

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

ОТЧЕТ
по лабораторной работе №4
по дисциплине «Машинное обучение»
Тема: Ассоциативный анализ

Студент гр. 8304

Холковский К.В

Преподаватель

Жангиров Т. Р.

Санкт-Петербург

2021

Цель работы

Ознакомиться с методами ассоциативного анализа из библиотеки MLxtend

Ход работы

Загрузка данных

Были загружены данные:

```
{'photo/film', 'soups', 'skin care', 'butter milk', 'sugar',  
169
```

Рис 1 – Загруженные данные

FPGrowth и FPMax

Данные были приведены к удобному для анализа виду при помощи TransactionEncoder:

```
Instant food products  UHT-milk  ...  yogurt  zwieback  
0          False      False  ...   False   False  
1          False      False  ...    True   False  
2          False      False  ...   False   False  
3          False      False  ...    True   False  
4          False      False  ...   False   False  
...           ...         ...  ...     ...     ...  
9830         False      False  ...   False   False  
9831         False      False  ...   False   False  
9832         False      False  ...    True   False  
9833         False      False  ...   False   False  
9834         False      False  ...   False   False  
  
[9835 rows x 169 columns]
```

Рис 2 – Удобные для анализа данные

Был проведен ассоциативный анализ используя FPGrowth при уровне поддержки 0.03:

	support	itemsets
0	0.082766	(citrus fruit)
1	0.058566	(margarine)
2	0.139502	(yogurt)
3	0.104931	(tropical fruit)
4	0.058058	(coffee)
..
58	0.033249	(pastry, whole milk)
59	0.047382	(other vegetables, root vegetables)
60	0.048907	(root vegetables, whole milk)
61	0.030605	(rolls/buns, sausage)
62	0.032232	(whipped/sour cream, whole milk)

Рис 3 – Результат FPGrowth при minSup=0.03

Был проведен анализ полученных данных:

```

Max for len: 1 is: 0.25551601423487547
Min for len: 1 is: 0.03040162684290798
Max for len: 2 is: 0.07483477376715811
Min for len: 2 is: 0.030096593797661414

```

Рис 4 – Анализ полученных данных

Был проведен ассоциативный анализ используя FPMax при уровне поддержки 0.03:

	support	itemsets
0	0.030402	(specialty chocolate)
1	0.031012	(onions)
2	0.032944	(hygiene articles)
3	0.033249	(berries)
4	0.033249	(hamburger meat)
..
45	0.038332	(soda, rolls/buns)
46	0.040061	(soda, whole milk)
47	0.042603	(other vegetables, rolls/buns)
48	0.056634	(rolls/buns, whole milk)
49	0.074835	(other vegetables, whole milk)

Рис 5 -Результаты FPMax

Был проведен анализ полученных данных:

```

Max for len: 1 is: 0.09852567361464158
Min for len: 1 is: 0.03040162684290798
Max for len: 2 is: 0.07483477376715811
Min for len: 2 is: 0.030096593797661414

```

Рис 6 – Анализ полученных данных

Алгоритмы работают одинаково, но в FPMax входят только максимальные наборы элементов. Набор элементов называется максимальным, если он является частым и не существует частого супер-шаблона, содержащего его. Поэтому результаты для наборов длины 2 не изменились.

Была построена гистограмма для товаров:

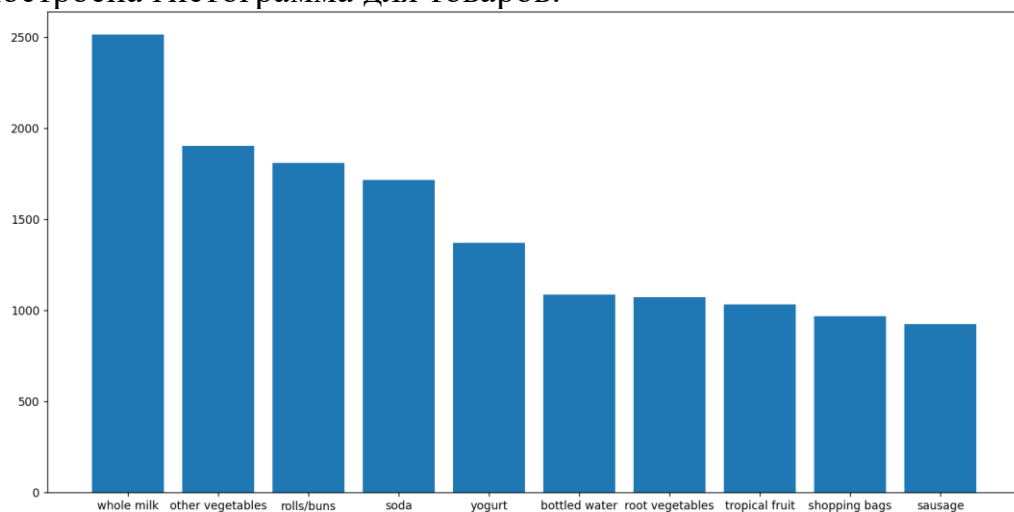


Рис 7 – Гистограмма для всех товаров

Был проведен анализ для нового набора данных:

	support	itemsets
0	0.082766	(citrus fruit)
1	0.139502	(yogurt)
2	0.104931	(tropical fruit)
3	0.255516	(whole milk)
4	0.193493	(other vegetables)
29	0.033249	(whole milk, pastry)
30	0.047382	(root vegetables, other vegetables)
31	0.048907	(whole milk, root vegetables)
32	0.030605	(rolls/buns, sausage)
33	0.032232	(whole milk, whipped/sour cream)

Рис 8 – FPGrowth для нового датасета

```

Max for len: 1 is: 0.25551601423487547
Min for len: 1 is: 0.05765124555160142
Max for len: 2 is: 0.07483477376715811
Min for len: 2 is: 0.030503304524656837

```

Рис 9 – Анализ полученных данных

	support	itemsets
0	0.057651	(pork)
1	0.032232	(whipped/sour cream, whole milk)
2	0.077682	(canned beer)
3	0.080529	(bottled beer)
4	0.030503	(whole milk, citrus fruit)
5	0.033249	(pastry, whole milk)
6	0.030605	(rolls/buns, sausage)
7	0.098526	(shopping bags)
8	0.035892	(other vegetables, tropical fruit)
9	0.042298	(whole milk, tropical fruit)
10	0.047382	(other vegetables, root vegetables)
11	0.048907	(whole milk, root vegetables)
12	0.034367	(bottled water, whole milk)
13	0.034367	(rolls/buns, yogurt)
14	0.043416	(other vegetables, yogurt)
15	0.056024	(yogurt, whole milk)
16	0.032740	(other vegetables, soda)
17	0.038332	(rolls/buns, soda)
18	0.040061	(whole milk, soda)
19	0.042603	(rolls/buns, other vegetables)
20	0.056634	(rolls/buns, whole milk)
21	0.074835	(other vegetables, whole milk)

Рис 10 – FPMaх для нового датасета

```

Max for len: 1 is: 0.09852567361464158
Min for len: 1 is: 0.05765124555160142
Max for len: 2 is: 0.07483477376715811
Min for len: 2 is: 0.030503304524656837

```

Рис 11 – Анализ полученных данных

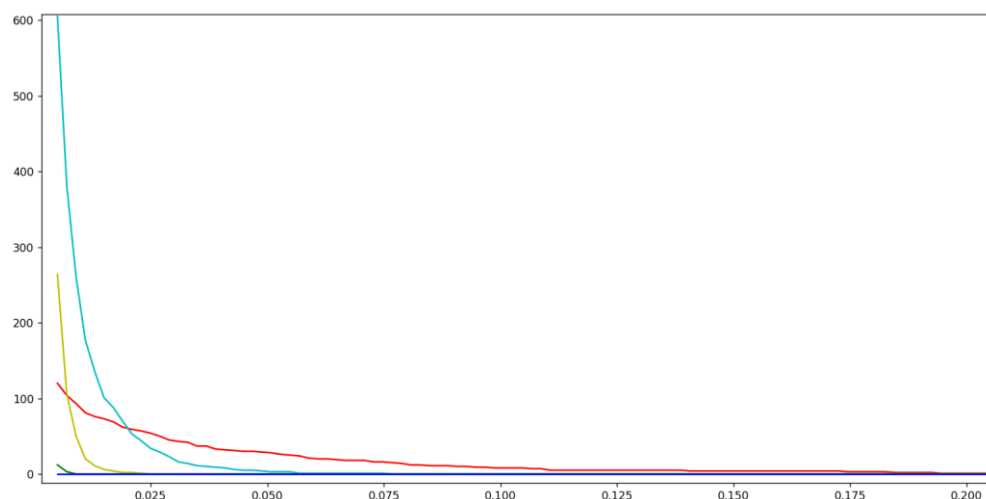


Рис 12 – График изменения количества правил от поддержки для разных длин наборов

1: красный, 2: циановый, 3: желтый, 4: зеленый, 5: blue.

