

# СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ ДЛЯ ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА

Сергей П. Д.

Кафедра программного обеспечения информационных технологий,  
Белорусский государственный университет информатики и радиоэлектроники  
Минск, Республика Беларусь  
E-mail: polina.sergey.d@gmail.com

*В данной работе рассматриваются статистические методы обнаружения мошенничества в банковской сфере, а именно неконтролируемые методы, также приводятся примеры некоторых известных способов анализа.*

## ВВЕДЕНИЕ

Статистические методы обнаружения мошенничества основаны на сравнении наблюдаемых данных и ожидаемых значений [1]. Цель статистического анализа заключается в том, чтобы в результате получить оценку подозрительности: чем выше оценка, тем больше вероятность того, что транзакция является мошеннической. Различают контролируемые и неконтролируемые методы обнаружения мошенничества. Контролируемые методы основаны обучении выявлению мошеннических и не мошеннических (законных) транзакций на обучающих наборах исходных данных, где заранее известно к какому классу - мошенническому или законному - принадлежит транзакция. Используя контролируемые методы, надо быть уверенным в достоверности обучающих данных. Проблемой для обучения является несбалансированность размеров классов, так как мошеннических транзакций значительно меньше законных. Вдобавок контролируемые методы можно использовать только для обнаружения тех типов мошенничеств, которые уже были замечены ранее. Неконтролируемые методы обнаружения мошенничества применяются при отсутствии обучающих наборов законных и мошеннических транзакций. Данные методы обычно являются сочетанием методов профилирования и отслеживания выбросов. Сначала моделируется базовое распределение, которое представляет нормальное поведение, а затем обнаруживается поведение, которое сильнее всего отклоняется от нормы.

## I. АНАЛИЗ ОДНОРАНГОВЫХ ГРУПП (PEER GROUP ANALYSIS)

Метод анализа одноранговых групп отслеживает объекты, которые ведут себя иначе, чем объекты, с которыми у них раньше было схожее поведение [2]. Каждый объект рассматривается отдельно и сравнивается с остальными объектами с использованием критериев сравнения. На основе сравнения выбирается одноранговая группа, которая больше всего схожа с рассматриваемым объектом. Далее поведение изучаемого объекта сравнивается с поведением его одно-

ранговой группы, те объекты, поведение которые наиболее сильно выделяется на фоне поведения группы, получают большую оценку подозрительности. На рисунке 1 представлен пример выброса при использовании метода анализа одноранговых групп. Черной линией обозначена сумма денежных расходов рассматриваемого объекта за неделю, оранжевыми линиями - расходы одноранговой группы, голубыми - расходы остальных объектов. Расходы изучаемого объекта резко увеличились по сравнению с расходами объектов одноранговой группы, что говорит о возможном мошенничестве и требует дальнейшего изучения данного случая.

## II. АНАЛИЗ ТОЧКИ ОСТАНОВА (BREAK POINT ANALYSIS)

Анализ точки останова - это метод обнаружения неконтролируемых выбросов на основе информации о транзакциях на одном банковском счете [3]. Точка останова - это наблюдение, которое считается аномальными для анализируемого счета. В анализе сравнивается последовательность транзакций, например, их сумма или частота в пределах движущегося окна транзакций фиксированной длины. Перед началом анализа необходимо установить длину окна и соотношение параметров старых и новых транзакций. Когда осуществляется новая транзакция, она попадает в окно, а самая старая транзакция в окне удаляется. Затем самые поздние транзакции из окна сравниваются с более ранними транзакциями из окна, чтобы понять, следуют ли они модели поведения ранних транзакций. На рисунке 2 показан пример точки останова. Черными вертикальными линиями обозначено окно наблюдений, транзакции до стрелки - транзакции из локальной модели, транзакции после стрелки отображают аномальное поведение для рассматриваемой модели. Преимуществом данного метода является то, что для него не требуются сбалансированные данные, так как рассматривается только транзакции одного счета и не требуется обобщать данные в фиксированные моменты времени, как в случае с методом анализа одноранговых групп.

## ЗАКЛЮЧЕНИЕ

Использование исключительно контролируемых методов для обнаружения мошеннической транзакции может быть неэффективно, так как они предотвращают только известные схемы мошенничества. Для увеличения надежности системы предотвращения мошенничества целесообразно применять также неконтролируемые методы, которые могут отследить аномалии в поведении объектов и тем самым помочь выявить новые виды мошенничества.

## СПИСОК ЛИТЕРАТУРЫ

1. Statistical fraud detection: a review / Bolton R. J, Hand D. J. // Statistical Science, vol. 17, no. 3, 2002, p. 235–255.
2. Peer Group Analysis / Bolton R.J. and Hand D.J. // Technical Report, Department of Mathematics, Imperial College, London, 2001, 15 p.
3. Unsupervised profiling methods for fraud detection / Bolton R.J. and Hand D.J. // Conference on Credit Scoring and Credit Control 7, Edinburgh, UK, 2001, 16 p.

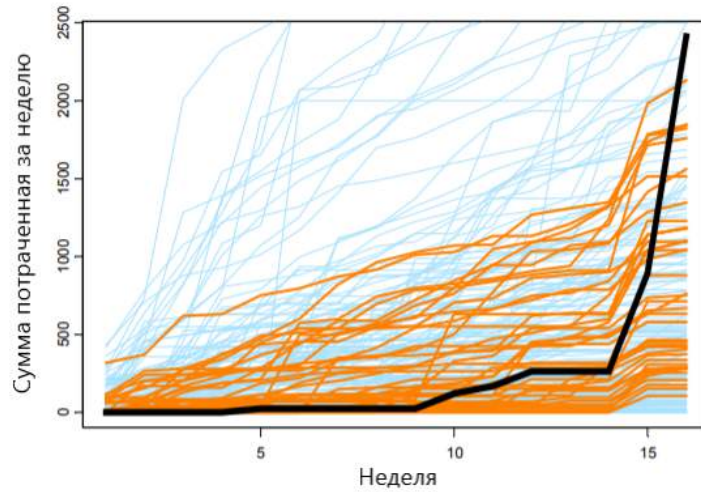


Рис. 1 – Пример выброса при использовании метода анализа одноранговых групп

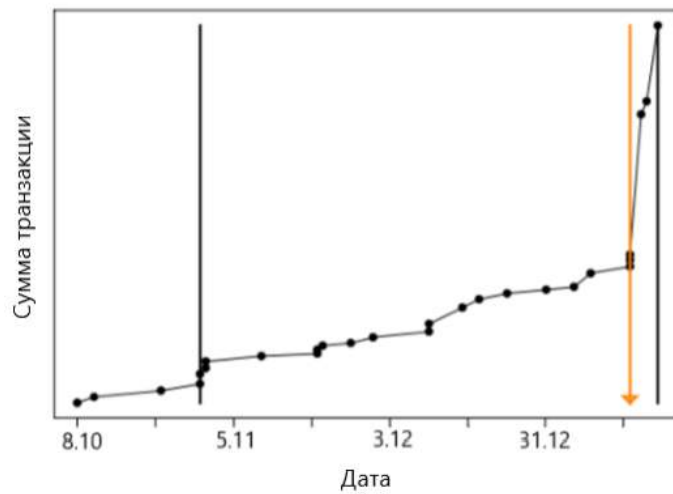


Рис. 2 – Пример точки останова