Завдання на курсову:

Типове завдання:
9 випадкове завдання
Нейромережа:
10-9
Спосіб генерування штучної змінної:
3-1
При перевірці визначити швидшу нейромережу.

Дати:

Одержання завдання : до 10 вересня
Огляд літератури: 10-30 вересня
Розробка структури вхідних і вихідних даних, та уточнення методів: 1-30 жовтня
Опис і розробка алгоритму симуляції штучної мережі: 1-10 листопада
Дата здачі: 1 грудня

Консультації:

Skype : Nick\_yatsyshyn

Теми лекцій:
1) Можливості інтелектуального аналізу
2) Недоліки технології інтелектуального аналізу даних для Data Mining OLAP
3) Сховища даних
4) Технології інтелектуальних обчислень і апаратне забезпечення
5) Види моделей інтелектуальних обчислень
6) Класифікація. Регресійний аналіз. Прогнозування часових послідовностей.
 Асоціація. Послідовність.
7) Методи інтелектуальних обчислень
8) Нейронні мережі. Дерева рішень.
9) Проблеми у процесі визначення нового знання.
10) Збір та підготовка даних.
11) Побудова моделей.
12) Використання та тестування моделей.
13) Генетичні алгоритми.
14) Задачі оптимізації
15) Нечітка логіка
16) Нечітка логіка
17) Нечітка логіка
18) Заключна лекція

Тема: Штучні нейронні мережі.
Особливості нейронних мереж
- Паралельні обчислення
- Можливість навчання на помилках
- Дуже низьке енергоспоживання
Перші математичні розрахування створення штучного нейрона сягають кінця 40-х років. Дослідження тривали близько 10 років, але в зв’язку з недостатнім розвитком ЕОМ в кінці 50-х фінансування припинилось, і нова хвиля досліджень почалась лише в кінці 80-х років.
Штучний нейрон складається з 4 основних частин:
- Приймаючі входи
- Блок перетворення отриманих сигналів
- Блок обробки сигналів
- Виходи
Приймаючі входи зображаються за допомогою вектора вхідних сигналів:



1) Вхід
2) Суматор
3) Реалізація передавальної функції
4) Вихід

Є 2 способи з’єднання нейронів:
- опосередкований



- Безпосередній



При створенні штучної нейронної мережі, як і в простому нейроні є зворотній зв’язок, який буває збуджуючий (+) та від’ємний (-). Ключовою відмінністю штучних нейронних мереж є можливість навчання. Є 3 способи навчання штучних нейронних мереж:
- З вчителем – означає, що ми маємо 2 вектори: Х={Х1,Х2,…,Хn} даних на
 вході і y={y1,y2,…,yn} даних на виході.
- Без вчителя – це навчання коли є тільки вхідна вибірка даних, а навчання проходить на основі саморегуляційних функції або карт у прихованому шарі.
- Змішана – частина нейронів у прихованому шарі вчаться з вчителем, частина без нього.

Задачі ШНМ(штучної нейронної мережі)
- Класифікація – поділ усієї множини вхідних даних на класи за спільними ознаками.



- Кластеризація - полягає у виділенні кластерів, де об’єднані класи за певною спільною ознакою.


- Апроксимація – виявлення математичної залежності для даних, які описують певний процес, але спотворені шумами.
- Прогнозування – полягає у можливості X(t1),X(t2),X(tn)
- Оптимізація – полягає у мінімізації чи максимізації цілої функції
Асоціативна пам’ять – дана проблема виникла у зв’язку із тим, що коли ми працюємо із програмними продуктами, то виклик до інформації іде тільки за адресом, а це завдання вирішує знаходження необхідної інформації як за адресом, так і за змістом ( навіть якщо він не повний). Керування сигналів полягає в створенні керуючих сигналів для управління технічних сигналів або соціальної системи.
Вимоги до ШНМ:
- Множина даних, які описують проблеми
- Тестова множина даних
- Вибір функції суматора
- Вибір функції передавальної
- Вибір способу навчання
- Вибір інструментального забезпечення навчання мережі
Основні компоненти ШНМ:
- Вагові коофіцієнти – це вектор важливості того чи іншого коду.
 W={w1,w2,…wn} . Все навчання в ШНМ здебільшого звоиться но вивчання вагових коофіцієнтів, котрі є як на входах, так і між нейронами.
- Суматор – зазвичай це є функція виду:

∑inXi\*wi

 - Передавальна ділиться на :
 а) Лінійні
 - Жорсткі
 - Знесичені
 б) Нелінійні (S-подібні): тангенсоїди, синусоїди,…
 - Маштабування даний компонент вирішує задачу зміщення можливості
 зв’язків між ними.

