

# Лекция 6. Ассоциативные правила

2024/2025 учебный год

Доцент кафедры ИВЭ, Махно В.В.

©Создано при помощи <https://sberuniversity.ru/>



# Ассоциации

Правил может получиться много. Как понять, какие из них заслуживают внимания? Нужно обратить внимание на два параметра. **Первый – поддержка.** Она означает число или процент транзакций, которые содержат один набор данных. Для примера возьмем пару «хлеб и молоко». Мы можем встретить четыре вида чеков: без хлеба и молока, с хлебом, но без молока, с молоком, но без хлеба, и с хлебом и молоком. Нас интересует последний, так как именно он содержит необходимый набор данных и помогает нам определить значение параметра поддержки.

**Второй параметр – достоверность.** Он определяет процент транзакций, где фигурируют оба продукта сразу. То есть, показывает, насколько высока вероятность, что в чеке, где есть хлеб, окажется и молоко. Если у правила эти два параметра достигают высоких значений, то оно берется в работу





---

## Ассоциации

Задача поиска ассоциаций или ассоциативных правил, формулируется так: нужно найти закономерности вида «если покупатель купил <некоторый набор товаров>, то он купит и <еще один товар>» — эта закономерность и называется ассоциативным правилом.

Программное обеспечение — язык Python и специальные библиотеки, например, Orange.

# Ассоциативные правила

Ассоциативные правила применяются для оптимизации размещения товаров на полках магазинов, для формирования персональных рекомендаций в онлайн-магазинах (хотя там больше используются рекомендательные системы, о которых мы поговорим ниже) и для планирования акций (можно сделать скидку на один товар, зная, что вместе с ним купят другой, более дорогой товар).

Поиск ассоциативных правил обычно выполняется в два этапа. **На первом этапе** находятся часто покупаемые наборы товаров: алгоритм проходит по всему набору данных много раз, вычисляет частоты различных комбинаций товаров и выделяет часто встречающиеся комбинации. **На втором этапе** происходит непосредственное выделение правил на основе найденных на первом этапе частых комбинаций. Самые известные алгоритмы поиска ассоциативных правил — **Apriori** и **FP-Growth**.



# Ассоциативные правила

При поиске правил часто отслеживают следующие числовые показатели комбинаций товаров: **поддержка** (support) и **доверие** (confidence, переводят как достоверность или значимость).

- Поддержка — это частота, сколько раз комбинация товаров встретилась в чеках, разделить на общее число чеков. Например, в какой доле чеков встретилась комбинация «книга, журнал, ручка».
- Доверие определяется для правила, например «книга, журнал  $\sim$  ручка» и означает, в скольких процентах случаев, в которых купили книгу и журнал, также купили и ручку. На первом этапе поиска ассоциативных правил комбинации товаров фильтруются по поддержке, а на втором ищутся ассоциативные правила, превышающие порог на доверие.

# Пример. Привести расчеты поддержки и доверия

ID	Мука	Разрыхлитель	Масло	Пиво	Хлеб	Молоко
1	1	1	1	0	0	0
2	0	0	1	0	1	1
3	0	0	0	1	1	0
4	1	1	1	0	0	1
5	1	1	0	0	0	0
6	0	0	1	1	0	0
7	1	1	1	0	1	1
8	0	1	0	1	1	0

