

Последовательный шаблон (Sequence pattern)

Последовательные шаблоны — одно из направлений [Data Mining](#), в котором решается задача обнаружения значимых связей между событиями, происходящими последовательно.

Последовательный шаблон можно определить как конструкцию вида:

$$A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n, t(A_i) \leq t(A_j), i \leq j,$$

где A — событие, или множество событий, t — время. Таким образом, последовательный шаблон это упорядоченная во времени последовательность событий, наблюдаемая настолько часто, что это позволяет говорить о наличии между ними связей.

Анализ таких связей позволяет обнаруживать правила вида: если в наблюдении появился набор одних событий из шаблона, то это с высокой долей вероятности повлечет за собой появление других событий или события из этого же шаблона. При этом события первой группы называются **основными**, а события, появление которых ожидается — **целевыми**. Основным применением последовательных шаблонов является исследование типичных моделей поведения клиентов.

Теория последовательных шаблонов происходит из теории [ассоциативных правил](#). Методы анализа ассоциативных правил и последовательных шаблонов во многом похожи: и в том, и в другом случае используются такие понятия как предметный набор и [транзакция](#), такие числовые характеристики, как [поддержка](#) и [достоверность](#), а для обнаружения частых шаблонов применяются различные модификации [алгоритма Apriori](#). Однако, между ассоциативными правилами и последовательными шаблонами есть принципиальное различие.

В ассоциативных правилах представляет интерес факт совместного появления предметов в [транзакции](#) и не рассматривается порядок их появления. Иными словами, важно, что товар A и товар B были приобретены по одному чеку, т.е. появились в одной транзакции. В последовательных шаблонах, напротив, последовательность событий играет решающую роль, поскольку считается, что предыдущие события влияют на вероятность появления последующих.

Задача поиска последовательных шаблонов была впервые решена [Р. Агравалом](#) и [Р.Срикнатом](#), авторами популярного алгоритма поиска ассоциативных правил Apriori. Они предложили 3 алгоритма для решения задачи открытия последовательных шаблонов на больших массивах данных — GSP, AprioriSome и AprioriAll.

Типичным приложением последовательных шаблонов является предсказание того, будет ли востребован на рынке новый продукт компании (например, новый тарифный план) на основе наблюдаемой динамики потребления старых продуктов (изменения времени разговоров по старым тарифным планам). В этом случае основными событиями могут быть рост, спад или неизменность уровня потребления за последние четыре недели, а целевым событием — приобретение или отказ от приобретения нового продукта.

Например, возможна последовательность:

Рост ⇒ Спад ⇒ Спад ⇒ Неизменно ⇒ Отказ,

где первые четыре события являются **основными**, а последнее — **целевым**.

Анализ большого числа таких последовательностей позволит определить типичный характер поведения клиентов в отношении новых продуктов компании.