- Вхідна функція – даний компонент призначений для виявлення нейрона (групи нейронів) які будуть мати найвагоміше значення при обрахунку результуючого значення ШНМ.
- Функція похибки – призначена для створення зворотнього зв’язку у межах нейрона чи між ними.
- Функція навчання – полягає у тому, щоб ми вибрали правильний спосіб навчання для одержання необхідного результату.

ШНМ:
- Необхідно створити структуру нейронної мережі: Структура нейронної мережі – це спосіб зв’язків між нейронами
- Створити архітектуру нейронної мережі. Архітектура включає в себе тип нейронів, та зв’язки між ними.
- Вибір філософії (парадигми) – спосіб навчання, передавання функцій, суматори.

Існує 2 величезні групи нейронних мереж:
- Прямого поширення:
1) Перцептрони
2) Back Propagation
3) Мережа зустрічного поширення
4) Карта Кохонена
5) Рекурентні мерехі
6) Мережі Хопфілда
- Зворотнього

Алгоритм навчання перцептрона
1) Ініціалізація даних (таблиця істинності)
2) Введення констант (швидкість навчання η{0;0,1], поріг спрацювання (µ))
3) Обрахунок суматора – шукаємо активаційну функцію (а=∑inXi\*wi). Порівняння активаційної функції з пороговим значенням. Якщо а>θ правдива, то шукаємо δ=t-y. Далі шукаємо ∆w1=ηδX1.
Властивості перцептрона:
1) На входах є бінарні чи аналогові значення
2) На виході ми можемо отримати 2 типи даних: 0 або 1, 1 або -1
3) Використовується для задач класифікації.
 Переваги: простота архітектури і висока швидкість навчання.
 Недоліки: обмеженість сфери застосування.

Єдність – це тип з’єднання у вхідному та вихідному шарах….
Складність зав’язків – це на скільки нам відомо, закон розподілу за яким вхідна(навчальна вибірка ) може бути описана.
Обчислювальна складність – це вид функції суматора, об’єкт порогової функції, та вигляд передавальної функції.
Існує 5 основних правил навчання:
- Правило Хебба – полягає у підвищенні сили ваги нейрона за умови високого вихідного сигналу одночасного, сусідніх нейронів.
- Правило Хопфілда – це модифікація правила Хебба, яке враховує підсилення або послаблення ваги нейрона навіть за умови неодночасного високого(низького) сигналу сусідніх нейронів.
- Правило Дельта – правило за яким працює перцептрон
- Правило градієнтного спуску – суть полягає у зміні сили ваги нейронів, на основі знаходження похідної від передавальної функції.
- Правило змагання – суть цього методу полягає у знаходженні у кожному прошарку (вхідний, вихідний, прихований) одного, максимум 2 нейронів з найвищою вагою.

Back propagation

Дана нейронна мережа відрізняється від звичайного перцептрона активаційними функціями: сигноїда, сигноїда ферма і гіперболічний тангенс

F(s)=1/(1+e-2λs) [1.1 - сигноїда]
F(s)=1/(|s|+λ) [1.2 - сигноїда ферма]
F(s)=(es/λ-e-s/λ/( es/λ+e-s/λ)) [1.3 - гіперболічний тангенс]

Дана мережа отримала декілька модифікацій:

Delta Bar Delta

Використовує метод найкрутішого спуску, який полягає у знаходженні глобальних мінімумів, не зважаючи на локальні. Для цього при знаходженні зміни ваг змінюється швидкість навчання. Для зміни швидкості навчання вводяться такі параметри:
- Величина градієнт зміни
- Константа збільшення швидкості і константа зменшення швидкості

Extended Delta Bar Delta

При розрахунку зміни ваги вводиться коефіцієнт зміни моменту.

