу користувачеві отримати важливу статистику та іншу інформацію, що стосується застосування торгових стратегій і нейронних мереж до даних.

NeuroShellSeries. Нейромережева архітектура, яка лежить в основі програм цієї серії, є найостаннішими досягненнями наукового пошуку, результатом якого виявилось створення алгоритму “самопобудови” нейронної мережі, то володіє рекордними швидкостями навчання. Ці програми надзвичайно прості у використанні. Тепер користувач повинен зосередитися тільки на формулюванні завдання, все інше програми цієї серії зроблять самі. До складу серії входять:

- NeuroShell Predictor — Провісник;
- NeuroShell Classifier — Класифікатор;
- NeuroShellRun-timeServer Засоби автономного використання мереж, отриманих в NeuroShellPredictor і NeuroShellClassifier.

Пакет NeuroShellPredictor дає можливість створювати системи для

розв'язання задач прогнозування на основі наявної бази даних. Це можуть бути прогнози наступних значень параметрів часового ряду, наприклад, прогноз курсу акцій, або оцінка будь-якої величини, визначуваної набором незалежних чинників, зокрема оцінка вартості квартир або вживаних.

NeuroShellPredictor настільки простий у використанні, що відпадає необхідність у інструкції для користувача. Замість неї у програмі є “Інструктор”, який проведе через всі етапи створення моделі для прогнозів. Він дає можливість вибрати стовпці у файлі даних, які використовуватимуться як входи, і вказати один стовпець як вихід.

Пакет NeuroShellClassifier призначений для розв'язання задач розпізнавання образів, пов'язаних із визначенням належності образу (ситуації) до тієї або іншої категорії. Наприклад, щодо набору біржових показників виробляти сигнал для купівлі або продажу акцій тієї або іншої компанії. Пакет був розроблений, подібно до NeuroShellPredictor, для тих, хто не мав попереднього досвіду роботи з нейронними мережами. На відміну від програми NeuroShellPredictor, яка на виході нейронної мережі дає безперервне значення величини, що передбачається, нейронна мережа NeuroShellClassifier має декілька виходів, які визначають вірогідність належності пред'явленого образу до кожної з декількох категорій. Як приклади категорій можна навести такі, як {купити, продати, утриматися від операцій} або {ракова пухлина, доброякісна пухлина}. У NeuroShellClassifier реалізовано два ефективні алгоритми класифікації: один є спеціальною нейронною мережею, а інший — статистичним класифікатором, керованим генетичним алгоритмом. Ці алгоритми були розроблені спеціально для того, щоб зробити моделі класифікації точнішими, а також щоб “не ломати голову” кожного разу при створенні моделі, добираючи її параметри.

Пакет NeuroShellRun-timeServer містить низку програм, які дають можливість використовувати мережі, створені за допомогою NeuroShellPredictor і NeuroShellClassifier, або з робочих листів Microsoft Excel, або у власних програмах.

Програмний комплекс DEDUCTOR. DeductorStudio є аналітичним ядром платформи Deductor. Він містить повний набір механізмів імпорту, обробки, візуалізації й експорту даних для швидкого й ефективного аналізу інформації. У пакеті використовуються найпотужніші технології, такі як багатовимірний аналіз, нейронні мережі, дерева рішень, карти Кохонена, спектральний аналіз і багато інших. При цьому акцент зроблений на самонавчальних методах і машинному навчанні, що дає змогу будувати адаптивні системи, тобто здатні реагувати на зміну ситуації. Використання самонавчальних методів і майстрів для налаштування дозволяє знизити вимогу до підготовки персоналу, роблячи сучасні технології доступними широкому колу користувачів. Всі механізми уніфіковані і виконуються за допомогою майстрів.

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 за алгоритмом зворотного поширення помилки. Основним призначенням бібліотеки є інтеграція нейронних мереж в інформаційні системи з метою розширення аналітичних можливостей систем. Реалізація нейронних мереж у вигляді компонентів, наявність відкритого коду дає можливість легко вбудовувати їх в інші програми, ефективність, — це цільовий показник, а всі інші, що впливають на нього, називатимемо чинниками (причому значення показника і чинників можуть бути як числовими, так і символічними). Початкові дані аналізу подаються у вигляді таблиці, один зі стовпців якої — цільовий показник (наприклад, прибуток або обсяг продажу), а інші — чинники впливу (витрати на рекламу, пора року, регіон тощо). Пакет будує кількісну модель залежності значень показника від значень чинників, після чого надається можливість проводити перехресний аналіз (проглядати в графічній і аналітичній формі залежність модельованого показника від будь-якого з обраних 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. У цьому прикладі досліджувався вплив різних чинників (регіон, масштаб і види діяльності) на ефективність роботи філії компанії. До регіональних чинників можна віднести категорію регіону і тип населеного пункту, де розташована торгова філія: 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 філії. Для ефективного управління компанією необхідно розуміти, від чого залежать витрати в кожній філії та що є вирішальним чинником. На підставі таблиць, що характеризують дані за філіями, 4Thought дає можливість будувати модель впливу обраних стовпців чинників на величину витрат філії. Це досить складне завдання, яке неможливо вирішити традиційними статистичними методами. Система використовує дані про роботу філій для навчання нейронної мережі, яка моделює залежність їхніх витрат від різних чинників. Готову модель можна використовувати для кількісного аналізу впливу чинників на витрати філій. Зокрема, побудувавши залежність модельних значень витрат від обороту (перехресний аналіз), можна оцінити результат економії на масштабах: чим більший оборот, тим менші витрати. Модель дає змогу досліджувати навіть ті чинники, які не були в неї введені. Наприклад,

якість управління філією не можна пов'язувати безпосередньо з її прибутковістю — на прибутковість впливають й 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 й інших чинників. Якщо очікувані витрати вищі від фактичних, отже, ефективність роботи філії вище середньої, а це, напевно, 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? Як вплине впровадження нової технології на доходи підприємства і продуктивність співробітників? Відповіді на ці питання складно — термін окупності часто дуже великий, а на ефективність роботи компанії впливають й інші чинники. За приклад візьмемо мережу магазинів роздрібної торгівлі, де встановлювалася інформаційна система для касових терміналів на 2000 точок збуту. Обсяг інвестицій становив 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 філії, нещодавніх витрат на ремонт та ін. На всіх торгових точках були зібрані дані про перераховані чинники, а також інформація про зростання доходів і про те, скільки місяців минуло з моменту впровадження інформаційної системи. Щоб оцінити ефективність впроваджуваної системи, потрібно було побудувати модель зростання доходів залежно від комбінації цих чинників. За допомогою системи 4Thought був проведений аналіз взаємозалежності з метою одержання кількісної оцінки дії кожного чинника окремо.

