



МАШИННОЕ ОБУЧЕНИЕ &

---

МАРКЕТИНГ

КАШНИЦКИЙ ЮРИЙ, ФКН НИУ ВШЭ

# АССОЦИАТИВНЫЕ ПРАВИЛА

- ▶ Человек купил пиво -> он купил чипсы
- ▶ Такие правила можно находить в данных автоматически
- ▶ Ассоциативное правило - это четверка  $(A, B, c, s)$  вида  $A \rightarrow B$  с поддержкой  $c$  и достоверностью  $s$ . Здесь  $A$  и  $B$  - какие-то подмножества, например, множеств купленных товаров.

ТАБЛИЦА ПОКУПОК

	Хлеб	Молоко	Сыр	Пиво	Сок
Игнат	Х	Х	Х	Х	
Осип		Х	Х	Х	Х
Изя		Х	Х		
Глеб		Х	Х	Х	
Борис	Х		Х	Х	



ТАБЛИЦА ПОКУПОК

	a	b	c	d	e
1	X	X	X	X	
2		X	X	X	X
3		X	X		
4		X	X	X	
5	X		X	X	

# ПОИСК ЧАСТЫХ МНОЖЕСТВ ТОВАРОВ

- ▶ Найдем в таблице покупок все *частые* множества товаров: поддержка больше  $\text{min\_supp} = 40\%$
- ▶ В начале отберем все частые товары (которые купили 40% и более покупателей). Это все, кроме e

a	40 %
b	80 %
c	100 %
d	80 %

# ОСНОВНОЕ СВОЙСТВО APRIORI

- ▶ Если множество  $\{g1, g2\}$  часто ( $\text{sup}(\{g1, g2\}) > s$ ), то есть товары  $g1$  и  $g2$  купили не менее  $s$  процентов людей, то с добавлением элементов поддержка не может возрасть (очевидно)
- ▶  $\text{sup}(\{g1, g2\}) > s \Rightarrow \text{sup}(\{g1, g2, g3, \dots\}) \leq s$
- ▶ Здравый смысл: увеличиваем множество товаров, значит, людей, купивших все эти товары, может стать только меньше (или остаться столько же)

# ОСНОВНОЕ СВОЙСТВО APRIORI

- ▶ В обратную сторону:  $\{g_1, g_2, g_3\}$  - частое подмножество товаров (то есть  $\text{sup}(\{g_1, g_2, g_3\}) \geq s \Rightarrow \{g_1, g_2\}$  - частое,  $\{g_2, g_3\}$ , ...,  $\{g_3\}$  - частое.
- ▶ То есть все подмножества частого множества тоже частые
- ▶ Значит, частыми множествами могут быть только те, у которых любые подмножества частые

## ВТОРОЙ ШАГ

- ▶ Частые пары товаров:

a,c	40 %
a,d	40 %
b,c	80 %
b,d	60 %
c,d	80 %



### ТРЕТИЙ ШАГ

- ▶ Кандидаты в частые тройки товаров
- ▶ Какие тройки могут быть частыми?
- ▶ Те, у которых, все подмножества частые
- ▶  $\{a, c, d\}$ ,  $\{b, c, d\}$  - это кандидаты в частые тройки. Других не может быть.
- ▶ Например,  $\{a, b, d\}$  не может быть частой тройкой, поскольку  $\{a, b\}$  - не часто.

### ТРЕТИЙ ШАГ

- ▶  $\{a, c, d\}, \{b, c, d\}$  - кандидаты в частые тройки.  
Проверяем. Оба множества частые (но могло быть и по-другому).
- ▶ Итак,  $\{a, c, d\}, \{b, c, d\}$  - тройки товаров, часто идущие вместе. Частых четверок нет (почему?)
- ▶ В нашем примере частые тройки товаров -  $\{\text{хлеб, сыр, пиво}\}$  (поддержка - 40%) и  $\{\text{молоко, сыр, пиво}\}$  (поддержка - 60%)

# АССОЦИАТИВНЫЕ ПРАВИЛА НА ОСНОВЕ APRIORI

- ▶ Частые множества товаров  $\{a, c, d\}$ ,  $\{b, c, d\}$  помогают найти ассоциативные правила:  
 $\{a, c\} \rightarrow \{d\}$ ,  $\{a, d\} \rightarrow \{c\}$ ,  $\{c, d\} \rightarrow \{a\}$ ,  
 $\{b, c\} \rightarrow \{d\}$ ,  $\{b, d\} \rightarrow \{c\}$ ,  $\{b, d\} \rightarrow \{a\}$
- ▶ У таких правил поддержка не ниже найденных поддержек частых подмножеств товаров. Но достоверность может снизиться
- ▶ Например, у правила  $\{b, d\} \rightarrow \{a\}$  поддержка 40%, но достоверность - 33%
- ▶ Часто ставят ограничение и на достоверность - `min_conf`

# РЕКОМЕНДАЦИЯ ТОВАРОВ (ПРИМИТИВНАЯ)

- ▶ Например, нашли правило {молоко, сыр} -> {пиво}, верное для 60% покупателей
- ▶ Покупателю, взявшему молоко и сыр, можно «рекомендовать» пиво
- ▶ Можно товары, хорошо идущие вместе, ставить рядом (мерчендайзинг). Пример - Nutella в хлебном отделе



ЗАДАНИЕ

- ▶ Найти множество товаров, часто покупаемых вместе  
 $\text{min\_sup} = 20\%$ ,  $\text{min\_conf} = 80\%$

	Ябл.	Печ-е	Бан.	Творог	Мясо	Орехи	Вино
1		x		x		x	
2	x		x	x		x	
3	x	x		x	x		x
4		x	x	x	x		x
5		x	x		x		

# РЕКОМЕНДАЦИЯ УСЛУГ

- ▶ Другой пример - поиск услуг, часто подключаемых вместе и рекомендация услуг с помощью Apriori и ассоциативных правил

	serv_1	serv_2	serv_3
client_1		x	
client_2	x		x
client_3		x	x



ЗАДАНИЕ

- ▶ Найти множество услуг, часто подключаемых вместе  
 $\text{min\_sup} = 20\%$ , найти правила с  $\text{min\_conf} = 80\%$

	s1	s2	s3	s4	s5	s6	s7
1	x	x			x	x	
2	x	x	x			x	
3			x	x	x		
4		x				x	x
5			x	x	x		x

# АКТУАЛЬНЫЕ ДЛЯ МАРКЕТИНГА ЗАДАЧИ

- ▶ Показ онлайн-рекламы и баннеров
- ▶ Предсказание оттока клиентов
- ▶ Кластеризация/сегментирование пользователей/товаров
- ▶ Поиск необычных клиентов (поиск выбросов)
- ▶ Рекомендация товаров и услуг
- ▶ Кредитный скоринг
- ▶ Семантический анализ слоганов
- ▶ ....

## СОРЕВНОВАНИЯ KAGGLE

- ▶ «Give me some credit»
- ▶ Otto group «Classify products into correct category»
- ▶ «Avito Context Ad Clicks»
- ▶ Cristo Labs «Display Advertising Challenge»
- ▶ Avazu «Click-Through Rate Prediction»
- ▶ Rossmann Store Sales

## ДРУГИЕ СОРЕВНОВАНИЯ

- ▶ Соревнование Tinkoff по кредитному скорингу
- ▶ Соревнование Tinkoff «Угадай мерчанта»
- ▶ Контест Билайн по предсказанию возрастных групп клиентов
- ▶ ...