Скерований випадковий пошук

В порівнянні з Back propagation для навчання цієї мережі вводяться 4 поняття:
- Випадковий крок – це компонента, яка використовується для знаходження зміни величини ваги нейрона (0…1).
- Крок реверсування – ця компонента використовується за умови, якщо при оцінці величини похибки при навчанні нейромережі збільшується.
- Скерована компонента – застосовується у випадках, якщо перед нейромережею стоїть задача класифікації.
- Саморегулююча відмінність – застосовується з метою неможливості зациклення навчання нейромережі.

Мережа Кохонена

Дана мережа має лише вхідний та вихідний етап. Даний тип нейромереж використовує правило змагання. Щоб реалізувати дане правило використовується теорія графів, а саме – знаходження найменшої відстані.

Dij=∑in(Xi\*ωij)z

Необхідною умовою для даної нейромережі є наперед визначена кількість кластерів. Використовується кластеризація.

Counter Propagation

Дана мережа характеризується наявність шара Кохонена та шара Гросперга. Шар Кохонена працює класично з визначенням найкращого нейрона (1 буде тільки на одному вихідному нейроні мережі Кохонена). Шар Гросперга навчається зчитувати. Шар Гросперга характеризується тим, що швидкість навчання починається з 0.7 і поступово змінюється.

Мережа Кхопфілда

Використовується для визначення задач. Кожен шар у даній нейромережі має однакову кількість нейронів. При виконанні задач з класифікації, необхідно враховувати що кількість класів буде обмежена 15% від кількості нейронів.

Машина Вольцмана

Дана нейромережа при знаходженні похибки навчання використовує теорію ймовірності.

P(c)=e(-c/KT)
K – константа Вольцмана
Т – штучна температура

Штучна температура – це коефіцієнт, що вводиться в межах від 0 до 1 при обрахунку ймовірності змінної та величини зміни ваги.

Мережа Хемінда

Навчання є аналогічне до Back Propagation тільки на вхід її даються бінарні вектори, а результат отримуємо цілими числами. На практиці використовується для підрахунку бітів, які співпадають з навчальним вектором.

Двоскерована асоціативна пам’ять

У якості передавальної функції використовують:

Bj=F∑jaj\*wij
w – матриця ваг між нейронами
F – передавальна функція

Використовується для задач розпізнавання даних із втратами. Дана мережа має обмеження: кількість запам’ятовуваних векторів L :

L=n/(2log2n)

Мережа адитивної резонансної теорії

Має 2 модифікації: ART1 – для бінарних значень, ART2 – для аналогових значень. Парадигма даної нейромережі полягає у порівнянні нового значення(даних) до уже відомих кластерів(класів), якщо вони не підходять до цього класу то створюється новий клас.

Метод нечіткої логіки:
Нечітка логіка використовується для сукупності даних, які з певною ймовірністю розміщені на векторі від 0 до 1 . Суть методу полягає у поділі набору даних на окремі групи за спільними властивостями.
Недоліки:
- відсутність стандартної методики для побудови нечітких систем
- математичні описи не відповідають класичним канонам математики.
Задачі які вирішує:
- нелінійний контроль за виробництвом
- задачі класифікації
- розпізнавання образів
- дослідження даних



Генетичні алгоритми:
Суть даних методів полягає у математичному описі, що описує популяцію. Тобто знаходимо булеві функції , що описують дослідуваний процес.
Недоліки:
- При реалізації тяжко обчислити оптимальність та адекватність отриманого рішення.
- Важко сформувати ефективний спосіб реалізації обчислень основних критерій.

Еволюційне програмування:
Даний спосіб є розвитком генетичних алгоритмів, з відмінністю в тому, що ми програмуємо цільову функцію, яка генерується за допомогою вектора коефіцієнтів, які знаходяться переважно експериментально.

Візуалізація даних:
Це сукупність методів, котрі допомагають користувачу наочно оцінити результати своєї роботи (блок-схеми, UML діаграми…).

Кожен з методів має свої переваги та недоліки, і на практиці доцільно використовувати їх комбінацію. У процесі розвитку інтелектуального аналізу даних генерують набір підходів.