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

Оптимізація цін. Один із найкорисніших додатків нейроаналітики — оптимізація цін. Відомо, що чим вища ціна на товар, тим більший прибуток від його продажу. Проте у міру підвищення ціни обсяг продажу 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 продаж за минулий період з'ясовується, як змінюються витрати і обсяги продажу залежно від ціни на товар. Величина оптимальної ціни залежить від реакції клієнтів на зміну цін, від собівартості обсягу продукції, а також від зовнішніх чинників.

Компанія, що займається ремонтом автомобілів, використовувала 4Thought для встановлення оптимальних цін на свої послуги. Щоб визначити чутливість клієнтів до зміни цін, система повинна встановити, як у минулому змінювався попит на продукт залежно від різних чинників. Щодо компанії з ремонту автомобілів, то необхідно було змодельювати зміну попиту залежно від цін на послуги самої фірми, середнього рівня цін конкурентів, пори року, витрат на рекламу, а також певних змін у законодавстві, які вплинули на структуру ринку. Були зібрані дані за три роки. Інформацію про ціни конкурентів вдалося отримати від менеджерів філій, які добре знайомі з діяльністю інших фірм. Щоб визначити, які чинники впливають на попит, була побудована модель зміни обсягу продажу відповідно до цін на послуги, рівнем цін конкурентів, порою року і витратами на рекламу. Ця інформація використовувалася для оцінки співвідношення ціна/обсяг продажу.

На рис. 3.25 показано, як зростає попит на послуги (вертикальна вісь) за зниження цін (горизонтальна вісь). Щоб визначити оптимальну ціну, необхідно побудувати модель витрат, що показує, яким чином змінюються витрати залежно від обсягу продажу. Витрати — розрахункова величина, яку можна ввести в модель за допомогою обчислюваного стовпця. Комбінуючи моделі продажу і витрат, можна отримати повну картину впливу різних чинників на прибутковість бізнесу. Зокрема, можна встановити залежність між ціною і прибутком.

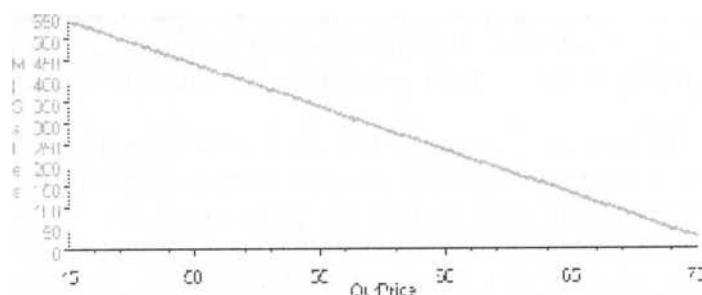


Рисунок 3.25 - Залежність попиту від цін на послуги

У такий спосіб вдається визначити оптимальну ціну для будь-якої іншої галузі бізнесу. Компанія з ремонту автомобілів використовувала отриману інформацію для встановлення цін на свої послуги, при цьому

тільки за рахунок правильного ціноутворення їй вдалося підвищити прибуток на декілька мільйонів доларів.

Система SENNSales. Цей 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 у таблицях. SENNSales використовується для розв'язання задач профілізації клієнтів, маркетингового аналізу, прогнозування попиту та ін. Розглянемо досвід її застосування для прогнозування фінансових рядів.

Надійне і точне прогнозування ф'ючерних курсів обміну валют, кредитних ставок, цін і динаміки продажу дає можливість укласти вигідні міжнародні валютні угоди, формувати фінансові й інвестиційні пакети, давати точну оцінку поточної ситуації на ринку та ін. Дані за минулі періоди (індекси, курси обміну, кредитні ставки і криві продажу) 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 цієї 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 кредитні ставки для кожної з країн (Англія, Франція, Німеччина, Японія і США). При цьому були взяті до уваги зміни в системі рахунків у Європі та країнах "великої сімки", викликані появою на фінансових ринках нової європейської валюти. За традиційного підходу до управління портфелем прогнозується дохід на кожну зі 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.

3.4 Сучасна практика та перспективні напрями застосування нейротехнологій

Нейронні мережі дедалі частіше застосовуються в реальних бізнес-додатках. У деяких сферах, таких як виявлення фальсифікацій і оцінка ризику, вони стали безперечними лідерами серед використовуваних методів. У системах прогнозування і маркетингових досліджень постійно зростає ефективність застосування цих систем. Варто зазначити, що оскільки економічні, фінансові і соціальні системи дуже складні та є результатом дій і протидій різних процесів, то дуже складно створити

повну математичну модель з урахуванням усіх можливих дій і протидій. Практично неможливо детально апроксимувати модель, засновану на таких традиційних параметрах, як максимізація корисності або максимізація прибутку.

У системах подібної складності природно і найбільш ефективно використовувати моделі, які безпосередньо імітують поведінку суспільства і економіки. А це якраз те, що здатна запропонувати методологія нейронних мереж. Розглянемо деякі важливі сфери, в яких ефективність застосування нейронних мереж доведена на практиці.

Застосування нейронних мереж у менеджменті. Протягом останнього десятиліття в журналах і газетах, таких як “ManagementScience”, “ManandCybernetics”, “DecisionSciences”, “Computers&OperationsResearch”, “EuropeanJournalofOperationalResearch” було надруковано безліч пропозицій із застосування нейронних мереж у бізнесі і дослідженні операцій. Більшість варіантів застосування нейронних мереж у менеджменті стосуються задач, що належать до таких чотирьох категорій: класифікація, побудова емпіричної кривої і аналіз часових рядів, кластеризація і оптимізація. Нижче наведені приклади кожної категорії.

