

Ассоциативные правила

Лекция 6

Поиск покупательских шаблонов

Если пара товаров X и Y часто покупается вместе, то:























- реклама товара X может быть направлена на покупателей товара Y ;
- товары X и Y могут быть размещены на одной и той же полке, чтобы побудить покупателей одного товара к приобретению второго;
- товары X и Y могут быть скомбинированы в некий новый продукт, такой как X со вкусом Y

Поддержка, достоверность и лифт

Мера 1: поддержка.

Поддержка показывает то, как часто данный товарный набор появляется, что измеряется долей покупок, в которых он присутствует. В табл. 1 {яблоко} появляется в четырех из восьми покупок, значит, его поддержка 50 %. Товарные наборы могут содержать и несколько элементов. Например, поддержка набора {яблоко, пиво, рис} — два из восьми, то есть 25 %.























$$\text{Поддержка } \{\text{яблоко}\} = \frac{4}{8}$$

Покупка 1	   
Покупка 2	  
Покупка 3	 
Покупка 4	 
Покупка 5	   
Покупка 6	  
Покупка 7	 
Покупка 8	 

Мера 2: достоверность

Достоверность показывает, как часто товар Y появляется вместе с товаром X, что выражается как $\{X \rightarrow Y\}$. Это измеряется долей их одновременных появлений. Согласно табл. 1, достоверность $\{\text{яблоко} \rightarrow \text{пиво}\}$ соответствует трем из четырех, то есть 75 %.

$$\text{Достоверность } \{\text{яблоко} \rightarrow \text{пиво}\} = \frac{\text{Поддержка } \{\text{яблоко}, \text{пиво}\}}{\text{Поддержка } \{\text{яблоко}\}}$$























Покупка 1	   
Покупка 2	  
Покупка 3	 
Покупка 4	 
Покупка 5	   
Покупка 6	  
Покупка 7	 
Покупка 8	 

Мера 3: лифт

Лифт отражает то, как часто товары X и Y появляются вместе, одновременно учитывая, с какой частотой появляется каждый из них.

Таким образом, лифт {яблоко->пиво} равен достоверности {яблоко->пиво}, деленной на частоту {пива}.

$$\text{Лифт} \{ \text{яблоко} \rightarrow \text{пиво} \} = \frac{\text{Поддержка} \{ \text{яблоко}, \text{пиво} \}}{\text{Поддержка} \{ \text{яблоко} \} \times \text{Поддержка} \{ \text{пиво} \}}$$

Покупка 1	   
Покупка 2	  
Покупка 3	 
Покупка 4	 
Покупка 5	   
Покупка 6	  
Покупка 7	 
Покупка 8	 