3

5

IAD

4

2

1

6

8

7

1)Теорія баз даних
2)Алгоритми
3)Теорія статистики
4)Машинне обчислення
5)Візуалізація
6)Штучний інтелект
7)Розпізнавання образів
8)Комбіновані методи

Data Mining – суть даного поняття полягає, що для пошуку у великих базах даних неочевидних залежностей.
Задачі що реалізовуються:
- Класифікація
- Кластеризація
- Прогнозування
- Послідовність
- Оцінка відхилень
- Аналіз зв’язків
- Візуалізація

З метою оцінки отриманих результатів застосовують мякі обчислення. Дані обчислення ґрунтуються на основі використання теорії ймовірності та теорії графів.

P(HK\H)=(p(x\H)\p(H))/SUM(p(x\H)\*p(H))

Де х – вектор вхідних даних
Н – запропоновані гіпотези (можливі рішення)

За допомогою даної байєсовської теорії запропонували використовувати байєсівські мережі. Кожна мережа представлена за допомогою двійки, де G неорієнтований граф. В – властивості мережі. Під властивостями розуміють властивості, що знаходяться в таблицях інцидентності та суміжності.

P=<G,B>



Існує кілька видів мереж Баєса:
- Дискретні – мережі, які володіють такими властивостями: всі вершини мережі пов’язані, і описуються за допомогою матриці ймовірності. Якщо вершини не є зв’язані, то вважається, що ймовірність їх рівна 1.
- Динамічні – у яких стан вершин змінюється у часі. Найпростішим прикладом такої мережі є мережа Маркова. Основним застосуванням мережі Маркова є знаходження часу перехідного стану при роботі електро-техніки.
- Неперервні – вузли графів у даній мережі описуються за допомогою 1 та 0 (тобто вектором). Реалізовується за допомогою петель.


Дерева рішень

Дерева рішень використовуються для опису даних задач класифікації та регресії. Основними складовими дерева рішень є :
- об’єкт
- атрибут
- мітка класу
- вузол
- лист
- перевірка
Для побудови дерева рішень існує кілька етапів:
1) Етап розбиття – на даному етапі здійснюється розбиття навчальної вибірки на окремі групи (класи) за однаковими властивостями, або виконанням умов
2) Зупинка навчання – даний етап виникає, якщо при постановці задачі ми задаємо глибину руху по остову дерева (радіус).
3) Правило відсікання – даний етап (його властивості) враховує точність з якою має бути запропоноване рішення, і на ньому відсікається ті вузли, котрі є неінформативними для виконання поставленої задачі.
4) Створення правил – здійснюється побудова листа правил, який узгоджується із замовником або експертами у галузі.
Переваги дерев рішень:
- Швидкість побудови ланцюга міркувань
- На початковому етапі правила будуються на природній мові
- Висока точність прогнозу
Усі вузли у дереві рішень бувають 3 типів:
- Вузол рішень
- Ймовірнісні вузли
- Замикаючий вузол
Одним із ключових при побудові дерева рішень є запис у вузлах правил та функціональних залежностей на ребрах цільових функцій. При побудові цільових функцій дерева рішень використовується кілька методика:
1) Методика «Розділяй і владарюй» - суть даної методики полягає у розміщенні на початку дерева рішень змінних, які з кожним наступним рівнем вузлів класифікуються до тих пір, поки об’єкти не будуть належати одному класу.
Проблеми методики:
- Рання зупинка прийнятого рішення – це пов’язано із обмеженою кількістю правил для прийняття рішення.
- Відсікання гілок з верху до низу, що може привести до втрати правил або цільових функцій.
2) Методика на основі алгоритму ID3 – дана методика відрізняється від методики «Розділяй і владарюй» тим, що правила і цільові функції записуються за допомогою алгебри предикатів. Дана методика використовується тільки як класифікатор.
3) Методика на основі алгоритму CART – дана методика характеризується тим, що дерево рішень представляється у бінарному вигляді (0 і 1). Дана методика дозволяє здійснювати описи вхідних даних та будувати регресійні моделі. Даний підхід є простим у створенні кінцевого програмного продукту, та при прийнятті рішення дозволяє отримувати результат візуально.
4) Метод на основі алгоритму C45 – є модифікацією методики алгоритму ID3 з тою різницею, що вузли описуються не тільки за допомогою предикатів, а й за допомогою числових значень.
Недоліками дерева прийняття рішень є :
1) Велика кількість вхідних даних – є можливість великої деталізації – це зумовлює те, що у нас з’являються об’єкти, які з певною ймовірністю відносяться до 3 і більше класів, що знижує швидкість роботи дерева рішень.
2) Велика кількість вузлів зменшує наочність і зрозумілість побудованого дерева. Цей недолік вирішується утворенням кількох дерев рішень, які дають проміжний результат, і створення головного дерева.
Дерева рішень застосовуються у банківській системі, НГ промисловості, молекулярній біології.
Всі дерева прийняття рішень є основою для створення експертних систем, у яких застосовують найпростіші предикати (if - else).

