УДК 004.93(045)

*ТЕЛЕНИК С.Ф.,*
*МАЛЮКОВ П.Н.,*
*ПИЩАЕВА Н.Н.*

# НЕЙРОСЕТЕВОЙ МЕТОД КЛАССИФИКАЦИИ АНОМАЛИЙ В ТРАФИКЕ ПРОВАЙДЕРА МОБИЛЬНОЙ СВЯЗИ

Рассмотрены популярные нейронные сети для классификации фродов в трафике провайдера. Проанализированы алгоритмы для определения лучшего варианта. При помощи разработанных программ проведены исследования для сравнения быстродействия и качества классификации.

Popular neural networks for classification of fraud in the traffic of the operators are considered. Algorithms for definition of the best variant are analyzed. By means of the developed programs researches for comparison of speed and quality of classification are carried out.

## Введение

По мере увеличения масштабов и сложности современных сетей связи растет, и число попыток использовать широко разветвленную телекоммуникационную инфраструктуру в противоправных целях. Несанкционированные, сомнительные и откровенно мошеннические действия приводят не только к ощутимым потерям в доходах операторов, но и к снижению объемов и качества предоставляемых услуг.

Цепочка формирования доходов провайдера связи начинается с сетевого оборудования, обеспечивающего собственно предоставление услуги связи и фиксирующего это событие. Затем эта информация проходит через разнообразные промежуточные программные и аппаратные комплексы, агрегирующие и трансформирующие данные. Завершается все биллинговой системой, которая тоже состоит из нескольких модулей, ответственных за тарификацию услуг, прием платежей, выставление счетов и т.д. К этому следует добавить систему обслуживания клиентов, отчетности, управление услугами и т.п., то видно, насколько сложна дорога от предоставления телекоммуникационных услуг до получения за нее денег оператором. На каждом участке этого пути информацию поджидают сбои, конфликты оборудования, ошибки в настройках систем и, конечно же, воровство или фрод (мошенничество в телекоммуникационной сфере).

Задача противодействия мошенничеству и гарантирования доходов актуальна для телекоммуникационных компаний и сегодня эта тема выходит на новый уровень значимости по нескольким причинам:

– мировой фінансово-экономический кризис, выдвигающий особенно жесткие требования к эффективности ведения бизнеса;

– насыщения рынка телекоммуникационных услуг и обострение конкуренции;

– новая волна государственных регуляторов к тому, как компании осуществляют мониторинг и отчетность своей выручки;

– стремительное развитие технологий и маркетинговых инноваций, за которыми развитие методов и инструментов контроля просто не успевают.

## Анализ проблемы

Одной из актуальных задач, стоящих перед компаниями – операторами связи является минимизация финансовых потерь, которые могут бать следствием мошеннических действий абонентов. Под воздействием фрода падает эффективность использования сетевых ресурсов, страдает имидж компании, уходят ценные клиенты, замедляется продвижение новых телекоммуникационных услуг. Основные услуги сотовой связи включают передачу голосового трафика, передачу факсимильных, коротких и мультимедийных сообщений, передачу данных, предоставление всевозможных контентных сервисов, доступ к сети интернет, мобильное телевидение и.т.п. К дополнительным услугам можно отнести переадресацию вызова, сохранение вызова и ожидание вызова, услугу конференц – связи позволяющей вести разговор по телефону одновременно несколькими абонен-

тами, услугу запрета (или ограничения) автоматического определения вызывающего абонента, голосовую почту, роуминг и т. д.[1]

Основные типы потерь от мошенничества в телекоммуникационных сетях таковы:

– отсутствие корректной полной информации о предоставленных услугах телекоммуникаций;

– неверная обработка такой информации;

– использование технических и логических ошибок в работе билинговой системы;

– несобранная выручка (задолжности абонентов);

– недополученная прибыль (оператор телекоммуникаций получил меньше, чем мог бы). Оператору необходима классификация видов мошенничества, которая помогла бы ему упорядочить деятельность по борьбе с фродом в своей сети. Классификация фрода в телекоммуникациях чрезвычайно затруднена, поскольку мошенники постоянно совершенствуют свои навыки и способы отъема денег у операторов и клиентов связи. Одним из критериев дифференциации видов фрода можно использовать направленность действий злоумышленников: в одних случаях пострадавшей стороной являются абоненты, в других – сами операторы. На сегодня Ассоциациями CFCA определено более 200 видов мошенничества в телекоммуникациях.