1. Класифікація. Належним чином розроблена нейронна мережа може використовуватись як класифікатор. Після навчання розпізнавання історичних даних нейронна мережа може визначати клас належності деякої характерної межі. Нейронні мережі можна використовувати при аналізі кредитоспроможності, щоб передбачити банкрутство фірми. Нейронні мережі можуть також і пінити 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 розглядатись як побудова емпіричної кривої. Крім того, нейронні мережі можуть використовуватись для визначення моделі коливань часового ряду. Аналітики сфери маркетингу, використовуючи нейронні методи мережі, можуть визначати ринкові функції відгуку, основані 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 апріорного знання класів. Ідентифікація споживчих сегментів і групування технологічних деталей можуть бути прикладом у цій прикладній категорії.

4. Оптимізація. Оскільки процес навчання в нейронних мережах

повинен мінімізувати наперед певну помилку або енергію, то нейронні мережі можуть використовуватися для розв'язання задач оптимізації. Задачі на зразок оптимального планування робіт, оптимального планування роботи магазину і мінімізації втрат можуть бути розв'язані з використанням нейронних мереж.

Прогноз ризиків і рейтингування. Є дві базові інвестиційні стратегії: активна, основана на прогнозах прибутковості тих або інших активів, і пасивна, в якій ринок вважають непередбачуваним, і головною метою є мінімізація ризиків. Оцінка інвестиційного ризику, таким чином, є одним з наріжних каменів фінансового аналізу. Ці методики використовують два основні підходи: навчання з учителем на прикладах експертних оцінок або збанкрутілих фірм, і навчання без учителя — шляхом катетеризації наявних даних.

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

Рейтинг корпоративних облігацій. Істотну частину ринку цінних паперів становлять корпоративні облігації — позики корпорацій під фіксований відсоток. Тільки на Нью-Йоркській фондовій біржі в 2005 р. перебували в обігу облігації близько 15 000 компаній із загальною номінальною вартістю понад 260 млрд дол. Для оцінки ризику невиконання відсотків або неповернення грошей за облігаціями 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 надають свої фінансові звіти цим двом організаціям. Рейтинги цих агентств надзвичайно авторитетні, від них безпосередньо залежать процентні ставки за облігаціями: чим нижчий рейтинг емітента, тим дорожче обходиться емітенту обслуговування свого боргу, оскільки інвестори бажають отримати плату за додатковий ризик. До того ж у 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, більше того, агентства стверджують, що він не оснований у 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 зацікавлені у володінні власними алгоритмами рейтингування, що “емулюють” рейтинги великої двійки, — принаймні з трьох причин. По-перше, не для кожної облігації є офіційний рейтинг. Багато паперів, що залишились без уваги великих рейтингових агентств, можуть у результаті

виявитися дуже привабливими для інвестицій, якщо зуміти грамотно оцінити ступінь їх ризикованості. Подруге, 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 на поліпшення якості оцінки фінансового стану емітентів шляхом інтенсивнішого статистичного аналізу, отримавши таким чином перевагу над тими, хто користується офіційними рейтингами.

Наведені вище доводи обґрунтовують таку постановку задачі для нейроаналізу на основі загальнодоступної фінансової звітності компаній-емітентів спробувати відтворити рейтинги “Standard&Poog’s” або “Moody’s”. Незважаючи на наявність неформальної компоненти, вірогідно, алгоритмічна складова цих рейтингів досить велика. Врешті-решт, загальна кількість аналітиків у обох провідних агентствах разом узятих не перевищує 100 осіб. Отже, впоратися 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 ХХст. і базувалися на методі лінійної регресії. Типовий відсоток вгадування рейтингу в цих моделях становив приблизно 60 %. Оскільки можливості нелінійного нейромережевого моделювання ширші, не дивно, що перші спроби застосувати нейромережі показали набагато кращі результати — на рівні 88 % для відтворення окремої градації рейтингу. Складніші нейромережеві моделі здатні з прийнятною точністю відтворювати широкий діапазон рейтингів облігацій за набором ключових фінансових індикаторів фірм-емітентів. Незважаючи на непогані загалом результати, такі нейромережеві моделі дуже компактні. Як вхідні змінні зазвичай використовуються 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&Poog’s”, досягнута цією моделлю.

Оцінка акцій. На відміну від облігацій акції корпорацій не гарантують повернення відсотків і основної суми боргу. Проте оцінка перспективності різних активів у пакетах акцій є одним з головних завдань будь-якого інвестора. Зтіним з підходів до цього є рейтингування акцій, основане на ширшому колі Фінансових показників компаній, доступних з їх фінансових звітів. Результативність такого підходу ілюструють рейтинги провідного консультативного агентства СТТТА з інвестицій в акції ValueLine. Раз на тиждень це агентство її збиває акції близько 1700 компаній за 5 рейтинговими категоріями. Статистичні дослідження підтверджують значущість рейтингу ValueLine. А саме пакети, що складаються з акцій вищої рейтингової категорії, систематично дають

великий прибуток протягом найближчого кварталу. Є підстави припускати, що квартальні звіти корпорацій впливають на курс акцій. Зокрема, несподівано високі прибутки (збитки) статистично значущо корелюють з підвищенням (зниженням) курсів акцій. Причому ця кореляція існує достатньо довго — протягом принаймні двох місяців від 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 є кількісне визначення фінансового благополуччя. Можна, як і у випадку з облігаціями, скористатися для навчання мережі рейтингами, наприклад згаданого вище агентства ValueLine для відтворення цієї, загалом суб'єктивної, оцінки компанії. Можна спробувати використовувати як індикатор благополуччя об'єктивніший критерій — ринковий курс акцій у найближчому майбутньому. Проте ринковий курс може бути схильний до значних флуктуацій суто спекулятивного характеру. Нарешті, можна скористатися вказівками найсудовішого вчителя, досліджуючи крайню форму прояву фінансового "нездужання" банкрутство. Аналіз банкрутства, таким чином, може слугувати джерелом об'єктивних оцінок стійкості фінансового становища фірм.

