

Задание I «Поиск ассоциативных правил»

1. Загрузите файл «assc_TRANSACTION.csv». В нем три колонки Customer – ID клиента, Product – покупка, Time – временная метка (для задания не нужна). Определите (написав соответствующий код) сколько различных значений принимают переменные Product и Customer.
2. Найдите частые эпизоды с ограничением на размер правила равным 4, с использованием алгоритма и порога на поддержку согласно вашему варианту.
3. Найдите самый большой (где больше всего элементов) частый эпизод, содержащий продукт согласно вашему варианту. Какая у него поддержка?
4. На основе найденных частых эпизодов постройте ассоциативные правила с порогом на достоверность согласно вашему варианту. Найдите правило с максимальным лифтом, содержащем продукт из вашего варианта в левой части правила. Дайте ему письменную словесную интерпретацию, укажите и объясните его числовые показатели: поддержку, достоверность и подъем.
5. Используя только двухместные правила постройте ориентированный граф, где вершины элементы правила, их цвет (или размер) – поддержка элемента (item support), дуги – импликации (ориентированы в направлении от условия к следствию), веса дуги – достоверности.
6. Для данного графа рассчитайте меры центральности согласно вашему варианту и найдите элемент с самой высокой мерой, а также какую меру имеет продукт из вашего варианта.
7. Постройте числовую матрицу со счетчиком числа покупок в ячейках, клиентами по строкам и продуктами по столбцам.

Задание 2 «Выявление скрытых структур в данных»

8. С помощью метода из вашего варианта постройте линейную проекцию набора данных на плоскость (2 компоненты) цветом укажите транзакции, содержащие продукт вашего варианта.
9. С помощью метода из вашего варианта постройте нелинейную проекцию набора данных на плоскость цветом укажите транзакции, содержащие продукт из вашего варианта. Не указанные в задании параметры (например, размер решетки для SOM или число слоев в автоэнкодере можно выбирать на свое усмотрение для получения наиболее удобной визуализации). Дайте письменный комментарий, чем с вашей точки зрения для вашего примера лучше или хуже нелинейная проекция).
10. Из исходной матрицы (из пункта 7) согласно вашему варианту отберите указанное число независимых переменных с использованием заданного метода.

Запишите и перешлите для проверки JN реализующий шаги 1-10.

| ВАРИАНТ | ПУНКТ 2 | ПУНКТ 3 | ПУНКТ 4 | ПУНКТ 6 | ПУНКТ 8 | ПУНКТ 9 | ПУНКТ 10 |
|---------|-------------|----------|---------|-------------|---------|-------------|------------|
| 1 | Apriori, 2% | artichok | 50% | Betweenness | PCA | tSNE | VarClus, 3 |
| 2 | FPTree, 2% | baguette | 40% | Closeness | NMF | SOM | VarClus, 3 |
| 3 | Apriori, 3% | bourbon | 30% | Clust. coef | PCA | AutoEncoder | VarClus, 4 |
| 4 | FPTree, 3% | coke | 20% | Authority | NMF | tSNE | VarClus, 4 |
| 5 | Apriori, 4% | cracker | 10% | Hub | PCA | SOM | VarClus, 5 |
| 6 | FPTree, 4% | heineken | 50% | Betweenness | NMF | AutoEncoder | VarClus, 5 |
| 7 | Apriori, 5% | ice_crea | 40% | Closeness | PCA | tSNE | VarClus, 6 |
| 8 | FPTree, 5% | peppers | 30% | Clust. coef | NMF | SOM | VarClus, 6 |
| 9 | Apriori, 6% | soda | 20% | Authority | PCA | AutoEncoder | VarClus, 7 |
| 0 | FPTree, 6% | turkey | 10% | Hub | NMF | tSNE | VarClus, 7 |