

Лабораторна робота № 6

Знайомство з методами побудови асоціативних правил

Мета роботи: Ознайомитися та набути навички побудови асоціативних правил за допомогою WEKA.

Завдання: Побудувати асоціативні правила для тестових екземплярів даних за допомогою зазначеного методу.

Загальні відомості

Асоціативне правило — закономірність появи заданих значень атрибутів у вигляді комбінації. Методи побудови асоціативних правил призначені для пошуку комбінацій значень атрибутів екземплярів даних, на базі яких за значеннями атрибутів першої множини можна спрогнозувати значення атрибутів другої множини.

На відміну від методів побудови моделей класифікації, методи пошуку асоціативних правил не потребують задання окремого атрибуту в якості класу, класом вважається унікальна комбінація значень атрибутів. Пошук асоціативних правил часто виконують супермаркети під час аналізу споживчого кошику для визначення продуктів, які покупці часто купують разом, після чого знайдені комбінації товарів намагаються розміщувати поруч, щоб збільшити ймовірність їх придбання.

Для ілюстрації ідеї пошуку асоціативних правил виконаємо аналіз тестового файлу `contact-lenses.arff` (розміщується у директорії `data` каталогу WEKA) з записами особливостей підбору контактних лінз за фізіологічними характеристиками пацієнтів. Для пошуку виберемо метод Apriori.

Метод Apriori — класичний метод пошуку шаблонів значень атрибутів та асоціативних правил. Метод Apriori перебирає усі можливі комбінації значень заданої кількості атрибутів, наприклад, *комбінація* $l = \{\text{атрибут } A \text{ значення } A1,$

атрибут Б значення Б1}, та вибирає ті комбінації, які перевищують мінімальне значення критерію підтримки (support criterion, задається у параметрах методу).

Наприклад, у файлі contact-lenses.arff міститься 24 екземпляри даних, отже можна послідовно шукати шаблони, які містять від 1 до 24 атрибутів. Для кожної комбінації задається мінімальне значення критерію підтримки, нехай буде 20% від загальної вибірки з 24-х екземплярів, тоді мінімальна допустима кількість екземплярів, у яких буде знайдено конкретний шаблон має перевищувати чи дорівнювати $24 \times 0.1 = 4.8 \geq 5$ екземплярів.

Виконаємо у пакеті WEKA пошук асоціативних правил за допомогою методу Apriori з заданими параметрами (див. рисунок 1). Для відображення шаблонів вкажіть значення параметру outputItemSets=True.

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm. More Capabilities

car: False

classIndex: -1

delta: 0.05

doNotCheckCapabilities: False

lowerBoundMinSupport: 0.2

metricType: Confidence

minMetric: 0.9

numRules: 10

outputItemSets: True

removeAllMissingCols: False

significanceLevel: -1.0

treatZeroAsMissing: False

upperBoundMinSupport: 1.0

verbose: False

Open... Save... OK Cancel

Рисунок 1. Тестові значення параметрів методу Apriori у WEKA

Метод Apriori починає пошук з шаблонів, які містять лише одне значення атрибуту. Далі, в процесі пошуку, виходячи з того факту, що менші комбінації атрибутів зустрічаються частіше, кількість атрибутів у шуканих шаблонах

поступово збільшується. Пошук зупиняється за умови відсутності знайдених екземплярів для поточної кількості атрибутів.

Результат роботи пошуку для різних кількостей атрибутів зображено на рисунку 2.

