

Анализ отклонений (Deviation detection)

Синонимы: Определение выбросов

Разделы: [Бизнес-задачи](#)

Анализ отклонений — это поиск в данных редких, нетипичных объектов или наблюдений, которые не соответствуют логике поведения анализируемого [бизнес-процесса](#) или модели описывающих его данных.

Задачей анализа отклонений является их обнаружение и генерация формальных правил, объясняющих причины появления отклонений.

Отклонения в числовых данных можно интерпретировать как векторы в многомерном пространстве, расположенные на значительном удалении от основного «облака» данных по одному или нескольким [признакам](#). В [деревьях решений](#), например, отклонениями можно считать наблюдения, попавшие в узлы, содержащие всего несколько объектов. В [ассоциативных правилах](#) в качестве отклонений можно рассматривать ассоциации с очень низкой [поддержкой](#) (т.е. содержащие объекты крайней редко появляющиеся совместно).

Таким образом, поиск отклонений можно производить с помощью аналитических моделей, отказавшись от [обобщающей способности](#), чтобы модель обнаруживала не только наиболее общие зависимости и закономерности, но и частные, редкие.

Следует четко понимать разницу между отклонением и [аномальным значением](#). Отклонение — это нетипичный, редкий объект, наблюдение или шаблон. Аномальное значение (выброс) — это просто значение, сильно отличающееся от других значений одного и того же признака.

Т.е. аномальное значение всегда ищется путем сопоставления значений внутри столбца таблицы, а отклонения — путем сопоставления всех наблюдений набора данных.

Появление в наблюдении аномального значения (выброса) не делает автоматически данное наблюдение отклонением. Поэтому анализ отклонений (deviation analysis) и обнаружение аномальных значений (outlier detection) — это две разные задачи [Data Mining](#). Первая ищет отклонения и пытается их объяснить. Вторая больше связана с [очисткой данных](#).

Возможны два случая:

1. Отклонение является «информационным мусором», т.е. появилось вследствие ошибки, технического сбоя и т.д. Например, цена товара в 10 раз отличается от его

аналогов. Если это ошибка (добавили лишний ноль на ценнике), то объект не является отклонением, а является типичным объектом с неверно указанным признаком. В этом случае следует просто произвести очистку аномального значения и продолжить обработку наблюдения в обычном порядке.

2. Отклонение является объектом, имеющим свойства, значительно отличающие его от других объектов. Например, отличие цены на товар от аналогов является не ошибкой, а следствием какого-то эксклюзивного исполнения товара, придающего ему принципиально новые потребительские свойства. В этом случае задача заключается в обнаружении правила, которое объясняет эти свойства (например, «если часы являются противоударными и влагонепроницаемыми, то стоят в 10 раз больше обычных»).

К задачам, в которых анализ отклонений является важной частью решения, можно отнести:

- обнаружение мошенничеств с кредитными картами и страховками (фрод);
- обнаружение вирусных атак на основе нетипичной сетевой активности;
- выявление эксклюзивных клиентов, которым можно предложить обслуживание на особых условиях, и др.

В Logipom обработчик редактирование выбросов предназначен для автоматической корректировки значений набора данных, существенно отклоняющихся от средних.