

## Поиск ассоциативных правил

Кузьмичева Ольга, КЭ-403

## Задание

1. Доработайте программу из задания Поиск частых наборов, чтобы она также выполняла поиск ассоциативных правил. Список результирующих правил должен выдаваться в удобочитаемом виде (антецедент→консеквент) с указанием поддержки и достоверности каждого правила. Дополнительные параметры программы: порог достоверности, способ упорядочивания результирующего списка наборов (по убыванию значения поддержки или лексикографическое).

2. Проведите эксперименты на наборах из задания 1. В экспериментах зафиксируйте значение пороговое значение поддержки (например, 10%), варьируйте пороговое значение достоверности (например, от 70% до 95% с шагом 5%).

3. Выполните визуализацию полученных результатов в виде следующих диаграмм:

- сравнение быстродействия поиска правил на фиксированном наборе данных при изменяемом пороге достоверности;
- общее количество найденных правил на фиксированном наборе данных при изменяемом пороге достоверности.

4. Подготовьте список правил, в которых антецедент и консеквент суммарно включают в себя не более семи объектов (разумное количество).

Проанализируйте и изложите содержательный смысл полученного результата.

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

- формулировка задания;
- гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами;
- рисунки с результатами визуализации;
- пояснения, раскрывающие смысл полученных результатов.

В соответствии с заданием была доработана программа поиска частых наборов, чтобы она также выполняла поиск ассоциативных правил. В [репозитории](#) размещено три файла: 2apriori1.ipynb, 2apriori2.ipynb – реализация поиска ассоциативных правил на разных датасетах, results.xls – результаты экспериментов и графики. [Датасеты](#) размещены в репозитории предыдущей лабораторной работы: dataset1.txt – датасет, используемый в 2apriori1.ipynb, Online\_Retail.xls – датасет, используемый в 2apriori2.ipynb.

Предобработка данных датасетов, код алгоритма, а также результаты каждого эксперимента представлены в google colab файлах, размещенных в репозитории.

### Эксперимент 1

В результате поиска наборов в первом датасете (наборы retail - сведения о покупках в супермаркете), были получены графики, представленные на рисунках 1 и 2. Значения порога поддержки взяты следующие: 1%, 3%, 5%, 10%, 15%.

На рисунке 1 можно заметить, что, в целом, время выполнения алгоритма уменьшается при увеличении порога поддержки. При небольшой разнице между порогами поддержки, время выполнения может колебаться, что можно заметить на графике при значении порога 0.1. Тем не менее можно сделать вывод, что быстродействие на фиксированном наборе данных при увеличении порога поддержки увеличивается. Это можно объяснить тем, что при большем числе параметра поддержки, составляется меньше наборов.

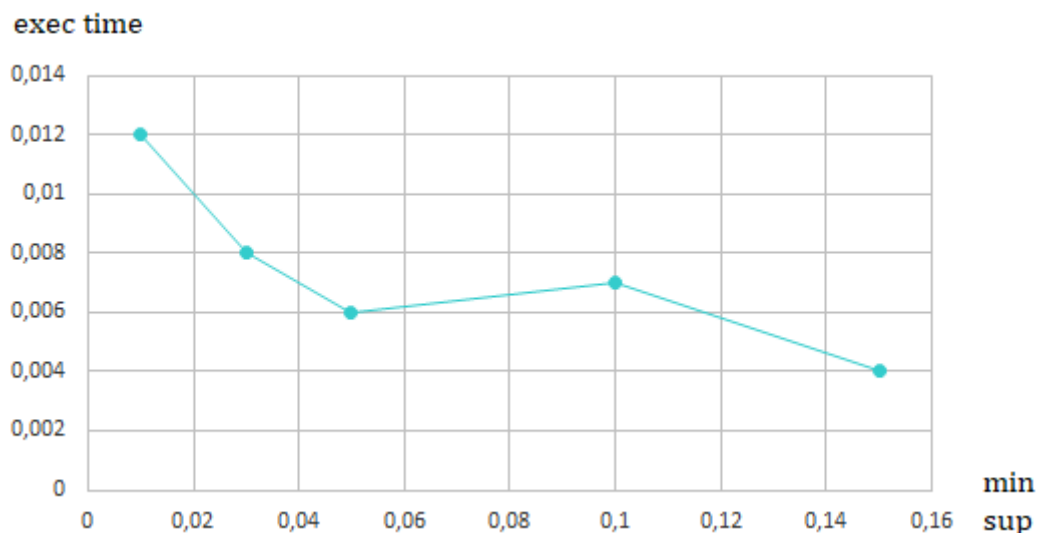


Рисунок 1 – Быстродействие при изменяемом пороге поддержки

На рисунке 2 можно заметить, что количество найденных правил на фиксированном наборе данных при увеличении порога поддержки уменьшается.