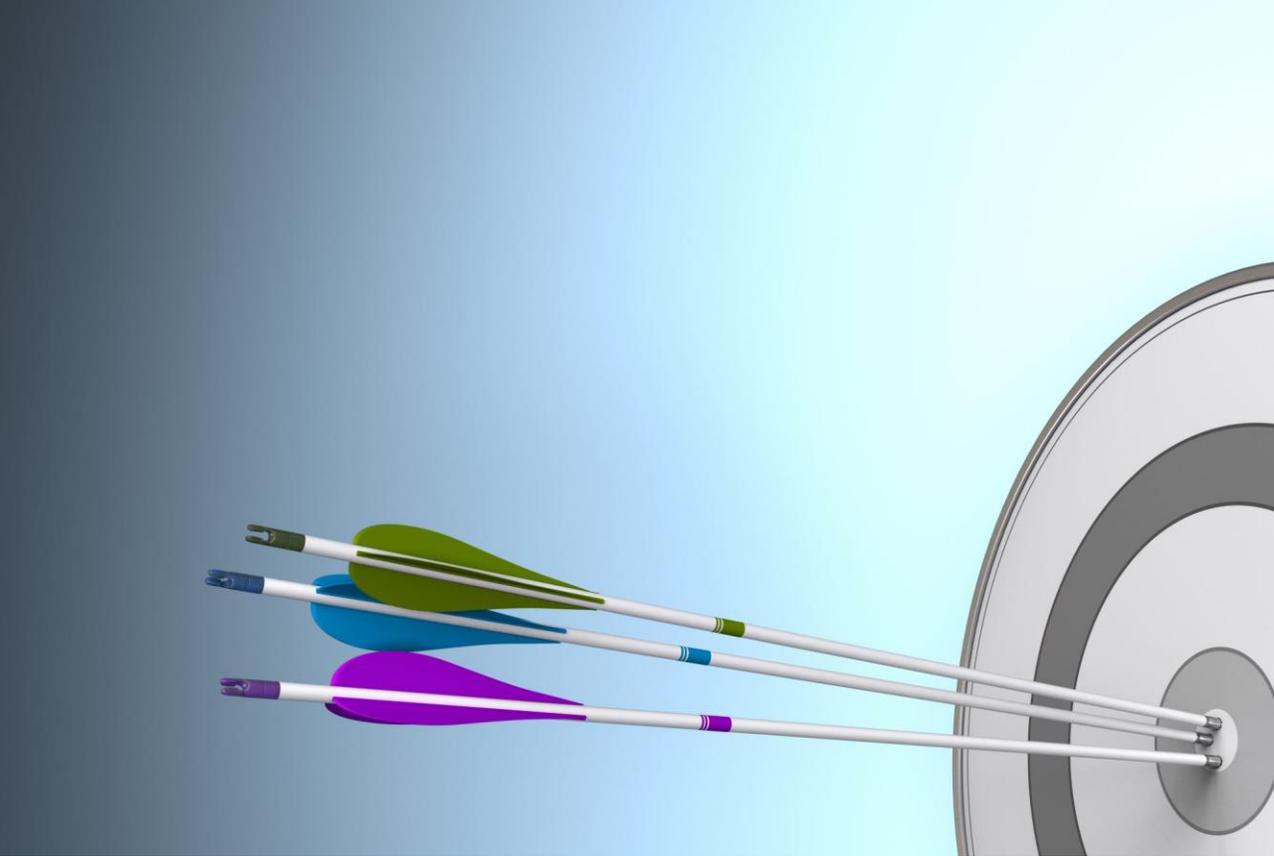
## Рекомендательные системы

Ассоциативные правила позволяют обнаружить, какие товары покупают часто вместе с другими, на основе этого можно предложить клиенту, что ему еще купить. Такая постановка задачи гораздо более широкая и называется задачей построения рекомендательных систем, или задачей рекомендаций.

# Рекомендательные системы

Данные в задаче рекомендаций выглядят как таблица, по строкам которой клиенты, по столбцам — товары (фильмы, книги и т.д.), и для некоторых ячеек известна оценка клиента для товара (для большинства ячеек таблицы оценка не известна, поскольку каждый клиент, скорее всего, взаимодействовал только с небольшим количеством товаров).



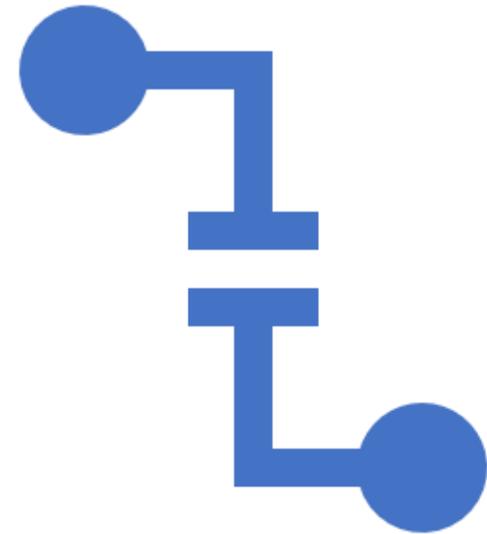


## Контентный подход

Самый простой способ построить рекомендательную систему — свести задачу к задаче регрессии. В данном случае объектом будет пара (клиент, товар), признаки — совокупность признаков товара и клиента, а целевая переменная — оценка товара клиентом. Для решения этой задачи регрессии можно использовать любые методы, например линейные или решающие деревья. Чтобы выполнить рекомендацию для клиента, мы выполним предсказания для всех пар (этот клиент, товар), то есть для всех товаров, и выберем те, для которых предсказано наибольшее значение. Описанный подход называется контентным (Content-based).

# Коллаборативная фильтрация

Другой подход называется основанным на памяти (Memory-based), его еще называют коллаборативной фильтрацией. Его идею мы уже упомянули выше: чтобы порекомендовать клиенту товары, алгоритм выделяет товары, которые клиент уже купил, выделяет клиентов, купивших эти товары и находит товары, которые покупали эти клиенты. Такой подход активно использовался до 2009 года, после чего ему на смену пришли подходы, основанные на скрытых переменных.



# Скрытые переменные

Подходы, основанные на скрытых переменных, работают так: они создают новые признаковые описания для клиентов и для товаров и оценивают нужность товара клиенту на основе этих признаков описаний. Про создание новых признаков представлений мы подробно поговорили в предыдущем разделе: в рекомендациях выполняется примерно то же самое, но только с акцентом на поиск товаров для клиентов, и называются эти признаки скрытыми переменными. Именно системы этой группы сегодня чаще всего используются на практике.



Онлайн-курс СберУниверситета

# Генеративное искусство

Подробнее о курсе



Бесплатный курс от Сбера по генеративному искусству

[https://courses.sberuniversity.ru/generative-art?utm\\_source=tg&utm\\_medium=organic&utm\\_campaign=courses&utm\\_content=gen\\_i&utm\\_term=01\\_09\\_2023](https://courses.sberuniversity.ru/generative-art?utm_source=tg&utm_medium=organic&utm_campaign=courses&utm_content=gen_i&utm_term=01_09_2023)