**Метод асоціації та метод аналогії**

**Метод асоціації** – це сукупність методів, які ґрунтуються на представленні інформації про предметну область за допомогою функції приналежності. Існує 5 методів асоціації:

**Метод мозкового штурму**

Полягає у використанні психологічних та педагогічних закономірностей для математичного опису. Даний метод характеризується особливістю фізичного контакту між розробниками.

**Метод евристичних питань**

Суть даного методу полягає у формуванні ключових питань, які необхідно визначити. Даний метод характеризується такими принципами:
1) Проблемність і оптимальність
2) Дроблення інформації
3) Визначення основних цілей
Переваги даного методу:
- простота реалізації
- Ефективність використання

**Метод організованих стратегій**

В основі цього методу лежать:
- принцип відсторонення від досліджуваного об’єкту
- принцип автономності кожного елементу системи
Даний метод покладений в основу теорії ігор.

**Метод вільних асоціацій**

Суть даного методу полягає у вирішенні раніше не розглядуваних задач. Даний підхід покладений в основу усіх стартапів.

**Метод аналогій**

Сукупність даних методів полягають у встановленні зв’язків еквівалентності між двома різними системами, стан яких є різним у часі.

**Існує 4 особливості аналогій**:
- Пряма аналогія – коли ми вирішуємо проблему за подібним алгоритмом, раніше реалізованим.
- Суб’єктивна аналогія – даний принцип побудований на вживанні в образ іншого об’єкта. Тобто створення таких функцій еквівалентності між 2 системами, які будуть відтворювати таку модель, яка реалізовує проблему з точки зору іншого підходу.
- Символічна аналогія – суть даного принципу полягає у представленні всіх елементів системи за допомогою символів або візуального відображення.
- Фантастична аналогія - суть даного принципу полягає у отримання результату у най неймовірніших випадках.
**Існує 3 критерії методу аналогій**:
- Час – дана категорія використовується з метою вивчення поведінки системи на протязі певного періоду часу. Практична реалізація – часові ряди.
- Місце – це критерій, який характеризує навколишні збуджуючі елементи, які діють на нашу систему.
- Напрямок – напрямок буває 2 видів:
 - Диференціальний – характеризується розбиттям часу на окремі проміжки.
 - Інтегральний – сукупність часових проміжків.
**Перевагами методу аналогій є:**
- Наочність представлення результатів роботи у вигляді таблиця, графіків, структурних карт тощо.
**Недоліком методу аналогій є**:
- Точність отриманих результатів коливається у межах 60% - 80%

**Нечітка логіка**

Основною перевагою людського мозку при прийнятті рішень є видання результату за наявності неповної та нечіткої інформації. У зв’язку з цим в середині 50-х років 20 століття Заде запропонував математичний апарат який назвав FASY-логіка. Даний математичний апарат передбачав прийняття значення змінної від [0;1]. Тобто всі змінні у так званій терм-множині знаходяться у цьому проміжку.
Терм-множина – множина значень для x є А , яке набуває значень:

А=µ(х)/хі

Де µ(х) – це функція приналежності.
Функція приналежності – це характер зміни значень нечітких множин. **Характер зміни характеризується такими видами**:
- трикутна
- трапецевидна
- пилоподібна
- сигмоїдальна
Будь-яка тер-множина характеризується такими величинами:
- **Висота** – значення функції приналежності для відповідної випадкової змінної. Висота називається нормальною, якщо функція приналежності µ(х) = 1, в іншому випадку висота називається суб-нормальною. Якщо при записі тер-множини існує тільки 1 нечітка змінна, висота якої є 1, то така тер-множина називається унімоїдальною.
- **Точка переходу** – це таке значення нечіткої змінної, при якому перехідна функція має значення 0.5.

