

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ОТЧЕТ
по лабораторной работе №3
по дисциплине «Машинное обучение»
Тема: Частотный анализ

Студент гр. 8304

Холковский К.В

Преподаватель

Жангиров Т. Р.

Санкт-Петербург

2021

Цель работы

Ознакомиться с методами частотного анализа из библиотеки MLxtend

Ход работы

Загрузка данных

Были загружены данные и получена следующая информация о датасете:

```
Count(ids): 1139
Count(items): 38
```

Рис 1 – Результаты загрузки данных

Подготовка данных

Данные были подготовлены для анализа

	all- purpose	aluminum foil	bagels	...	vegetables	waffles	yogurt
0	True	True	False	...	True	False	True
1	False	True	False	...	True	True	True
2	False	False	True	...	True	False	False
3	True	False	False	...	False	False	False
4	True	False	False	...	True	True	True
...
1134	True	False	False	...	False	False	False
1135	False	False	False	...	True	False	False
1136	False	False	True	...	True	False	True
1137	True	False	False	...	True	True	True
1138	False	False	False	...	True	False	False

Рис 2 – Обработанные данные

Теперь каждому id соответствует булевый список одинаковой длины.

Ассоциативный анализ с использованием алгоритма Apriori

Был применен алгоритм apriori с минимальным уровнем поддержки 0.3 и получен список из 52 элементов.

0	0.374890	(all- purpose)	1	26	0.367867	(sandwich bags)	1
1	0.384548	(aluminum foil)	1	27	0.349429	(sandwich loaves)	1
2	0.385426	(bagels)	1	28	0.368745	(shampoo)	1
3	0.374890	(beef)	1	29	0.379280	(soap)	1
4	0.367867	(butter)	1	30	0.390694	(soda)	1
5	0.395961	(cereals)	1	31	0.373134	(spaghetti sauce)	1
6	0.390694	(cheeses)	1	32	0.360843	(sugar)	1
7	0.379280	(coffee/tea)	1	33	0.378402	(toilet paper)	1
8	0.388938	(dinner rolls)	1	34	0.369622	(tortillas)	1
9	0.388060	(dishwashing liquid/detergent)	1	35	0.739245	(vegetables)	1
10	0.389816	(eggs)	1	36	0.394205	(waffles)	1
11	0.352941	(flour)	1	37	0.384548	(yogurt)	1
12	0.370500	(fruits)	1	38	0.310799	(aluminum foil, vegetables)	2
13	0.345917	(hand soap)	1	39	0.300263	(bagels, vegetables)	2
14	0.398595	(ice cream)	1	40	0.310799	(cereals, vegetables)	2
15	0.375768	(individual meals)	1	41	0.309043	(cheeses, vegetables)	2
16	0.376646	(juice)	1	42	0.308165	(dinner rolls, vegetables)	2
17	0.371378	(ketchup)	1	43	0.306409	(dishwashing liquid/detergent, vegetables)	2
18	0.378402	(laundry detergent)	1	44	0.326602	(eggs, vegetables)	2
19	0.395083	(lunch meat)	1	45	0.302897	(ice cream, vegetables)	2
20	0.380158	(milk)	1	46	0.309043	(laundry detergent, vegetables)	2
21	0.375768	(mixes)	1	47	0.311677	(lunch meat, vegetables)	2
22	0.362599	(paper towels)	1	48	0.331870	(poultry, vegetables)	2
23	0.371378	(pasta)	1	49	0.305531	(soda, vegetables)	2
24	0.355575	(pork)	1	50	0.315189	(waffles, vegetables)	2
25	0.421422	(poultry)	1	51	0.319579	(yogurt, vegetables)	2

Рис 3 – Frequent itemsets for min sup = 0.3

Был применен алгоритм *apriori* с минимальным уровнем поддержки 0.3 и максимальным количеством элементов в *itemset* – 1 и получен список из 37 элементов.