Виды мошенничества в сетях мобильной связи:

– клонирование SIM – карт мобильных телефонов (для организации междугородних переговорных пунктов, одновременного использования неограниченных тарифных планов рядом лиц, противоправных действий под видом истинного владельца SIM карты, если ее клонировали без ведома владельца);

– использование реквизитов подставных лиц при регистрации в сети (для заключения контракта на корпоративное обслуживание группы абонентов с кредитной формой оплаты) ;

– незаконное завершение соединения;

– создание «чужих» SIM – центров (в результате «свой» SIM – центр не получает информацию об отправленном сообщении, и оно не тарифицируется);

– роуминговый фрод (генерация трафика на дорогостоящие PRS направления из расчета на задержку своевременной тарификации роумера в билинге домашней сети);

– Bypass (нарушение порядка маршрутизации международного трафика в результате чего дорогостоящие роуминговые звонки тарифицируются оператором как его условно бесплатные внутресетевые);

– мошенничество со службами заказа билетов, оплаты мелких товаров ,услуг, перевод денег с банковских счетов при помощи SMS – транзакций и т.д.

Виды внутреннего мошенничества – мошенничество сотрудников самого оператора телекоммуникаций:

– умышленная активация неоплаченных сервисов;

– мошенничество при использовании предоплаченных услуг;

– ошибочная тарификация.

Для активного противодействия фроду необходимы регулярные сбор и анализ информации о действии злоумышленников, выявление тенденций, всесторонний анализ тарифных планов и маркетинговых ходов, прогнозирование новых схем мошенничества и создание активных способов защиты. Для борьбы с проявлениями мошенничества в телекоммуникациях применяются специализированные системы класса Fraud Management System (FMS), которые способны обнаруживать проявления мошенничества.

По мере развития систем сотовой мобильной связи развиваются методы и приемы фрода. С одной стороны, постоянно возникают новые услуги, новые технологии и, как следствие, новые виды фрода. С другой стороны, телекоммуникационное мошенничество, в отличие от традиционных угроз информационной безопасности, весьма «операторо специфично». Оно слишком сильно завязано на конкретные реализации тех или иных услуг определенного оператора, на его системы, его процессы и т.д. Помимо общих проблем фрода, у каждой телекоммуникационной компании будет свой специфический набор фродовых схем, присущих только ей. Поэтому при планировании своей деятельности, операторам связи предлагается воспользоваться топологизацией фродовых схем на основе методов их выявления. Такая классификация представляет собой законченный, ограниченный набор классов фрода. Каждая вновь возникающая, в том числе уникальная для данного оператора, фрод схема может быть отнесена к одному из этих классов. [2]

### Цель работы

Целью данной работы является исследование и реализация нейросетевой технологии классификации аномалий трафика провайдера, для повышения достоверности решения задачи обнаружения фрода и обеспечения ее работы в реальном времени.

### Постановка задачи

Обычно защита основана на анализе телефонии контрольных выборок абонентов и обнаружении разного рода закономерностей не характерных профилю эталонного абонента подобного тарифного плана, анализе подозрительных событий, например, сильный рост трафика абонента и т.д. Таким образом, речь идет о сканировании большого объема информации и выделении в нем определенного паттерна, с большой вероятностью соответствующему фроду. Это типичная задача классификации, для решения которой могут использоваться как классические методы [3], так и нейронные сети. Внедрение существующих систем противодействия мошенничеству – дело довольно сложное и дорогое. Требуется серьезная интеграция в инфраструктуру оператора и обработка больших объемов разнообразных данных из разных систем, которые у каждого оператора свои. Настройка этого процесса требует привлечения специалистов высокой квалификации. Может оказаться и так, что внедрение дорогостоящей системы будет просто не нужно, так как наиболее приоритетный риск удастся закрыть другими способами с меньшими затратами. [4]

Проблема классификации (выделения отдельных групп) рассматривается в статистике как задача группировки исходных данных. Идея структурности реализована в кластерном анализе, основная цель которого – выделение в многомерных данных такие однородные подмножества, чтобы объекты внутри групп были близки (похожи) в известном смысле друг от друга, а объекты из разных групп – не похожи.

### Метод решения задачи классификации

Нейронные сети характеризуются нечетким представлением и переработкой информации, высоким параллелизмом обработки информации и распределенным хранением информации. В нашем исследовании мы ограничились однослойной нейронной сетью (персептроном**)** и многослойной нейронной сетью с прямыми и обратными связями.