Прогноз ризиків банкрутств. Наведемо декілька цифр, що ілюструють ціну питання". Світовий ринок тільки міжбанківських кредитів оцінюється в 58 трлн дол. США. Це майже в два рази перевищує світовий обсяг цінних паперів. Природно, що оцінка ризику неповернення кредитів має для банків першочергове значення. Кількість банкрутств у США протягом 2000—2005 рр.

Зростала щорічно приблизно на 14 %. Отже, прогноз банкрутств, особливо в кризових економічних умовах, є основним завданням економічного аналізу.

Якщо в проблемі рейтингування завданням нейромережі було відтворення думки експертів про надійність корпорації, то нейромережевий прогноз банкрутств ґрунтується на статистичній обробці конкретних прикладів банкрутств. У такій постановці задачі нейромережі важливо самій стати експертом, що визначає фінансову стабільність корпорації, ґрунтуючись лише на об'єктивній інформації — показниках фінансової звітності. Зазвичай нейромережа повинна оцінити вірогідність банкрутства через певний проміжок часу за доступною на цей момент фінансовою звітністю. Як входи використовують фінансові індикатори — співвідношення балансових статей, що найбільш повно відображають певні сторони фінансового становища фірми. Найбільш загальний підхід полягає у використанні як входів логарифмів укрупнених статей балансів і звітів про прибутки (збитки). Нейромережа в цьому випадку сама обере

найбільш значущі лінійні комбінації входів, яким відповідатимуть найбільш значущі співвідношення різних статей у потрібних пропорціях. Використання індикаторів, з іншого боку, допомагає в інтерпретації результатів нейромоделювання, якщо скористатися, наприклад, технікою проріджування зв'язків і витягання правил. Зазначимо, що використання описаних індикаторів також лежить в основі загальноприйнятої методики рейтингування банків CAMEL.

Узагальнюючи досвід порівняльного аналізу прогнозів банкрутств різними методиками, зазначимо:

- нейромережеве моделювання забезпечує якнайкращу точність прогнозу банкрутств: близько 90 % порівняно з 80—85 % точності для інших статистичних методик;
- за бажання можна підвищити “підозрілість” нейромережі, забезпечив точність виявлення банкрутів аж до 99 % — за рахунок зниження вимог до іомилок;
- банкрутства можна впевнено передбачати за декілька років до їх фактичного настання, причому точність прогнозу за два роки практично не відрізняється від точності прогнозу за рік. Отже, неявні сигнали неблагополуччя наявні у фінансовій звітності фірми задовго до її краху.

Корисність навчання мережі на прикладах збанкрутілих фірм полягає також у тому, що така мережа виробляє функцію — числовий 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 тривалому існуванні фірми, й в отриманні достатнього прибутку. Важливим є не тільки стан фірми на сьогодні, а і характеристики наявних тенденцій. Значущим може бути інший набір чинників, що дає іншу оціночну функцію. Так, висока прибутковість може забезпечити підвищення надійності в майбутньому. Але, незрозуміло як можна навчати нейромережу на “майбутній успіх” за відсутності такого самого чіткого критерію ітгіху, яким є банкрутство для невдачі. Ці об'єктивні труднощі можна, якщо пригадати, що фірма існує не сама по собі, а в співтоваристві подібних до неї 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 із показниками решти учасників цього ринку. Такий підхід не потребує знання готових відповідей, оскільки оснований на навчанні без учителя.

Порівняльний аналіз фінансового стану фірм. Порівняльний аналіз, на відмінну від рейтингування, припускає введення не однієї, а декількох оціночних координат. Це дає змогу краще використовувати наявну інформацію, точніше позиціонувати фірму серед інших. З іншого боку, для осяжності результатів порівняльного аналізу кількість параметрів

порівняння повинна бути за можливості мінімальною. У вузькому значенні “осяжність” потребує 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.

У математичному аспекті ця задача зводиться до оптимального стиснення інформації про фінансовий стан фірми, тобто відображення інформації мінімальною кількістю параметрів при заданому рівні огрублення або мінімізації втрат інформації при заданому числі узагальнених координат. З метою візуалізації вигідно обмежитися двопараметричним уявленням. Це вже істотний крок вперед порівняно 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 за 30 фінансовими показниками — відношенням балансових статей до загальної суми активів банку. Подібна нормалізація приводить всі статті до єдиного масштабу, згладжуючи відмінності між великими і дрібними банками. Згідно з таким підходом надійність банку будемо характеризувати одним фінансовим показником — відношенням власного капіталу до залученого. Більш загальний підхід — використання не двох окремих компонент, а двох лінійних комбінацій усіх 30 початкових параметрів, які найкраще представляють наявні дані (рис. 3.26).

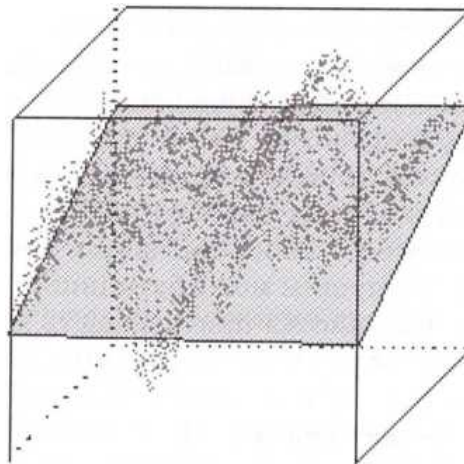


Рисунок 3.26 - Лінійна апроксимація багатовимірних даних

Кожен банк представлений точкою у 30-вимірному просторі і задача полягає у 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 дуже відхиляється від площини, цей метод може дати непогане початкове наближення. Проте виявляється, що в цьому випадку не так. Середньоквадратичне відхилення для випадку двох головних компонент дорівнює майже половині від загальної дисперсії. Отже, навіть оптимальний варіант лінійного стиснення не дає можливості візуалізувати фінансове становище банків.

У цій ситуації природно звернутися до нелінійного статистичного аналізу, тобто до нейромережевого моделювання. Нагадаємо, що методом, який 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 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 зі схожим фінансовим станом. Чим далі на карті координати банків, тим більше відрізняється один від одного їх фінансовий портрет.