	support	itemsets
0	0.374890	(all- purpose)
1	0.384548	(aluminum foil)
2	0.385426	(bagels)
3	0.374890	(beef)
4	0.367867	(butter)
5	0.395961	(cereals)
6	0.390694	(cheeses)
7	0.379280	(coffee/tea)
8	0.388938	(dinner rolls)
9	0.388060	(dishwashing liquid/detergent)
10	0.389816	(eggs)
11	0.352941	(flour)
12	0.370500	(fruits)
13	0.345917	(hand soap)
14	0.398595	(ice cream)
15	0.375768	(individual meals)
16	0.376646	(juice)
17	0.371378	(ketchup)
18	0.378402	(laundry detergent)
19	0.395083	(lunch meat)
20	0.380158	(milk)
21	0.375768	(mixes)
22	0.362599	(paper towels)
23	0.371378	(pasta)
24	0.355575	(pork)
25	0.421422	(poultry)
26	0.367867	(sandwich bags)
27	0.349429	(sandwich loaves)

28	0.368745	(shampoo)
29	0.379280	(soap)
30	0.390694	(soda)
31	0.373134	(spaghetti sauce)
32	0.360843	(sugar)
33	0.378402	(toilet paper)
34	0.369622	(tortillas)
35	0.739245	(vegetables)
36	0.394205	(waffles)
37	0.384548	(yogurt)

Рис 4 – Frequent itemsets for minSup=0.3 maxLen=1

Был применен алгоритм apriori с минимальным уровнем поддержки 0.3 и количеством элементов в itemset – 2 и получен список из 14 элементов.

	support	itemsets	length
38	0.310799	(vegetables, aluminum foil)	2
39	0.300263	(vegetables, bagels)	2
40	0.310799	(vegetables, cereals)	2
41	0.309043	(vegetables, cheeses)	2
42	0.308165	(vegetables, dinner rolls)	2
43	0.306409	(dishwashing liquid/detergent, vegetables)	2
44	0.326602	(vegetables, eggs)	2
45	0.302897	(vegetables, ice cream)	2
46	0.309043	(vegetables, laundry detergent)	2
47	0.311677	(vegetables, lunch meat)	2
48	0.331870	(vegetables, poultry)	2
49	0.305531	(vegetables, soda)	2
50	0.315189	(vegetables, waffles)	2
51	0.319579	(vegetables, yogurt)	2

Рис 5 – Frequent itemsets for minSup=0.3 Len=2

Был построен график количества наборов от уровня поддержки:

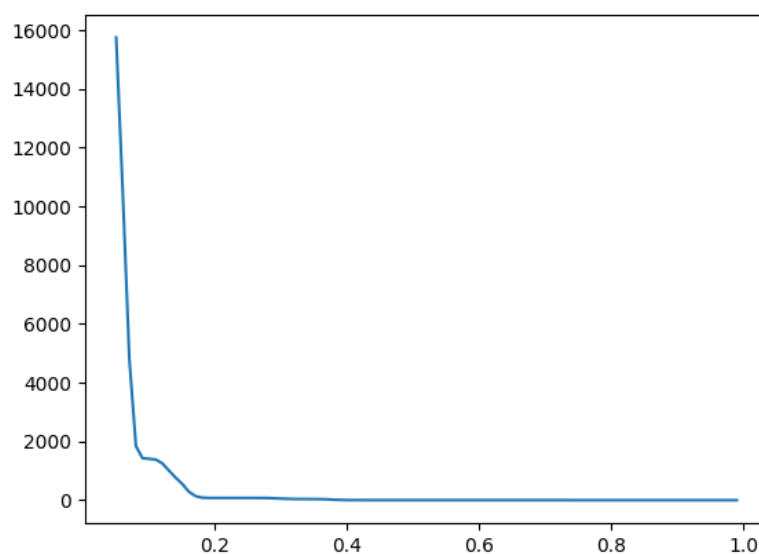


Рис 6 – График количества наборов от уровня поддержки

На графике из предыдущего пункта были отображены точки прекращения генерации наборов разного количества.