**Операції над нечіткими множинами**

**Об’єднання** – це така множина, коли значення функцій приналежності максимізуютьсяю.
**Перетин** – така величина, коли значення функції приналежності мінімізуються
**Різниця** – це множина, результати якої мінімізуються з використанням інвертування другої множини.
**Симетрична різниця**

АUВ = max(µa(х)/x; µb(х)/x)
А∩В = min(µa(х)/x; µb(х)/x)
A/B = min(µa(х); 1-µb(х))
A B = (A\B)U(B\A)=max(min(µa(х)/x; 1-µb(х));min(1- µa(х); µb(х)))

**<a,X,A>**
а – ім’я множини
Х – межі, в яких змінюється нечітка множина
А – нечітка множина – значення функції приналежності.

**<b,T,X,G,M>** - лінгвістична множина
b – найменування лінгвістичної змінної
T – множина значень нечітких змінних
G – синтаксична процедура, яка дозволяє здійснювати операції над нечіткими змінними.
Х – межі, в яких змінюється нечітка множина
М – семантична процедура, яка дозволяє перетворити нове значення лінгвістичної змінної у нечітку.

Всі нечіткі множини підпорядковуються тим ж законам, що й класичні множини.



1 – Блок фазифікації – на даному етапі множина лінгвістичних змінних перетворюється у нечіткі змінні.
2 – Блок прийняття рішень – сукупність правил та процедур (семантичних та лінгвістичних)
3 – Блок дефазифікації – перетворення прийнятого рішення у зрозумілий для людини формат
4 – Блок збереження необхідних констант оголошених лінгвістичних змінних, та правил, що задають обмеження на прийняті рішення.
Симуляцію для створення нечітких оболонок дозволяють такі середовища як: MathLab, Surfer.
**Переваги нечітких величин**:
- Можливість роботи із даними, які характеризуються з неповними та нечіткими даними.
- Можливість роботи з даними, область визначення яких є невизначеною.
- Можливість перетворювати висловлювання людини на зрозумілу для ЕОМ мову.
**Недоліки**:
- Складність при визначенні характеристики функції приналежності
- Кількість рекомендованих рішень на виході з блоку 3 може бути декілька

**Теорія Гри**

Перший математичний апарат для теорії гри був започаткований у 1944 році. Під грою розуміють формальний опис вирішення конфліктних ситуацій.
Всі ігри підлягають таким правилам:
 - Перелік можливих варіантів дій гравців
 - Об’єм інформації про дії одного суперника іншим
 - Послідовність чергування ходів
 - Результат гри
Метою теорії ігор є – вироблення рекомендацій по поведінці учасників конфлікту (визначення оптимальних стратегій та поведінок гравців)
Всі ігри поділяються на методи теорії статистичних рішень («ігри з природою») та методи теорії ігор.
До методів теорії ігор належать прямі та агреговані методи теорії прийняття рішень.
Всю теорію ігор можна класифікувати в залежності від таких параметрів:
 - В залежності від виду ходів (стратегічні і випадкові)
 - В залежності від кількості гравців (парні і множинні)
 - В залежності від взаємозв’язку ( безкоаліційні, коаліційні і кооперативні)
 - Коаліційні – коли кожен гравець коаліції отримує свій виграш
 - Кооперативний – суму виграшу кооперації ділять на кожного з гравців
- В залежності від кількості стратегій (скінчення і безкінечні)
- В залежності від кількості інформації (ігри з повною інформацією та ігри з
 неповною інформацією)
- В залежності від виду опису гри (позиційні та нормальної форми)
- В залежності від суми виграшу усіх гравців (з 0 сумою та з ненульовою сумою)
Стратегія гравця – це сукупність правил, які описують поведінку гравця підчас гри. Найбільше використовується так звані «матричні» ігри.
Матричні ігри характеризуються такими параметрами:
 - Нижня ціна гри – це максимальний з мінімальних виграшів.
 - Верхня ціна гри – це мінімальний з максимальних виграшів.
 - Сідлова точка – це параметр, який характеризує оптимальну стратегію
 поведінки двох гравців.
Нехай є така матриця:

|  |  |  |  |
| --- | --- | --- | --- |
|  | А1 | А2 | А3 |
| Б1 | 8 (верхня ціна) | 5 (нижня ціна) | 6 |
| Б2 | 8 | 6 (сідлова точка) | 7 (верхня ціна) |
| Б3 | 4 | 3 | 2 (нижня ціна) |

Де А1, А2, А3 – це перелік транспортних засобів необхідних для перевезення певної продукції, Б1,Б2,Б3 – продукція. Нижня ціна гри – найменший елемент в рядочку, верхня ціна гри найбільший елемент в стопці. Якщо мінімальний елемента максимальний збігаються, то це сідлова точка – вигідний сценарій для обох гравців.

Процес знаходження нового знання:
 - Опис поставленої задачі
 - Збір та підготовка даних:
 - Оцінка даних – це виявлення кількісних і якісних показників про поставлену
 задачу.
 - Об’єднання і очищення даних – виявлення дубльованих даних, та
 некоректно визначених.
 - Відбір даних – групування даних за ступенем важливості
 (основні та опосередковані)
 - Перетворення – приведення всіх даних до уніфікованого вигляду.
 - Побудова каркасу інтелектуальної системи видобування знань.
 Каркас – це вибір структури для видобування знань, яка підходить до
 поставленої задачі.
 - Перевірка побудованого каркасу на адекватність отриманих результатів
 - Тестування роботи побудованої системи
 - Використання

Сховища даних. Вітрини даних.
Системи VI:
 - Побудова сховищ даних
 - Обробка довільних запитів
 - Інструменти OLA і Data Mining
 - Засоби розповсюдження звітів
На даний момент найбільш розповсюдженими середовищами для створення VI систем є Oracle, Buisines Obgects і Microsoft. З метою очищення даних використовують консолідацію.
**Консолідація** – це сукупність методів і процедур призначених для отримання даних з різних джерел із забезпеченням необхідного рівня інформативності та перетворення в необхідний формат.
**Критеріями оптимальності при консолідації є**:
 - Швидкість доступу
 - Компактність зберігання
 - Автоматична підтримка цілісності структури даних
 - Контроль несуперечливості даних
**При консолідації вирішуються такі завдання**:
 - Вибір необхідних джерел
 - Вибір стратегії консолідації
 - Оцінка якості даних
 - Покращення інформативності даних
 - Перенесення у сховище даних
Оцінка якості даних виконує роль за допомогою аналітичних алгоритмів і методів засобу виявлення непотрібних даних.
**Збільшення інформативності даних** – це процес доповнення існуючої інформації, що дозволить використовувати отриману інформацію інформативніше.
**При створенні теоретичної основи для сховища даних виявили що необхідно мати такі додаткові інструменти**:
 - Сукупність джерел інформації (БД, облікові системи, звичайні документи)
 - Математичні алгоритми та методи для оцінки якості даних і підвищення
 інформативності
 - Інструментальні та програмні засоби для отримання трансформації та
 завантаження до сховища даних
Перше визначення сховища даних запропонував Інмон в кінці 90-х.
**Сховище даних** – це різновид зберігання даних, орієнтованих на підтримку процесу аналізу даних, та прийняття рішень, що забезпечує цілісність, несуперечливість і хронологію даних.