Розташування на карті банків із відкликаною ліцензією. Переваги карти Кохонена починають виявлятися після нанесення на неї будь-якої графічної інформації. На рис. 3.27 наводиться зразок карти Кохонена, на якій відмічені ячейки, що містять банки з відкликаними за наслідками 2004 р. ліцензіями. Видно, що банки з відкликаними ліцензіями групуються в правому верхньому кутку карти — “зоні ризику”. Ми побачимо, що ця зона має й інші ознаки неблагонадійності.

Зазначимо, що на відміну від аналізу банкрутств інформація про банкрутства не брала участь у навчанні мережі. Вона зображена на вже готовій карті, гудучи лише індикатором області параметрів із підвищеним ризиком банкрутства. Ця особливість описуваної методики дає змогу виявити галузь ризику ітосовно невеликої кількості прикладів.

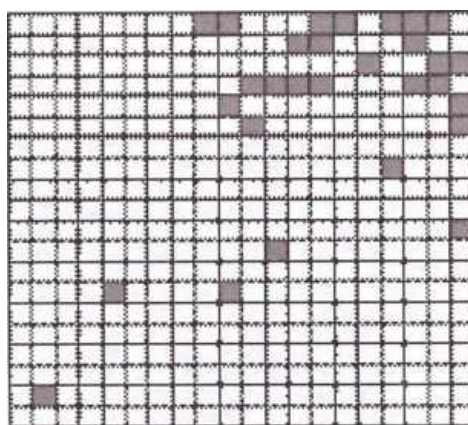


Рисунок 3.27 - Ячейки, що містять хоч один банк із відкликаною в 2004 р. Ліцензією

Карта розмірів банків. Розглянемо, наприклад, як розташовуються на побудованій карті банки різних розмірів (рис. 3.28). Розміри банків беруться логарифмічній шкалі, причому клітини, що відрізняються на одну колірну градацію, містять банки з п'ятикратним співвідношенням активів. Нагадаємо, що величина активів банків була спочатку виведена з набору параметрів, оскільки вона використовувалася для нормування решти статей. Незважаючи на це, банки різних розмірів розташовуються не хаотично, а регулярно, що свідчить про значущість розміру банку при виборі ним своєї фінансової стратегії.

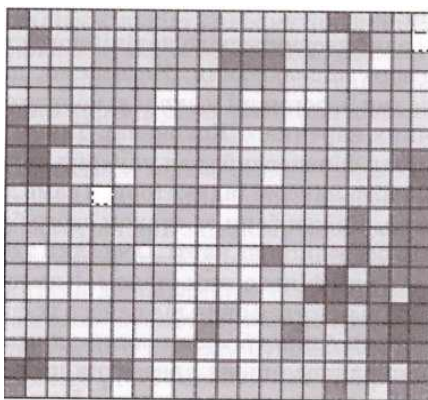


Рисунок 3.28 - Розмір активів українських банків

Візуально на карті можна виокремити такі великі групи банків: великі (низ карти), малі (група ліворуч і група праворуч) і середні банки.

Забарвлення, що відображає відносний розмір статутного фонду показує, що між двома групами малих банків є істотні відмінності: банки в нижньому правому кутку карти практично.

Порівняння з розташуванням банків-банкрутів показує, що вірогідність банкрутства як великих, так і малих банків у 2004—2005 рр. була невелика.

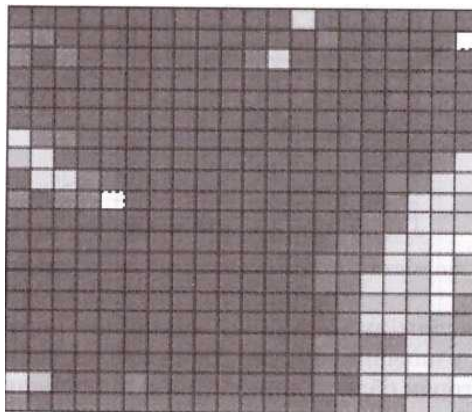


Рисунок 3.29 - Статутний фонд банків

Наведені вище забарвлення в сукупності утворюють атлас, що відображає фінансовий стан банків або інших фірм, що займаються схожими видами бізнесу. Цей атлас дає графічне відображення становища будь-якої конкретної фірми серед конкурентів і може використовуватися як зручний засіб фінансового аналізу. Зокрема, можна розглядати еволюцію фінансового стану окремої фірми в часі, виявляти наявні тенденції і цикли. З погляду макроекономіки зручність таких карт полягає в тому, що площі на цій карті приблизно пропорційні частці фірм через більш-менш рівномірне заповнення комірок. Отже, можна наочно уявляти собі, наприклад, частку великих банків або банків, що зазнають труднощів із поверненням кредитів.

Передбачення фінансових часових рядів. Прогноз фінансових часових рядів — необхідний елемент будь-якої інвестиційної діяльності. Сама ідея інвестицій — вкладання грошей зараз із метою одержання 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. Наведемо декілька цифр, що ілюструють масштаб цієї індустрії прогнозів. Денний оборот ринку акцій тільки в США перевищує 40 млрд дол. Депозитарій DTC (Depositary Trust Company) в США, де зареєстровано цінних паперів на суму 21 трлн дол., реєструє на день операцій приблизно на 750 млрд дол. Ще активніше йде торгівля на світовому валютному ринку FOREX. Його денний обіг перевищує 1000 млрд дол. Це приблизно 1/50 всього сукупного капіталу людства. Відомо, що 99 % всіх операцій — спекулятивні, тобто спрямовані не на обслуговування реального товарообігу, а розміщені з метою витягання прибутку за схемою “купив дешевше — продав дорожче”. Всі вони ґрунтуються на прогнозах зміни курсу учасниками операції. Причому, прогнози учасників кожної операції протилежні один одному. Отже, обсяг спекулятивних операцій характеризує ступінь відмінностей у прогнозах учасників ринку, тобто реально — ступінь непередбачуваності фінансових часових рядів.