Пример: ведение продуктовых продаж

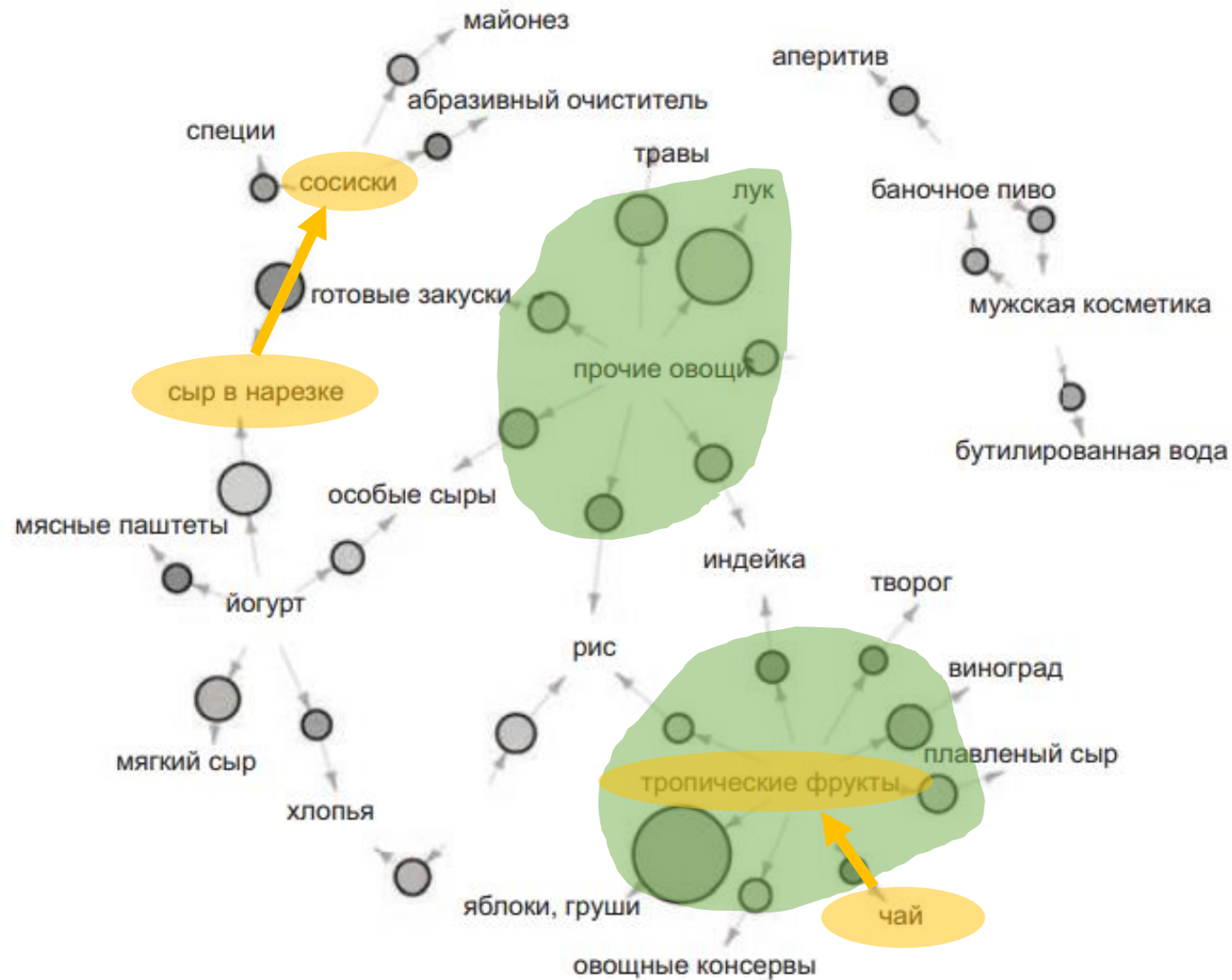


Таблица 2. Ассоциативные метрики для трех правил, связанных с пивом

Покупка	Поддержка	Достоверность	Лифт
Пиво → Газировка	1,38 %	17,8 %	1,0
Пиво → Ягоды	0,08 %	1,0 %	0,3
Пиво → Мужская косметика	0,09 %	1,2 %	2,6

Правило {пиво->газировка} имеет высокую достоверность — 17,8 %.

Однако и пиво, и газировка вообще часто появляются среди покупок, поэтому их ассоциация может оказаться **простым совпадением**.

Таблица 3. Значение поддержки для отдельных товаров в правилах, связанных с пивом

Покупка	Поддержка
Пиво	7,77 %
Газировка	17,44 %
Ягоды	3,32 %
Мужская косметика	0,46 %

С другой стороны, правило {пиво->мужская косметика} имеет низкую достоверность.

Тем не менее если кто-то покупает ее, он, вероятно, купит также и пиво.

Для пары {пиво->ягоды} верно обратное.

Если кто-то покупает пиво, то он, скорее всего, не возьмет ягод.

Принцип A priori

Если какой-то товарный набор редкий, то и большие наборы, которые его включают, тоже должны быть редки

Поиск товарных наборов с высокой поддержкой

Шаг 0: начать с товарных наборов, содержащих всего один элемент

Шаг 1: вычислить поддержку для каждого товарного набора. Оставить наборы, удовлетворяющие порогу, и отбросить остальные.

Шаг 2: увеличить размер анализируемого товарного набора на единицу и сгенерировать все возможные конфигурации, используя товарные наборы из предыдущего шага.

Шаг 3: повторять шаги 1 и 2, вычисляя поддержку для возрастающих товарных наборов до тех пор, пока они не закончатся.