**Ключові параметри сховища даних (за Інмоном):**
 - Предметна орієнтованість
 - Інтегрованість
 - Незмінність
 - Хронологія
Під **предметною орієнтованістю** він розумів, що будь-яке сховище даних повинно розроблятися під конкретну задачу.
**Інтегрованість** – це можливість отримання даних різного формату та з різних джерел.
**Незмінність** – передбачає при завантаженні даних у сховище даних незмінюваність на протязі всього часу використання бази даних.
**Підтримка хронології** – при створенні запису у сховищі даних добавляється обов’язкове поле – «дата і час».
**При побудові сховища даних розрізняють**:
 - **Деталізовані** – це дані про практичну діяльність суспільства, людини, техніки.
 - **Агреговані** – це дані які оцінюють якість даних, взаємозв’язки між даними та їх
 важливість.
 - **Методальні** – це будь-яка інформація, необхідна для аналізу, проектування,
 побудови та впровадження сховища даних. Їх поділяють на 2 рівні:
 - **Технічні методальні** – це дані, необхідні для забезпечення функціонування
 сховища даних (логи маніпуляцій з сховищем даних, параметри моделі
 даних)
 - **Бізнес-методальні** – це опис специфіки аналізу сховища даних
 (бізнес-терміни та означення, якими буде керуватися користувач)
**Для того, щоб сховище даних існувало ефективно необхідно**:
 - Висока швидкість отриманих даних
 - Інструменти для online аналізу даних
 - Можливість отримання і порівняння зрізів даних
 - Щоб дані були не суперечливі
**Існує кілька видів сховищ даних за архітектурою**:
 - **Багатомірні сховища даних, які використовують для обробки MOLAP.**
 **Побудовані на концепції побудови багатомірних кубів та гіперкубів. В межах
 багатомірного сховища всі дані розділені на 2 групи**:
 - **Виміри** – найменування властивостей об’єктів.
 - **Факти** – це дані, що кількісно характеризують досліджуваний процес
**Перевагами БСД є**:
 - Наочність
 - Ширші можливості побудови аналітичних запитів
 - Формування складних запитів без використання SQL з боку користувача
 - **Реляційні сховища даних, які використовують для обробки ROLAP**.
 **Побудовані на концепції реляційних БД, і має за архітектурою 2 види**:
 - **Зірка**
 Переваги:
 - Простота, логічність, моделі
 - Простота наповнення даних вибірки

 Недоліки:
 Повільна обробка вибірок
 Висока ймовірність виникнення невідповідності даних
 - **Сніжинка**.
 Переваги:
 - Біль ефективне завантаження даних у багатомірні структури
 - Низька ймовірність появи помилок невідповідності даних
 Недоліки:
 Складна для розуміння архітектура сховища даних
 Ускладнена процедура додавання даних у сховище даних.
 **Переваги реляційних сховищ даних**:
 - Майже нескінченна к-сть даних
 - Легка адаптація до вже існуючих систем зберігання даних
 - Забезпечення високого рівню захисту та ієрархії користувачів.
 - Немає необхідності змінювати архітектуру СД при додаванні нових
 вимірів
 - **Гібридні сховища даних, які використовують для обробки HOLAP** Використовують реляційний підхід для завантаження даних, і багатомірний
 підхід для їх обробки.
 Переваги:
 - Реляційність дозволяє формувати стійкі і несуперечливі опорні точки для
 багатомірного сховища.
 Недоліки:
 - Ускладнене адміністрування сховища даних.
В останній час з поширенням інтернету виник ще один тип сховища даних – віртуальний.
**Віртуальне сховище даних** – це емуляція сховища даних, які приймають, обробляють і інтегрують дані у процесі виконання даних. Для цього використовують розрізнені та віддалені джерела інформації.
Переваги ВСД:
 - Мінімізується необхідність використання великих об’ємів дискової пам’яті
 - Аналітик абстрагується від процесу – «де і як взяти дані»
 - Можливість отримати дані аналізу без їх завантаження у сховище даних
Недоліки ВСД:
 - Джерела інформації не завжди можуть бути доступні
 - При використанні HOLAP систем різних джерел даних можуть бути отримані
 різні результати
 - Низька швидкість доступу до джерел даних
**Вітрина даних** – це спеціалізоване, тематичне сховище даних, яке підключене до загального сховища даних. Використовується з метою збільшення виконання запитів.
Переваги:
 - Безпосередній контакт з користувачем
 - Дані тематично орієнтовані на конкретного користувача
 - Малі затрати при організації і підтримці
 - Хороші можливості розмежування прав користувачів