Методика передбачення часових рядів. Почнемо з етапу завантаження. Незважаючи на те, що передбачення, здавалося б, є екстраполяцією даних, нейромережі, насправді, розв’язують задачу інтерполяції, що істотно підвищує надійність розв’язку. Передбачення часового ряду зводиться до типової задачі нейроаналізу — апроксимації функції багатьох змінних за їх даним набором прикладів — за допомогою процедури занурення ряду в багаті вимірний простір. Наприклад, d -мірний лаговий простір ряду X_t складаються із d значень ряду в послідовні моменти часу: $X_{t-d} = (X_{t-1}, \dots, X_{t-d})$.

Для динамічних систем доведена теорема Такенса. Якщо часовий ряд

динамічною системою, тобто значення X_t є довільною функцією стану такої системи, існує така глибина занурення d (приблизно дорівнює гтективному числу мір свободи цієї динамічної системи), яка забезпечує однорічний прогноз наступного значення часового ряду. Таким чином, вибравши самале 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 за набором прикладів, заданих історією цього часового ряду. Навпаки, для випадкового ряду знання минулого нічого не дає для передбачення майбутнього. Тому, згідно з теорією ефективного ринку, розкид значень ряду, що передбачається, на наступному кроці при зануренні в лаговий простір не зміниться.

Для навчання нейромережі недостатньо сформувати навчальні набори входів-виходів. Необхідно також визначити помилку прогнозів мережі. Середньоквадратична помилка, використовувана за замовчуванням у 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 із можливою нормою прибутку. Наприклад, для вибору ринкової позиції надійне визначення знака зміни курсу. Важливіше, ніж пониження середньоквадратичного відхилення. Хоча ці показники і пов’язані між собою, мережі, оптимізовані за одним із них, даватимуть гірші прогнози іншого.

Основна специфіка передбачення часових рядів полягає у галузі передоброби даних. Процедура навчання окремих нейромереж стандартна. Як завжди, наявні приклади розбиваються на три вибірки: навчальну, валідаційну. Перша використовується для навчання, друга — для вибору оптимальної архітектури мережі або для вибору моменту зупинки навчання. Нарешті, третя, яка взагалі не використовувалася у навчанні, слугує для контролю якості прогнозу навченої нейромережі. Проте для дуже зашумлених фінансових рядів істотну перевагу в надійності прогнозів може надати використання комітетів мереж. Поліпшення якості прогнозів можливе також за рахунок використання нейромереж з зворотними зв’язками. Такі мережі можуть володіти локальною пам’яттю, що зберігає інформацію про далеке минуле, ніж те, що є явно на входах.

Як вхідні змінні логічно вибрати найбільш статистично незалежні величини, наприклад зміни котирувань $5C_t$ або логарифм відносного приросту $\log(C_t / C_{t-x}) = \Delta C_t / C_{t-1}$. Останній вибір вдалий для тривалих часових рядів, коли вже помітний вплив інфляції. У цьому випадку прості різниці в різних частинах ряду матимуть різну амплітуду, оскільки фактично вимірюються в різних одиницях.

Одним з “слабких місць” у фінансових прогнозах є дефіцит прикладів для навчання нейромережі. Фінансові ринки загалом не стаціонарні. З’являються нові фінансові інструменти, для яких ще не накопичена історія, змінюється характер торгівлі на колишніх ринках. У цих умовах тривалість доступних для навчання нейромережі часових рядів дуже обмежена. Проте можна збільшити кількість 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. Ідеться про генерацію штучних прикладів, які отримують із вже наявних шляхом застосування до них різних перетворень.

Наприклад, психологічно виправдане таке припущення: гравці звертають увагу, в основному, на форму кривої цін, а не на конкретні значення за осями. Тому якщо трохи “розтягнути” по осі котирувань весь часовий ряд, то отриманий у результаті такого перетворення ряд також можна використовувати для навчання разом із результатом. Таким чином можна подвоїти кількість прикладів за рахунок використання апріорної інформації, що впливає з психологічних особливостей сприйняття часових рядів учасниками ринку. Більше того, не просто збільшити кількість прикладів, а й обмежити клас функцій, серед яких шукається рішення, що також підвищує якість передбачення.

Особливістю передбачення фінансових часових рядів є прагнення до одержання максимального прибутку, а не мінімізації середньоквадратичного відхилення, як це робиться у разі апроксимації функцій. У простому випадку щоденної торгівлі прибуток залежить від правильно вгаданого знака зміни котирування. Тому нейромережі потрібно орієнтувати саме на точність вгадування знака, а не самого значення. Обчислимо, як пов’язана норма прибутку з точністю визначення знака в простій постановці щоденного входження в ринок.

Як менші, так і більші значення ставок зменшують прибуток. Яня — ризикова гра може призвести до програшу за будь-якої передбачаючої здатності. Цей факт ілюструє рис. 3.30.

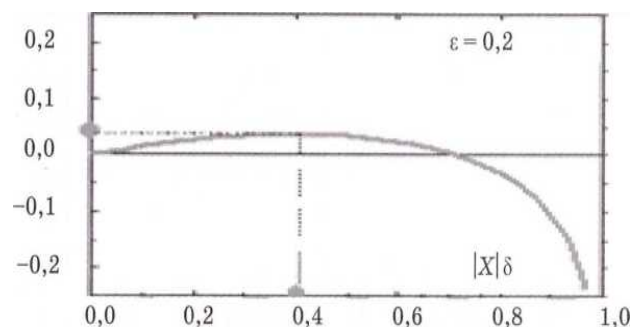


Рисунок 3.30 - Залежність середньої норми прибутку від вибору частки капіталу

Використання комітетів мереж. Внаслідок випадковості у виборі початкових значень синаптичних вагів, прогнози мереж, навчених на одній і тій вибірці, будуть відрізнятися. Цей недолік можна перетворити на перевагу, організувавши комітет нейроекспертів, що складається з різних нейромереж. Розкид у прогнозах експертів дасть уявлення про ступінь достовірності цих передбачень, що можна використовувати для правильного вибору стратегії гри.

Легко показати, що середнє значення комітету повинно давати кращі прогнози, ніж середній експерт із цього ж комітету. Нехай помилка i -го експерта для значення входу дорівнює D_i . Середня помилка комітету завжди менше середньоквадратичної помилки окремих експертів у зв'язку з нерівністю.