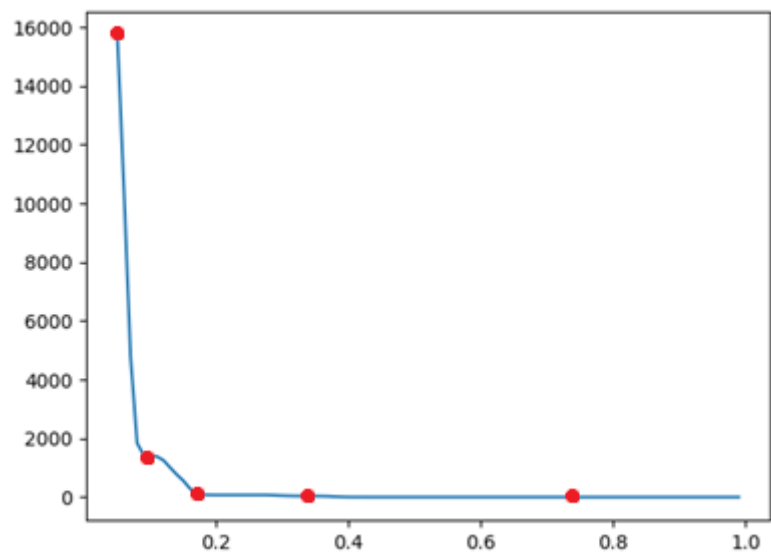


Рис 7 - График количества наборов от уровня поддержки с точками прекращения генерации наборов

Таблица 1 – Критические уровни поддержки для разного кол-ва наборов

Количество наборов	1	2	3	4	5-38
Уровень поддержки	0.74	0.34	0.17	0.09	0.05

Был построен датасет только из элементов, которые попадают в наборы размером 1 при уровне поддержки 0.38.

	support	itemsets	length
0	0.384548	(aluminum foil)	1
1	0.385426	(bagels)	1
2	0.395961	(cereals)	1
3	0.390694	(cheeses)	1
4	0.388938	(dinner rolls)	1
5	0.388060	(dishwashing liquid/detergent)	1
6	0.389816	(eggs)	1
7	0.398595	(ice cream)	1
8	0.395083	(lunch meat)	1
9	0.380158	(milk)	1
10	0.421422	(poultry)	1
11	0.390694	(soda)	1
12	0.739245	(vegetables)	1
13	0.394205	(waffles)	1
14	0.384548	(yogurt)	1
15	0.310799	(vegetables, aluminum foil)	2
16	0.300263	(bagels, vegetables)	2
17	0.310799	(vegetables, cereals)	2
18	0.309043	(vegetables, cheeses)	2
19	0.308165	(vegetables, dinner rolls)	2
20	0.306409	(vegetables, dishwashing liquid/detergent)	2
21	0.326602	(vegetables, eggs)	2
22	0.302897	(vegetables, ice cream)	2
23	0.311677	(lunch meat, vegetables)	2
24	0.331870	(poultry, vegetables)	2
25	0.305531	(vegetables, soda)	2
26	0.315189	(vegetables, waffles)	2
27	0.319579	(yogurt, vegetables)	2

Рис 8 – apriori для нового набора данных

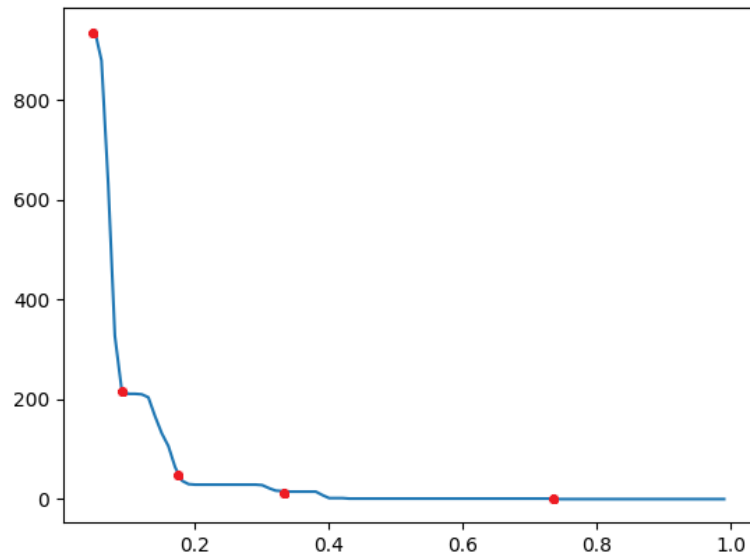


Рис 9 - График количества наборов от уровня поддержки с точками прекращения генерации наборов
Точки соответствуют точкам, для изначального набора данных.