В качестве функции активации $F$ , преобразующей взвешенную сумму входных сигналов S в выходной сигнал $y = F(S)$, использовались

– сигмоидная

$$y = \frac{1}{1 + e^{-cS}},$$

где $c>0$ – коэффициент ширины сигмоиды по оси абсцисс;

– гиперболический тангенс

$$y = th(cS) = \frac{e^{cS} - e^{-cS}}{e^{cS} + e^{-cS}}.$$

Взвешенная функция входных сигналов имеет вид

$$S = \sum_{j=0}^{n} w_j x_j ,$$

где $x_1..x_n$ – значения, поступающие на входы (синапсы) нейрона, $w_1..w_n$ – *веса* синапсов. [5]

В качестве среды проектирования был выбран продукт компании Sybase и язык высокого уровня C#. Отказ от использования указателей приводит к незначительному снижению производительности, но автоматический сбор мусора и расширенный функционал механизмов работы с коллекциями упрощают программную реализацию, делают программный код проще для поддержки и восприятия.

Программная часть состоит из библиотеки ClassificationLibrary.dll, реализующей методы классификации и графического интерфейса Classification.exe.

Основной класс Network для работы с односвязной и многослойной сетью прямого распространения (персептрон) содержит в себе методы, выполняющие следующие функции:

- добавления слоя с заданным количеством входов и выходов, задания функции активации с ее параметрами в качестве аргументов;
- вычисления выхода сети для объекта;
- возвращает номер нейрона с максимальным выходом;
- обучения многослойной сети методом обратного распространения. Функция принимает в качестве параметров список элементов, целевые значения выходов, количество шагов обучения и скорость обучения нейронной сети;
- создания и обучения однослойной нейронной сети. Функция использует дельта-правило.

С учетом выбранной структуры сети было принято решение отойти от принципов ООП и отказаться от ввода объекта для отдельного нейрона.

Работая с представлением слоя сети матрицей синаптических весов, мы получаем заметный прирост производительности за счет повышения скорости обращения к элементам матрицы в сравнении с поэтапным переходом к элементу в векторе снаптических весов.

Переход от общепринятой в нейронных сетях поэлементной функции активации к векторной позволяет реализовать слои с выходом типа «победитель забирает все», а также поднять производительность за счет уменьшения количества вызовов функции.

В процессе обучения сети необходимо вычислять производную от функции активации нейрона. Поскольку функция может принимать произвольную форму и вычисление ее производной по аналитической формуле может быть затруднительным, вместо производной используется первая конечная разность.

На основании этих решений был построен класс NetLayer, включающий методы: вычисления выхода слоя сети и вычисления первой конечной разности для последнего выхода сети. Свойства класса NetLayer содержат матрицу синаптических весов, копии последнего входного и выходного векторов. Сеть представляется списком объектов NetLayer, реализующих слой сети.

В процессе обучения нейронных сетей задавались обучающие выборки с различным количеством объектов, признаков и классов. Повторное обучение показало лучшие результаты, что свидетельствует о склонности алгоритма обучения «застревать» в локальных экстремумах целевой функции. Однослойная искусственная нейронная сеть показала себя достаточно хорошо на наборах с небольшим количеством классов.

Двухслойная сеть, обученная методом градиентного спуска, показывает схожие результаты. Среди отличий следует отметить значительно большее время обучения. Тестирование двухслойной сети показало ее высокую чувствительность к количеству нейронов скрытого слоя. Процент ошибки на однослойной сети не превысил 10% , для многослойной сети – 5%.

## Выводы

Сети показали свою высокую чувствительность к входным наборам данных, т.е. время обучения нейронной сети напрямую зависит от количества признаков, характеризующий трафик. Таким образом, для повышения качества классификации в дальнейшем необходимо рассмотреть вопрос сжатия пространства признаков.

## Список литературы


1. Попов В.И. Основы сотовой связи стандарта GSM. – М:Зко-Трендз,2005. – 296 с.
2. Шонин Д. Прагматическая классификация телекоммуникационного фрода, Jet Info, 11,2010 – С.10 -14
3. Малюков П.Н.Сравнительный анализ методов классификации аномалий в трафике провайдера мобильной связи /П.Н. Малюков, Н. Н. Пищаева / Електроніка та системи управління: сб. наук. праць – К.:НАУ, 2010. – Вип.2 (24). – С. 148 – 152
4. Шонин Д. Гарантирование доходов и противодействие мошенничеству в телекоммуникационных компаниях, Jet Info, 2,2010 – С.6 -11
5. Нейроинформатика / А.Н.Горбань, В.Л.Дунин-Барковский, А.Н.Кирдин и др. - Новосибирск: Наука. Сибирское предприятие РАН, 1998