Apriori

=====

Minimum support: 0.2 (5 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Large Itemsets L(1):

age=young 8

age=pre-presbyopic 8

age=presbyopic 8

spectacle-prescrip=myope 12

spectacle-prescrip=hypermetrope 12

astigmatism=no 12

astigmatism=yes 12

tear-prod-rate=reduced 12

tear-prod-rate=normal 12

contact-lenses=soft 5

contact-lenses=none 15

Size of set of large itemsets L(2): 21

Знайдено 21 шаблон

Large Itemsets L(2):

age=pre-presbyopic contact-lenses=none 5

знайдено у 5 екземплярів з 24-х можливих

age=presbyopic contact-lenses=none 6

spectacle-prescrip=myope astigmatism=no 6

spectacle-prescrip=myope astigmatism=yes 6

spectacle-prescrip=myope tear-prod-rate=reduced 6

spectacle-prescrip=myope tear-prod-rate=normal 6

spectacle-prescrip=myope contact-lenses=none 7

spectacle-prescrip=hypermetrope astigmatism=no 6

spectacle-prescrip=hypermetrope astigmatism=yes 6

spectacle-prescrip=hypermetrope tear-prod-rate=reduced 6

spectacle-prescrip=hypermetrope tear-prod-rate=normal 6

spectacle-prescrip=hypermetrope contact-lenses=none 8

astigmatism=no tear-prod-rate=reduced 6

astigmatism=no tear-prod-rate=normal 6

astigmatism=no contact-lenses=soft 5

astigmatism=no contact-lenses=none 7

astigmatism=yes tear-prod-rate=reduced 6

```

astigmatism=yes tear-prod-rate=normal 6
astigmatism=yes contact-lenses=none 8
tear-prod-rate=reduced contact-lenses=none 12
tear-prod-rate=normal contact-lenses=soft 5

Size of set of large itemsets L(3): 6

Large Itemsets L(3):
spectacle-prescrip=myope    tear-prod-rate=reduced    contact-
lenses=none 6
spectacle-prescrip=hypermetrope    astigmatism=yes    contact-
lenses=none 5
spectacle-prescrip=hypermetrope    tear-prod-rate=reduced
contact-lenses=none 6
astigmatism=no tear-prod-rate=reduced contact-lenses=none 6
astigmatism=no tear-prod-rate=normal contact-lenses=soft 5
astigmatism=yes tear-prod-rate=reduced contact-lenses=none 6

```

Рисунок 2. Шаблони для різних кількостей атрибутів

Асоціативне правило

Асоціативне правило складається з двох множин X (“причини”) і Y (“наслідку”) у вигляді конструкції if-then: $X \rightarrow Y$, тобто, якщо знайдені значення атрибутів множини X , тоді з ним ймовірно будуть знайдені значення атрибутів множини Y .

Критерій підтримки (support criterion) — частка кількості екземплярів, яка припадає на кожне асоціативне правило (від 0.0 до 1.0). Для 24 екземплярів підтримка у 20% (minSupport=0.2) дорівнює 5, тобто кожне асоціативне правило має виконуватися щонайменше для 5 екземплярів з усього набору даних.

Критерій достовірності (confidence criterion) — відношення підтримки атрибутів множини $(X \cup Y)$ до підтримки атрибутів множини X , що характеризує залежність між значеннями атрибутів множин X та Y .

Розробка програми з пошуку асоціативних правил

На рисунку 3 зображено код програми пошуку асоціативних правил з застосуванням методу Apriori.

```
import java.io.BufferedReader;
import java.io.FileReader;

import weka.core.Instances;
import weka.associations.Apriori;

public class AssociationRules {

    public static void main(String args[]) throws Exception {

        //load data
        Instances data = new Instances(new BufferedReader(new FileReader("contact-lenses.arff")));

        //build model
        Apriori model = new Apriori();
        model.buildAssociations(data);
        System.out.println(model);

    }
}
```

