Поиск частых наборов

Выполнила: Бадамшина А.Ф. КЭ-403

1. **Формулировка задания**

Разработайте программу, которая выполняет поиск частых наборов объектов в заданном наборе данных с помощью алгоритма Apriori (или одной из его модификаций). Список результирующих наборов должен содержать как наборы, так и значение поддержки для каждого набора. Параметрами программы являются набор, порог поддержки и способ упорядочивания результирующего списка наборов (по убыванию значения поддержки или лексикографическое).

Проведите эксперименты на наборе данных baskets.csv (сведения о покупках в супермаркете). В экспериментах варьируйте пороговое значение поддержки (например: 1%, 3%, 5%, 10%, 15%).

Выполните визуализацию результатов экспериментов в виде следующих диаграмм:

1. сравнение быстродействия на фиксированном наборе данных при изменяемом пороге поддержки;
2. количество частых наборов объектов различной длины на фиксированном наборе данных при изменяемом пороге поддержки.

Подготовьте отчет о выполнении задания и загрузите отчет в формате PDF в систему. Отчет должен представлять собой связный и структурированный документ со следующими разделами:

1. формулировка задания;
2. гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами;
3. рисунки с результатами визуализации;
4. пояснения, раскрывающие смысл полученных результатов.
5. **Гиперссылка на репозиторий**

Ссылка на репозиторий: https://github.com/AnjelinaBadamshina/Analytical-information-processing-technologies.git.

1. **Результаты визуализации**

На данном графике показано, как изменяется время выполнения алгоритма Apriori в зависимости от установленного порога поддержки. Видно, что при снижении порога поддержки вычисления занимают больше времени, так как алгоритму требуется обрабатывать больше возможных комбинаций.

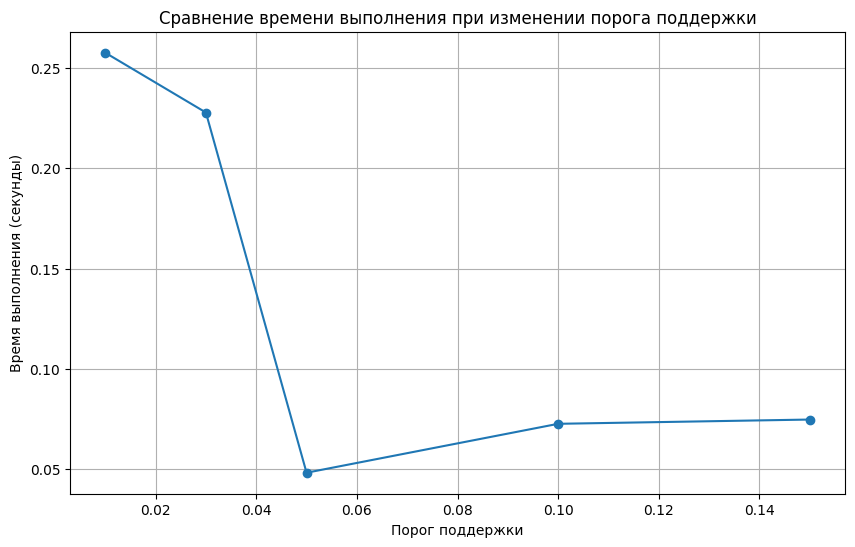
****

Рисунок 1 – Сравнение времени выполнения при изменении порога поддержки

Второй график показывает количество частых наборов различных длин при разных значениях поддержки. Чем ниже порог поддержки, тем большее количество частых наборов обнаруживается. Однако при высоких значениях поддержки остаются только самые популярные сочетания товаров.

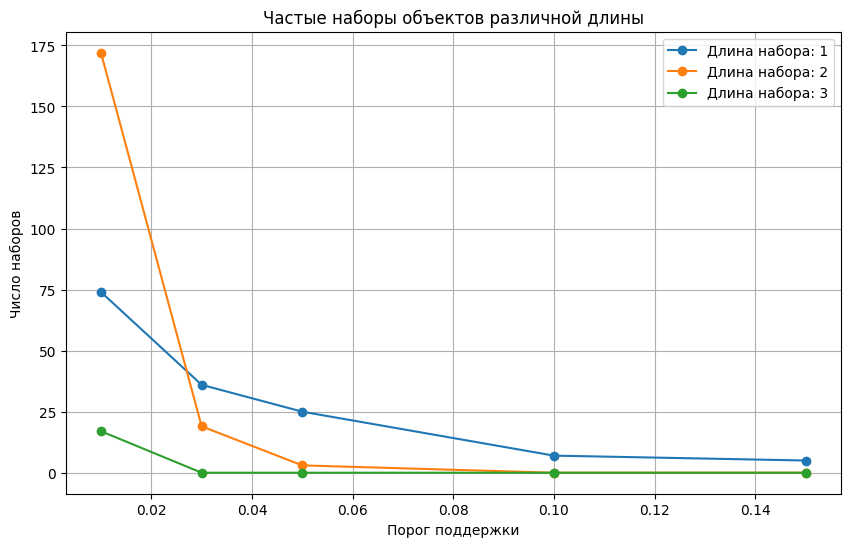


Рисунок 2 – Частые наборы объектов различной длины

1. **Анализ и пояснение результатов**

Влияние порога поддержки на производительность:

1. Чем ниже порог поддержки, тем больше комбинаций анализируется, что увеличивает время выполнения алгоритма.
2. При высоких значениях поддержки анализируется меньше комбинаций, что снижает нагрузку на вычисления.

Распределение частых наборов:

1. Большинство частых наборов состоит из одного или двух элементов.
2. При уменьшении порога поддержки появляются более длинные наборы, но их число сокращается из-за уменьшения встречаемости более сложных комбинаций.

Практическое значение анализа:

1. Метод Apriori используется для выявления закономерностей в данных о покупках, что помогает бизнесу в маркетинговых стратегиях.
2. Анализ позволяет оптимизировать выкладку товаров и разработку рекомендаций для покупателей.