Были выведены все наборы размер, которых больше 1 и в котором есть “yogurt” или “waffles”, всего 30 наборов.

```
frozenset({'aluminum foil', 'waffles'})
frozenset({'aluminum foil', 'yogurt'})
frozenset({'bagels', 'waffles'})
frozenset({'bagels', 'yogurt'})
frozenset({'cereals', 'waffles'})
frozenset({'cereals', 'yogurt'})
frozenset({'cheeses', 'waffles'})
frozenset({'yogurt', 'cheeses'})
frozenset({'dinner rolls', 'waffles'})
frozenset({'dinner rolls', 'yogurt'})
frozenset({'dishwashing liquid/detergent', 'waffles'})
frozenset({'dishwashing liquid/detergent', 'yogurt'})
frozenset({'eggs', 'waffles'})
frozenset({'yogurt', 'eggs'})
frozenset({'ice cream', 'waffles'})
frozenset({'ice cream', 'yogurt'})
frozenset({'lunch meat', 'waffles'})
frozenset({'lunch meat', 'yogurt'})
frozenset({'yogurt', 'milk'})
frozenset({'poultry', 'waffles'})
frozenset({'poultry', 'yogurt'})
frozenset({'waffles', 'soda'})
frozenset({'yogurt', 'soda'})
frozenset({'vegetables', 'waffles'})
frozenset({'yogurt', 'vegetables'})
frozenset({'yogurt', 'waffles'})
frozenset({'aluminum foil', 'yogurt', 'vegetables'})
frozenset({'yogurt', 'eggs', 'vegetables'})
frozenset({'lunch meat', 'vegetables', 'waffles'})
frozenset({'poultry', 'yogurt', 'vegetables'})
```

Рис 10 – Наборы для задания

Для набора данных, которые не попали в набор данных из п.6 был проведен анализ:

	support	itemsets	length
0	0.374890	(all- purpose)	1
1	0.374890	(beef)	1
2	0.367867	(butter)	1
3	0.379280	(coffee/tea)	1
4	0.352941	(flour)	1
5	0.370500	(fruits)	1
6	0.345917	(hand soap)	1
7	0.375768	(individual meals)	1
8	0.376646	(juice)	1
9	0.371378	(ketchup)	1
10	0.378402	(laundry detergent)	1
11	0.375768	(mixes)	1
12	0.362599	(paper towels)	1
13	0.371378	(pasta)	1
14	0.355575	(pork)	1
15	0.367867	(sandwich bags)	1
16	0.349429	(sandwich loaves)	1
17	0.368745	(shampoo)	1
18	0.379280	(soap)	1
19	0.373134	(spaghetti sauce)	1
20	0.360843	(sugar)	1
21	0.378402	(toilet paper)	1
22	0.369622	(tortillas)	1

Рис 11 – apriori для набора данных, которые не попали в набор данных из п.6

Был написан метод выводящий наборы, в которых хотябы 2 элемента начинаются на s.

```
frozenset({'vegetables', 'soda', 'sandwich loaves'})
frozenset({'vegetables', 'sandwich loaves', 'spaghetti sauce'})
frozenset({'vegetables', 'sandwich loaves', 'sugar'})
frozenset({'vegetables', 'shampoo', 'soap'})
frozenset({'vegetables', 'soda', 'shampoo'})
frozenset({'vegetables', 'shampoo', 'spaghetti sauce'})
frozenset({'vegetables', 'shampoo', 'sugar'})
frozenset({'vegetables', 'soda', 'soap'})
frozenset({'vegetables', 'soap', 'spaghetti sauce'})
frozenset({'vegetables', 'soap', 'sugar'})
frozenset({'vegetables', 'soda', 'spaghetti sauce'})
frozenset({'vegetables', 'soda', 'sugar'})
frozenset({'vegetables', 'sugar', 'spaghetti sauce'})
42
```

Рис 12 – Пример для min_sup = 0.1 у полного набора данных

Был написан метод выводящий все наборы, минимальная поддержка которых (0.1; 0.25]

```
{frozenset({'ketchup', 'butter'}), frozenset({'poultry', 'laundry detergent'})}
Кол-во элементов: 1331
```

Рис 13 – Пример для полного набора данных

Вывод

Ознакомились с методами частотного анализа из библиотеки MLxtend.