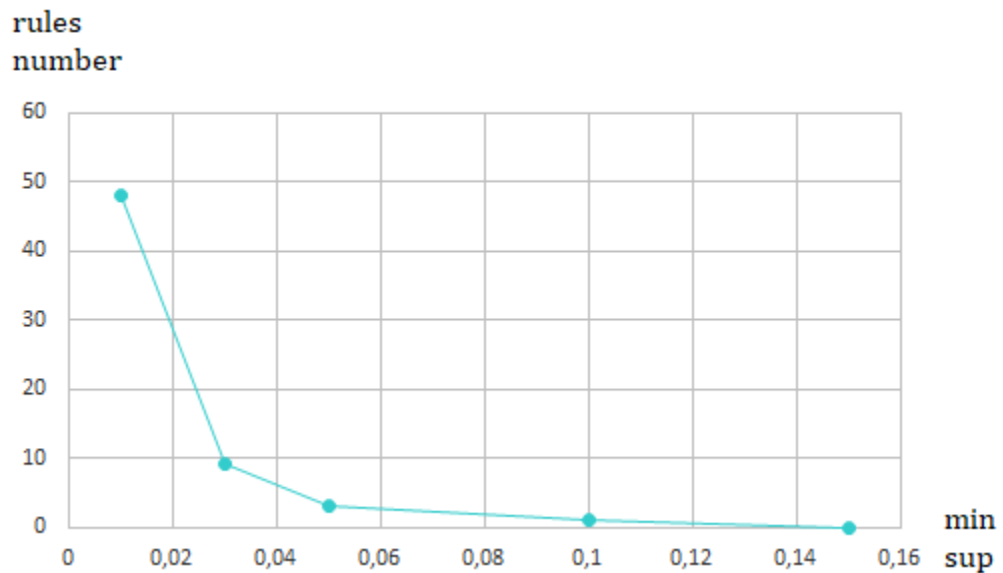


Рисунок 2 – Количество правил при изменяемом пороге поддержки

## Эксперимент 2

В результате поиска наборов во втором датасете (наборы online retail - сведения о покупках в супермаркете во Франции), были получены графики, представленные на рисунках 3 и 4. В связи со сложностью датасета, не удалось выполнить эксперимент на тех же значениях порога поддержки, что и в предыдущем эксперименте: при значениях ниже 17% происходило переполнение оперативной памяти. Были взяты следующие значения порога поддержки: 17%, 25%, 35%, 45%, 60%.

Несмотря на другие значения порога поддержки и отличный от первого эксперимента датасет, на рисунке 3 можно заметить, что график быстродействия ведет себя так же, как в первом эксперименте: быстродействие на фиксированном наборе данных при увеличении порога поддержки увеличивается.

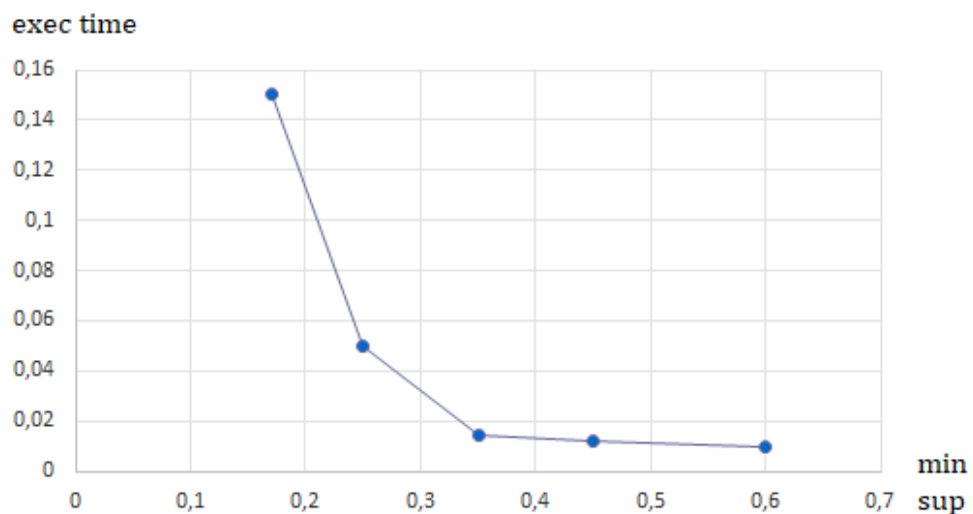


Рисунок 3 – Быстродействие при изменяемом пороге поддержки

На рисунке 3 можно заметить, что так же, как в эксперименте 1, количество найденных правил на фиксированном наборе данных при увеличении порога поддержки уменьшается.

