Лабораторна робота №2

ДЕРЕВА ВИРІШАЛЬНИХ ПРАВИЛ

Мета роботи: вивчити основні методи побудови дерев вирішальних правил, навчитися використовувати спеціалізовані програмні засоби для побудови дерев вирішальних правил.

### Завдання до роботи

1. Ознайомитися з конспектом лекцій та рекомендованою літературою, а також додатком Б, що містить опис програмного забезпечення для побудови дерев вирішальних правил.
2. Сформувати набір даних для обробки та аналізу.
3. Використовуючи рекомендоване програмне забезпечення здійснити обробку набору даних з метою побудови дерева вирішальних правил.
4. Використати побудоване дерево для прийняття рішень на конкретному прикладі.
5. Оформити звіт з роботи.
6. Відповісти на контрольні питання.

### Зміст звіту

1. Тема та мета роботи.
2. Короткі теоретичні відомості.
3. Набір даних для обробки (якщо він великий – навести фрагмент).
4. Лістинг основних функцій програми з коментарями.
5. Результати роботи програмного забезпечення (дерево вирішальних правил, результати прийняття рішень, інші характеристики).
6. Висновки, що містять відповіді на контрольні запитання, а також відображують результати виконання роботи та їх критичний аналіз.

### Контрольні питання

1. Що таке дерево вирішальних правил? Який спосіб подання правил в них використовується?
2. Дати означення основних понять, що відносяться до теорії дерев вирішальних правил: об’єкт, атрибут, мітка класу, вузол, лист, перевірка.
3. Навести основні класи задач, до яких можуть бути застосовані дерева вирішальних правил.
4. Яким чином відбувається побудова дерева рішень? Який метод використовується для цього?
5. В чому полягає процес навчання з учителем?
6. Порівняйте методи, що реалізують дерева вирішальних правил: CART та C4.5.
7. Поясніть принцип роботи “жадібних” алгоритмів.
8. Перелічіть основні аспекти, яким приділяється увага при побудові дерев рішень.
9. В чому полягає правило відбору ознаки для розбиття? Сформулюйте загальне правило для відбору атрибуту.
10. Виконайте порівняльний аналіз критеріїв оцінки якості розбиття множини на класи.
11. Що визначає правило зупину? Дайте порівняльну характеристику відомих критеріїв зупину побудови дерева вирішальних правил?
12. Для чого використовується правило відсіку?
13. Що розуміють під точністю та помилкою розпізнавання для дерева вирішальних правил?
14. Що необхідно зробити для добування правил з дерева вирішальних правил?
15. Які вимоги до структури та значень даних висуває метод С4.5?
16. Проаналізуйте алгоритм побудови дерева вирішальних правил за допомогою методу С4.5.
17. Яким чином визначається критерій вибору атрибуту в методі С4.5?
18. В якому випадку в процесі роботи методу С4.5 вузол помічається як лист? Що обирається в якості розв’язку листа?
19. Коли ентропія досягає свого максимуму (мінімуму) при використанні методу С4.5?
20. В яких випадках необхідно обрати поріг для порівняння значень атрибуту?
21. В чому полягає класифікація нових об’єктів? Звідки починається обхід дерева?
22. Порівняйте покращений критерій розбиття з класичним.
23. Яке евристичне правило використовується для зменшення ймовірності створення вузлів та листя, які містять незначну кількість об’єктів?
24. Проаналізуйте процедуру роботи з пропущеними даними.
25. Яким чином відбувається класифікація нових об’єктів у випадку відсутності значення певного атрибуту об’єкту, що класифікується?
26. Які переваги використання дерев вирішальних правил?
27. Який напрямок побудови дерева вирішальних правил використовується при використанні методу ID3?
28. Наведіть послідовність побудови дерева вирішальних правил за допомогою методу ID3.
29. В чому полягає вибір властивості на основі теорії інформації?
30. Проаналізуйте основні функції пакету Matlab для роботи з деревами вирішальних правил: внутрішня структура, параметри, основні змінні, методи, їх призначення та використання.

**Додаток Б**

**ДЕРЕВА ВИРІШАЛЬНИХ ПРАВИЛ У ПАКЕТІ MATLAB**

Дерева вирішальних правил у пакеті Matlab реалізовані у модулі Statistics Toolbox. Основними функціями для роботи з деревами є: treefit, treeprune, treedisp, treetest, treeval.

Таблиця Б.1 – Функції пакету Matlab для роботи із деревами вирішальних правил

|  |  |  |
| --- | --- | --- |
| № | Формат виклику | Призначення функції |
| 1 | t = treefit(X,y) | створює дерево вирішальних правил t на основі заданих значень незалежних змінних (матриця Х) та значень вихідного параметру (вектор у) |
| 2 | t1 = treeprune(t,'level',n) | створює дерево вирішальних правил t1 на основі заданого дерева t, скороченого до n-го рівня |
| t1 = treeprune(t,'nodes',nod) | створює дерево вирішальних правил t1 на основі заданого дерева t, видаляючи при цьому вузли, вказані в змінній nod |
| 3 | treedisp(t) | відображує дерево t у графічному вигляді |
| 4 | с = treetest(t,'test',X,y) | виконує тестування дерева t за допомогою тестової вибірки (X, y). |
| 5 | Ycalc = treeval(t,X) | за допомогою дерева t для масиву незалежних змінних Х розраховує значення вихідного параметру Ycalc |

Нижче наведено приклад створення дерева вирішальних правил на основі масивів даних meas та species, що зберігаються в структурі fisheriris.

load fisheriris; % завантажити з файлу fisheriris.mat змінні meas та species для створення дерева вирішальних правил

t = treefit(meas,species); % створити дерево на основі змінних meas та species

treedisp(t,'names',{'SL' 'SW' 'PL' 'PW'}); % вивести на екран побудоване дерево t у графічному вигляді. При цьому для підпису значень незалежних змінних, на основі яких було побудоване дерево t, використовується параметр names

[c,s,n,best] = treetest(t,'cross',meas,species); % протестувати дерево t

tmin = treeprune(t,'level',best); % мінімізувати дерево t

%розрахунок точності класифікації побудованого дерева

sfit = treeval(t,meas); % за допомогою дерева t отримати відповідні числові значення класів для екземплярів з масиву meas

sfit = t.classname(sfit); % отримати відповідні назви класів

mean(strcmp(sfit,species)) % розрахунок точності класифікації