Рисунок 3. Програма пошуку асоціативних правил

Результат роботи програми зображено на рисунку 4

```
Apriori
=====

Minimum support: 0.2 (5 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 21

Size of set of large itemsets L(3): 6

Best rules found:

1. tear-prod-rate=reduced 12 ==> contact-lenses=none 12 <conf:(1)> lift:(1.6) lev:(0.19) [4] conv:(4.5)
```

```

2. spectacle-prescrip=myope tear-prod-rate=reduced 6 ==> contact-lenses=none 6 <conf:(1)> lift:(1.6) lev:(0.09) [2]
conv:(2.25)
3. spectacle-prescrip=hypermetrope tear-prod-rate=reduced 6 ==> contact-lenses=none 6 <conf:(1)> lift:(1.6)
lev:(0.09) [2] conv:(2.25)
4. astigmatism=no tear-prod-rate=reduced 6 ==> contact-lenses=none 6 <conf:(1)> lift:(1.6) lev:(0.09) [2] conv:(2.25)
5. astigmatism=yes tear-prod-rate=reduced 6 ==> contact-lenses=none 6 <conf:(1)> lift:(1.6) lev:(0.09) [2]
conv:(2.25)
6. contact-lenses=soft 5 ==> astigmatism=no 5 <conf:(1)> lift:(2) lev:(0.1) [2] conv:(2.5)
7. contact-lenses=soft 5 ==> tear-prod-rate=normal 5 <conf:(1)> lift:(2) lev:(0.1) [2] conv:(2.5)
8. tear-prod-rate=normal contact-lenses=soft 5 ==> astigmatism=no 5 <conf:(1)> lift:(2) lev:(0.1) [2] conv:(2.5)
9. astigmatism=no contact-lenses=soft 5 ==> tear-prod-rate=normal 5 <conf:(1)> lift:(2) lev:(0.1) [2] conv:(2.5)
10. contact-lenses=soft 5 ==> astigmatism=no tear-prod-rate=normal 5 <conf:(1)> lift:(4) lev:(0.16) [3] conv:(3.75)

```

Рисунок 4. Результат роботи програми

На рисунку 4 зображено 10 асоціативних правил ($\text{numRules}=10$), виокремлених з 38 шаблонів за критерієм $\text{minMetric}=0.9$ ($\text{metricType}=\text{Confidence}$). Метод Аргіорі поступово зменшує мінімальну підтримку (minSupport) з 1.0 до заданих 0.2 (1.0, 0.9, 0.8, 0.7, ..., 0.2), поки не буде знайдена необхідна кількість правил ($\text{numRules}=10$) з заданим мінімальним значенням критерію достовірності ($\text{minMetric}=0.9$). Якщо 10 правил будуть знайдені з $\text{minSupport}=0.3$, тоді метод зупиниться та залишить шаблони, значення атрибутів яких містяться щонайменше у $24 \times 0.3 = 7.2 \geq 7$ екземплярів.

Розглянемо правило № 5 зображене на рисунку 5.

```

5.      astigmatism=yes      tear-prod-rate=reduced      6      ==>      contact-lenses=none      6      <conf:(1)>
lift:(1.6) lev:(0.09) [2] conv:(2.25)

```

Рисунок 5. Приклад знайденого асоціативного правила

Значення атрибутів множини X (“причини”) (astigmatism=yes tear-prod-rate=reduced) були знайдені у 6 екземплярів так само, як і значення атрибутів множини Y (“наслідку”) (contact-lenses=none), відповідно достовірність асоціативного правила дорівнює $\text{conf} = 6/6 = 1.0$, тобто виконується умова $\text{minMetric}=0.9$.

Розробіть програму та побудуйте перелік асоціативних правил за допомогою метода, зазначеного у варіанті.

Таблиця 1

Варіанти завдання

1	Apriori	11	Generalized Sequential	21	Apriori
2	FPGrowth	12	FPGrowth	22	PredictiveApriori
3	Generalized Sequential	13	Apriori	23	Tertius
4	PredictiveApriori	14	PredictiveApriori	24	FPGrowth
5	Tertius	15	Tertius	25	Generalized Sequential
6	Generalized Sequential	16	Apriori	26	PredictiveApriori
7	FPGrowth	17	FPGrowth	27	Generalized Sequential
8	Apriori	18	Generalized Sequential	28	FPGrowth
9	PredictiveApriori	19	PredictiveApriori	29	Apriori
10	Tertius	20	Tertius	30	Tertius

Звіт по роботі

1. Титульний аркуш.
2. Мета роботи.
3. Дані (відмінні від прикладу).
4. Код програми (параметри відмінні від прикладу).
5. Перелік знайдених асоціативних правил ($\text{minMetric} \geq 0.9$).
6. Висновки.