Ассоциативные правила

Был выполнен анализ:

	antecedents	consequents	...	leverage	conviction
0	(yogurt)	(whole milk)	...	0.009183	1.070481
1	(yogurt)	(other vegetables)	...	0.005150	1.033172
2	(tropical fruit)	(yogurt)	...	0.013156	1.102890
3	(tropical fruit)	(other vegetables)	...	0.008804	1.076706
4	(tropical fruit)	(whole milk)	...	0.005359	1.052495
5	(whole milk)	(other vegetables)	...	0.006849	1.025026
6	(other vegetables)	(whole milk)	...	0.006849	1.036649
7	(rolls/buns)	(whole milk)	...	-0.012801	0.930450
8	(bottled water)	(whole milk)	...	-0.010177	0.913309
9	(bottled water)	(soda)	...	0.007832	1.061153
10	(citrus fruit)	(whole milk)	...	-0.001349	0.984313
11	(citrus fruit)	(other vegetables)	...	0.008135	1.091192
12	(root vegetables)	(other vegetables)	...	0.028050	1.273671
13	(root vegetables)	(whole milk)	...	0.014031	1.141049
14	(sausage)	(rolls/buns)	...	0.010985	1.102730
15	(sausage)	(whole milk)	...	-0.011477	0.894062
16	(sausage)	(other vegetables)	...	-0.002776	0.975687
17	(whipped/sour cream)	(whole milk)	...	0.011419	1.189023
18	(whipped/sour cream)	(other vegetables)	...	0.015557	1.232002
19	(pastry)	(whole milk)	...	0.002304	1.027179

Рис 13 – Результаты анализа

Расчет проводился на основе метрики “confidence”

Метрика	Описание
support	$\text{support}(A \rightarrow C) = \text{support}(A \cap C)$

confidence	$\text{confidence}(A \rightarrow C) = \frac{\text{support}(A \rightarrow C)}{\text{support}(A)}$
lift	$\text{lift}(A \rightarrow C) = \frac{\text{confidence}(A \rightarrow C)}{\text{support}(C)}$
leverage	$\text{leverage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) \times \text{support}(C)$
conviction	$\text{conviction}(A \rightarrow C) = \frac{1 - \text{support}(C)}{1 - \text{confidence}(A \rightarrow C)}$

Были рассчитаны среднее значение, медиана и СКО для каждой метрики:

Метрика	Среднее	Медина	СКО
support	0.07468	0.06695	0.02254
confidence	0.28957	0.26443	0.10368
lift	1.04299	1.05608	0.18326
leverage	0.01553	0.01359	0.00606
conviction	1.01719	1.02285	0.08399

Был построен граф:

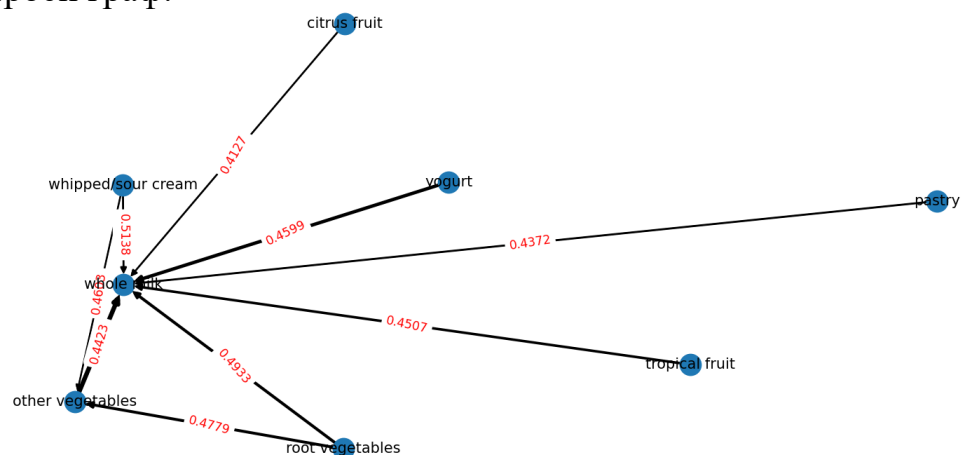


Рис 14 – Граф анализа

Вывод

Ознакомились с ассоциативного частотного анализа из библиотеки MLxtend.