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

Тому в прогнозах завжди краще спиратися на середні значення всього комітету (рис. 3.31). Виграш комітету (на рис. позначено кружечками) вищий, ніж виграш середнього експерта. Рахунок вгаданих знаків для комітету 59 : 41.

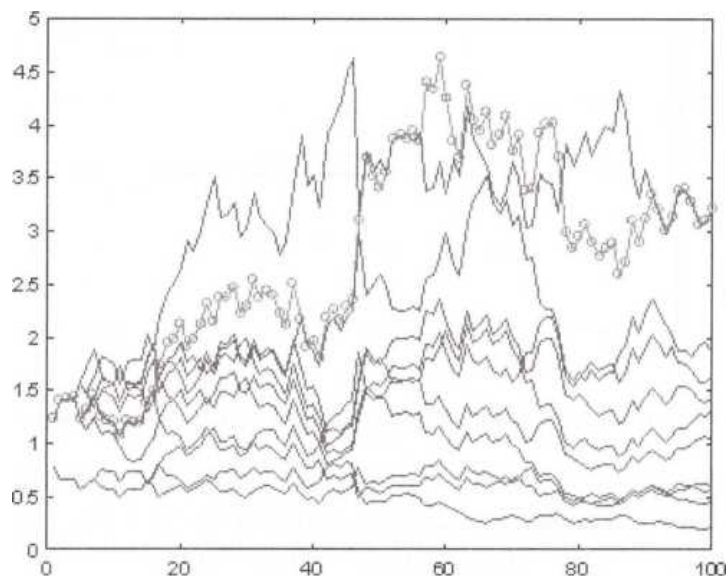


Рисунок 3.31 - Норма прибутку на останніх 100 значеннях ряду при передбаченні комітетом із 10 мереж

Отже, виграш комітету навіть вищий, ніж виграш кожного з експертів. Тож метод комітетів може істотно підвищити якість прогнозування. Звернемо увагу на абсолютне значення норми прибутку: капітал комітету зріс у 3,25 раза при 100 випадках входження на ринок.

Завершуючи огляд нейрокомп'ютерних технологій зупинимося на

деяких перспективних напрямках застосування нейронних мереж.

Штучні нейронні мережі та експертні системи. Останніми роками над штучними нейронними мережами домінували логічні і символічно-операційні дисципліни. Наприклад, широко пропагувалися експертні системи, в яких є багато помітних як успіхів, так і невдач.

Дехто говорить, що штучні нейронні мережі замінять сучасний штучний інтелект, але багато що свідчить про те, що вони існуватимуть, об'єднуючись у системи, де кожен підхід використовується для вирішення тих завдань, які краще виконує. Ця позиція підкріплюється тим, як люди функціонують у нашому світі. Розпізнавання образів відповідає за активність, що потребує 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 запросити додаткову інформацію, та займе більше часу, але якість отриманих у результаті рішень може бути вищою. Можна уявити собі штучну систему, що наслідуює такий поділ праці.

Штучна нейронна мережа реагувала б в більшості випадків відповідним чином на зовнішнє середовище. Оскільки такі мережі здатні визначати довірчий рівень кожного рішення, то мережа “знає, що вона не знає” і передає цей випадок для розв'язання експертній системі. Рішення, що приймаються на цьому вищому рівні, були б конкретними і логічними, але вони можуть потребувати збирання додаткових фактів для одержання остаточного висновку. Комбінація двох систем була б потужнішою, ніж кожна з систем окремо, дотримуючись при цьому високоефективної моделі, що дається біологічною еволюцією.

Надійність нейронних мереж. Перш, ніж штучні нейронні мережі може: буде використовувати там, де поставлено на карту людське життя або цінне майно, повинні бути вирішені питання, що стосуються їх надійності 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 нездійсненна і повинні використовуватися статистичні методи для оцінки функціонування. У деяких випадках це неприпустимо. Наприклад, що є допустимим рівнем помилок для мережі, яка управляє системою космічної оборони. Більшість людей скажуть, будь-яка помилка недопустима, оскільки веде до величезної кількості жертв і руйнувань. Це ставлення не змінюється від тієї обставини, що людина у схожій ситуації також може припускатися помилок. Проблема виникає через 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 розв’язують задачу. Внутрішнє уявлення, що формується в результаті навчання, часто настільки складне, що його неможливо проаналізувати, за винятком найпростіших випадків. Це нагадує нашу нездатність пояснити, як ми впізнаємо людину, незважаючи на відстань, кут освітлення чи минулі роки. Експертна система може прослідкувати процес своїх міркувань у 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, на думку багатьох користувачів, є те, що нейронна мережа розв’язує задачу, але не може розповісти як. Іншими словами, з навченої нейронної мережі не можна витягувати алгоритм рішення задачі. Проте спеціальним чином побудована процедура дає змогу вирішити це завдання.

Розглянемо клас мереж, які вважатимемо логічно прозорими (тобто такими, які розв’язують задачу зрозумілим для нас способом, для якого легко сформулювати словесний опис у вигляді явного алгоритму). Наприклад, визначимо, що всі нейрони мають не більше трьох вхідних сигналів. Використаємо нейронну мережу, 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 розв’язанню задач. Після цього вироблятимемо контрастування у декілька етапів. Процедура контрастування заснована на оцінці значущості вагів зв’язків у мережі. Основною її метою є спрощення технічної реалізації мережі і створення навички мережі, яка була б зрозумілою.

На першому етапі контрастуватимемо лише ваги зв’язків нейронів вхідного шару. Якщо після контрастування в деяких нейронів залишилося більше трьох вхідних сигналів, то збільшимо кількість вхідних нейронів. Потім аналогічну процедуру виконаємо по черзі для всіх останніх шарів. Після закінчення описаної процедури буде отримана логічно прозора мережа (рис. 3.32).

