

Лабораторна робота 5. Побудова асоціативних правил за алгоритмами Apriori та FP-росту

Недашківська Н.І.

Звіт має містити:

- результати по кожному пункту ходу виконання роботи,
- опис методу побудови асоціативних правил, який використовувався,
- у текстовому вигляді пояснення отриманих цифр (!).

Захист роботи:

- Демонстрація програми, яка реалізує завдання згідно з варіантом. Відповіді на питання по коду програми та отриманим результатам.
- Письмово теоретичне питання, задача та/ або виконання іншого варіанту завдання на комп'ютері.

1 Хід виконання роботи:

1. Взяти файл з даними у відповідності з варіантом.
2. Побудувати часті набори та асоціативні правила (АП), використовуючи алгоритм Apriori. Дослідити множини АП залежно від параметрів побудови АП:
 - мінімальної підтримки,
 - мінімальної достовірності,
 - максимальної потужності частих наборів.

Підібрати значення параметрів алгоритму Apriori, які призводять до значущих АП. АП R_k вважати значущим, якщо покращення (improvement або ліфт (lift)): $Improv(R_k) > 1$.

3. Побудувати часті набори та множину АП, використовуючи алгоритм FP-росту. Дослідити множини АП залежно від параметрів побудови АП, які було наведено вище для Apriori.

Підібрати значення параметрів алгоритму FP-росту, які призводять до значущих АП.

4. Зробити висновки щодо впливу параметрів алгоритмів Apriori та FP-росту на знайдені множини АП.
5. Знайти значення прогнозу на основі побудованої множини правил.
6. Порівняти результати, отримані алгоритмами Apriori та FP-росту.

Дозволяється використати наступні матеріали:

http://rasbt.github.io/mlxtend/api_subpackages/mlxtend.frequent_patterns/

<https://www.kaggle.com/povilass/analysis-of-transactions-from-a-bakery-and-apriori>

2 Варіанти:

1. Var_1_BreadBasket_DMS
<https://www.kaggle.com/sulmansarwar/transactions-from-a-bakery>
2. Var_2_groceries - groceries
The dataset contains 9835 transactions by customers shopping for groceries.
The data contains 169 unique items.
3. Var_3_groceries
4. Var_4_Market_Basket_Analysis_1
5. Var_5_BlackFriday
6. Var_6_1000_Transactions
7. Var_7_500_Transactions
8. Var_8_Market_Basket_Analysis_2
9. Var_1_BreadBasket_DMS
10. Var_2_groceries - groceries
11. Var_3_groceries
12. Var_4_Market_Basket_Analysis_1
13. Var_5_BlackFriday
14. Var_6_1000_Transactions
15. Var_7_500_Transactions
16. Var_8_Market_Basket_Analysis_2
17. Var_1_BreadBasket_DMS

18. Var_2_groceries - groceries
19. Var_3_groceries
20. Var_4_Market_Basket_Analysis_1
21. Var_5_BlackFriday
22. Var_6_1000_Transactions
23. Var_7_500_Transactions
24. Var_8_Market_Basket_Analysis_2
25. Var_1_BreadBasket_DMS
26. Var_2_groceries - groceries
27. Var_3_groceries
28. Var_4_Market_Basket_Analysis_1
29. Var_5_BlackFriday
30. Var_6_1000_Transactions

3 Контрольні питання до захисту роботи:

1. Постановка задачі аналізу ринкових кошиків. Поняття асоціативного правила. Показники корисності асоціативних правил.
2. Загальний метод пошуку асоціативних правил. Властивість антимонотонності
3. Знаходження частих наборів алгоритмом Apriori.
4. Побудова асоціативних правил: етап 2, коли множина частих наборів відома.
5. Модифікації алгоритму Apriori. Чим спричинена необхідність у цих модифікаціях?
6. Знаходження частих наборів алгоритмом Eclat. Переваги і недоліки Eclat.
7. Поняття FP-дерева. Ідея алгоритму FP-росту. Переваги і недоліки алгоритму.
8. Алгоритм FP-росту: етап 1 - побудова FP-дерева.
9. Означення умовного FP-дерева. Навести приклад. Алгоритм побудови умовного FP-дерева.

10. Алгоритм FP-росту: етап 2 - пошук частих наборів в FP-дереві. Проілюструвати на прикладі декілька перших ітерацій алгоритму.