

# Достоверность ассоциативного правила (Association Rule Confidence)

Синонимы: Доверие

В теории ассоциативных правил достоверность — это показатель, характеризующий уверенность в том, что ассоциация  $A \rightarrow B$  является ассоциативным правилом. То есть предположение о том, что появление события  $A$  влечет за собой появление события  $B$ , является достаточно достоверным.

Ассоциация  $A \rightarrow B$  рассматривается как потенциальное ассоциативное правило, если для нее выполняется условие минимальной поддержки:

$$S(A \rightarrow B) = \frac{N(A \cup B)}{T}.$$

Т.е. поддержка равна числу транзакций  $T$ , которые содержат оба элемента  $A$  и  $B$  к общему числу транзакций в базе данных. Однако, само по себе высокое значение поддержки еще не говорит о том, что  $A \rightarrow B$  является правилом, поскольку не учитывает сколько раз условие и следствие появляются независимо друг от друга.

Например, пусть ассоциация «Если хлеб, то сигареты» имеет поддержку  $S = 0.5$  (т.е. из 100 покупок в 50 хлеб и сигареты появляются совместно). Хлеб потребляют практически все, курят относительно немногие. Таким образом, сигареты и хлеб покупают только курящие, хлеб покупают все. Следовательно, можно ожидать, что в остальных 50 транзакциях появляется хлеб без сигарет (хотя, возможно, с какими-то другими продуктами).

Данный факт обесценивает сравнительно высокую поддержку ассоциации «Если хлеб, то сигареты». Чтобы учесть независимое появление условия и следствия, используется еще один показатель — достоверность, которая равна поддержке условия и следствия, к поддержке только условия:

$$C(A \rightarrow B) = \frac{N(A \cup B)}{S(A)}.$$

Если предположить, что в наборе из 100 транзакций сигареты и хлеб появляются совместно 50 раз, а хлеб всего 90 раз (как с сигаретами, так и без), то достоверность будет:

$$C(A \rightarrow B) = \frac{0.5}{0.9} = 0.55.$$

Оказывается, что хлеб появляется по отдельности почти так же часто, как и совместно с сигаретами. Таким образом, достоверность того, что  $A \rightarrow B$  является правилом всего 55%, что явно не достаточно. Хорошая достоверность для правила составляет 0.7-0.9.

Если бы сигареты и хлеб всегда покупались только вместе, то достоверность была бы  $0.5/0.5=1$ , что говорило бы об абсолютной достоверности правила.

В LogiTom существует специализированный обработчик ассоциативные правила, который выявляет ассоциативные правила в транзакционных данных. Подробнее в статье «Введение в анализ ассоциативных правил».