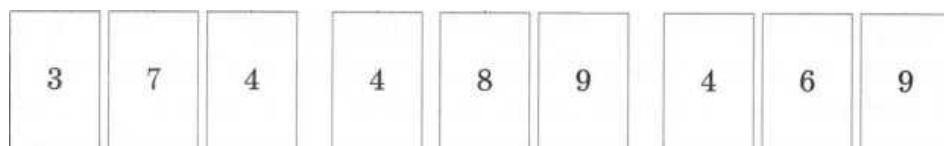


Рисунок 3.32 - Логічно прозора нейронна мережа

Як приклад, наведемо інтерпретацію алгоритму міркувань, отриманого за логічно прозорою мережею. Умова задачі така: за відповідями на 12 питань необхідно передбачити перемогу провладної або опозиційної партії. Відповіді на питання описують ситуацію на момент перед виборами. Негативний сигнал на виході мережі інтерпретується як передбачення перемоги провладної партії. У протилежному випадку відповіддю вважається перемога опозиційної партії. Всі нейрони реалізовували порогову функцію, що дорівнює 1, якщо сума алгебри вхідних сигналів нейрона більше або рівна 0, і -1 за суми, менше 0. Відповідь мережі базується на виявах синдромів політичної нестабільності і поганої політики. Отже, для перемоги провладної партії необхідна відсутність (-1) обох синдромів. На рис. 3.33 наведена структура логічно прозорої нейронної мережі, яка розв'язувала задачу про передбачення результатів виборів президента США.

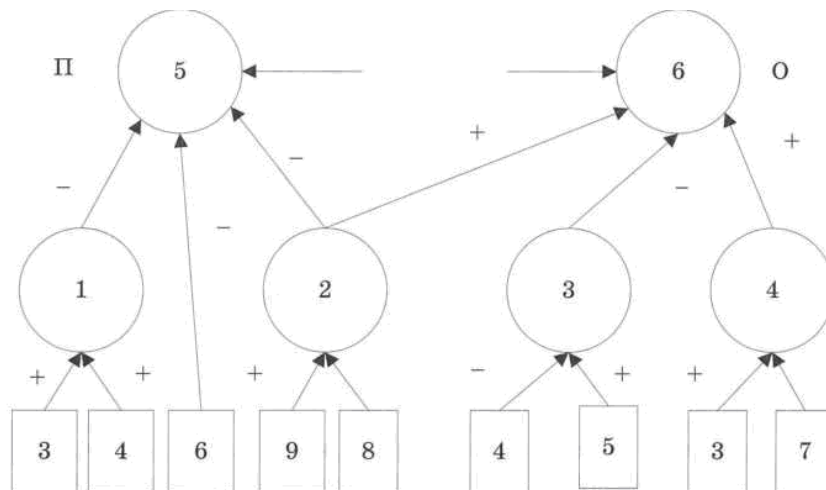


Рисунок 3.33- Логічно прозора нейронна мережа для задачі передбачення

Технологія здобуття явних знань з даних за допомогою навчених нейронних мереж має просту структуру. Перший етап: навчаємо нейронну мережу розв'язувати базову задачу. Зазвичай базовою є задача розпізнавання, передбачення та ін. У більшості випадків її можна трактувати як задачу про заповнення пропусків у даних. Другий етап: за допомогою аналізу показників значущості, контрастування і донавчання приводимо нейронну мережу до логічно прозорого вигляду — так, щоб отриману навичку можна було “прочитати”.

ВИКОРИСТАНА ЛІТЕРАТУРА

1. Нікольський Ю.В. Інтелектуальний аналіз даних – 2012.
2. Долінська Л.В., Ковальчук В.В. Вступ в теорію систем і теорію управління. – Одеса: Одеса: Видавець Букаєв Вадим Вікторович, 2010. – 205 с.
3. Долинская Л.В., Ковальчук В.В., Кузина Ю.В. Введение в теорию систем. Основы системного анализа. – Одеса: Видавець Букаєв Вадим Вікторович, 2011. – 108 с.
4. Долинская Л.В., Ковальчук В.В., Клименко А.М. Введение в теорию информационно-измерительных систем (применение статистических методов). – Одеса: Одеса: Видавець Букаєв Вадим Вікторович, 2011. – 112 с.
5. Ковальчук В.В., Панченко А.А., Жаровцев С.О. Інформаційно-вимірювальні системи. Лабораторний практикум (SciLab). Навчальний посібник. – Одеса: Видавець Букаєв Вадим Вікторович, 2011. – 164 с.
6. Долженков, В.А., Стученков А.Б. Microsoft Office 2007.- СПб.:БХВ-Петербург, 2007.-1200с.
7. Стоцкий Ю., Васильев А., Телина И. Office 2007. Изучаем самостоятельно. - СПб.: Питер, 2008. - 524 с.
8. Кусимов С.Т., Ильясов Б.Г., Исмагилова Л.А., Валеева Р.Г. Интеллектуальное управление производственными системами. – М.: Машиностроение, 2001. – 327 с.
9. Рей У. Методы управления технологическими процессами. – М.: Мир, 1993. – 368 с.
10. Вавилов А.А. и др. Эволюционный синтез систем управления. – Л.: ЛЭТИ, 1983. – 80 с.
11. Вережкин А.П., Дадаян Л.Г. Анализ и синтез автоматических систем регулирования сложных объектов нефтепереработки и нефтехимии. – Уфа: УНИ, 1999. – 94 с.
12. Ковальчук В.В. Основи наукових досліджень. – К: ВД „Слово”, 2009. – 237 с.
13. Hullmann A. Trends in Information and Communication Technology on Disability Policy and Practice. National Council on Disability, Washington DC., 2006.-P.1-55.
14. Барсегян АА. Технологии анализа данных: Data Mining, Visual Mining, Te Mining OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко. СПб. . БХВ Петербург, 2008.
15. Кулаичев А.П. Методы и средства комплексного анализа данных/ А.П. Кулаичев. — М. : ИНФРА-М, 2006.
16. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям / Н.Б. Паклин, В.И. Орешков. — СПб. : Питер, 2009.

Навчальне електронне видання

Ковальчук Володимир Володимирович

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Конспект лекцій

Видавець і виготовлювач

Одеський державний екологічний університет

вул. Львівська, 15, м. Одеса, 65016

тел./факс: (0482) 32-67-35

Е-mail: info@odeku.edu.ua

Свідоцтво суб'єкта видавничої справи

ДК № 5242 від 08.11.2016