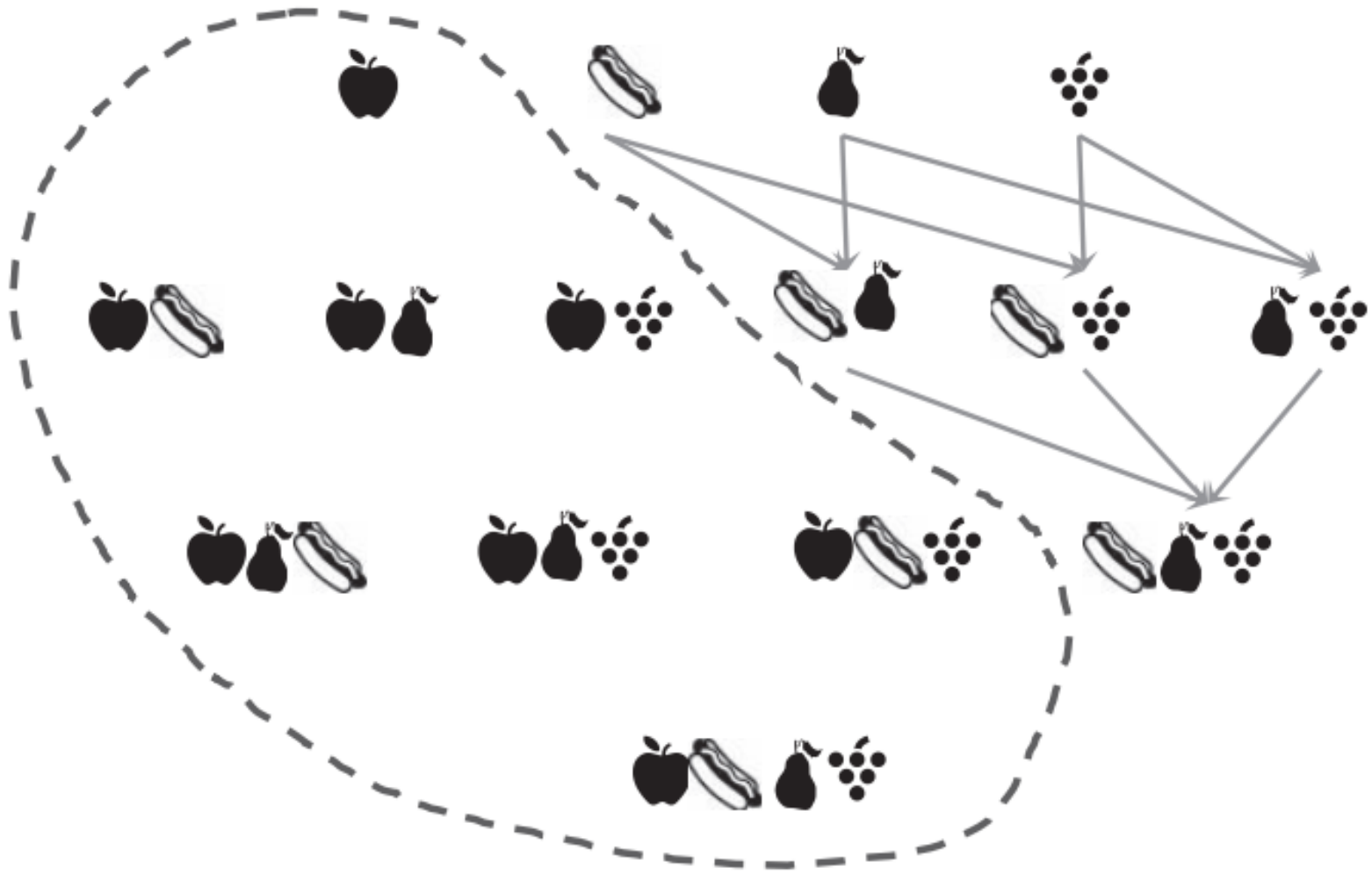


Рис. 5. Товарные наборы в пределах пунктирной линии будут отброшены

Поиск товарных правил с высокой достоверностью или лифтом

принцип Apriori также может помочь найти товарные ассоциации с высокой достоверностью или лифтом. Поиск этих ассоциаций требует меньше вычислений, поскольку если товарные наборы с высокой поддержкой известны, то достоверность и лифт вычисляются уже с использованием значения поддержки.

Ограничения

- Требует долгих вычислений
- Ложные ассоциации

Краткие итоги

- Ассоциативные правила выявляют то, как часто элементы появляются вообще и в связи с другими.
- Есть три основных способа оценки ассоциации:
 1. Поддержка $\{X\}$ показывает, как часто X появляется.
 2. Достоверность $\{X \rightarrow Y\}$ показывает, как часто Y появляется в присутствии X .
 3. Лифт $\{X \rightarrow Y\}$ показывает, как часто элементы X и Y появляются вместе по сравнению с тем, как часто они появляются по отдельности.
- Принцип Apriori ускоряет поиск часто встречающихся товарных наборов, отбрасывая значительную долю редких.