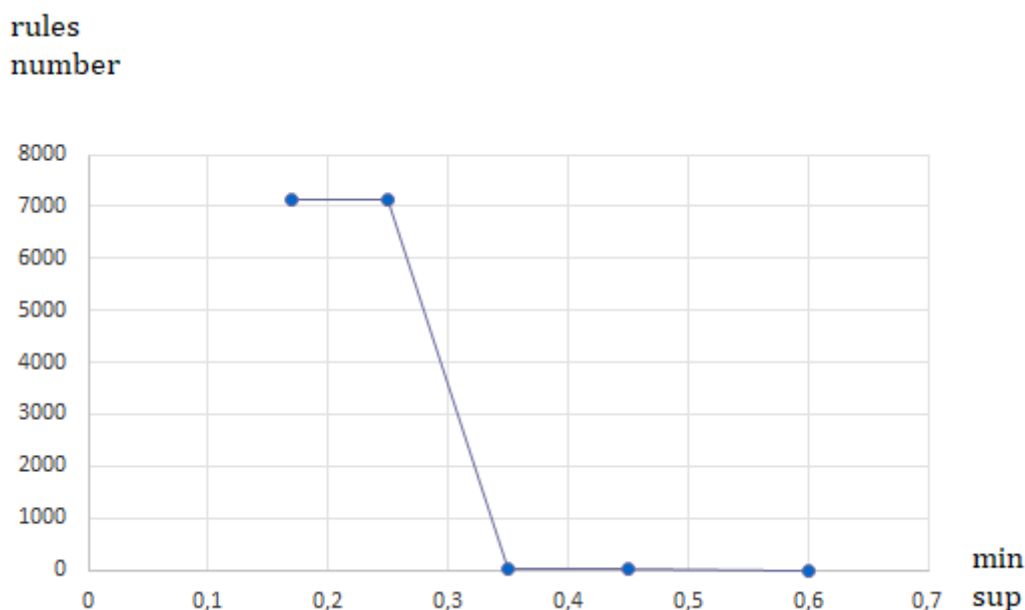


Рисунок 4 – Количество частых наборов при изменяемом пороге поддержки

Проанализируем одно из полученных правил, изображенное в таблице 1.

Было выбрано правило из второго эксперимента ( $\text{min support} = 0.45$ ,  $\text{index} = 7$ ), включающее в себя 3 объекта: SET OF SALT AND PEPPER TOADSTOOLS (набор из перечницы и солонки), LUNCH BOX WITH CUTLERY RETROSPOT (ланч бокс со столовыми приборами) и POSTAGE (марка).

antecedents	consequents	ant. sup.	con. sup.	support	confidence	lift	leverage	conviction
frozenset({'SET OF SALT AND PEPPER TOADSTOOLS', 'LUNCH BOX WITH CUTLERY RETROSPOT'})	frozenset({'POSTAGE'})	0.5	1.0	0.5	1.0	1.0	0.0	Infinity

Таблица 1 – полученное правило

Можно заметить, что покупка набора из перечницы и солонки, ланч бокс со столовыми приборами ведет к покупке марки. При этом поддержка антецедента равна 0.5, поддержка консеквента равна 1.0, поддержка набора равна 0.5 – этот показатель демонстрирует, насколько часто набор объектов обнаруживается в базе данных. Доверие в данном наборе равно 1.0 – этот показатель демонстрирует, насколько часто правило оказывается верным. Лифт в данном наборе равен 1.0 – это означает, что событие в левой части независимо от события в правой части. Усиление в данном наборе равно 0.0 – это также означает, что события независимы. Если два события независимы,

полученное правило является бесполезным. Уверенность в данном наборе равна бесконечности – это означает, что правило с бесконечной частотой дает неправильное предсказание. Итак, можно сделать вывод, что данное правило не является полезным из-за независимости антецедента и консеквента.

## **Вывод**

- Быстродействие на фиксированном наборе данных при увеличении порога поддержки увеличивается.
- Количество найденных правил на фиксированном наборе данных при увеличении порога поддержки уменьшается.