

УДК 004.93'1

НЕЙРОСЕТЕВОЙ МЕТОД КЛАССИФИКАЦИИ НЕЛЕГИТИМНОГО ИСПОЛЬЗОВАНИЯ ТРАФИКА ОПЕРАТОРА МОБИЛЬНОЙ СВЯЗИ

Теленик С.Ф., Малюков П.Н., Оконский И.В., Перов Б.Г.

Национальный технический университет Украины «Киевский политехнический институт»
кафедра автоматизации и управления в технических системах
E-mail: Logrus.private@gmail.com

Аннотация

Теленик С.Ф., Малюков П.Н., Оконский И.В., Перов Б.Г. Нейросетевой метод классификации нелегитимного использования трафика оператора мобильной связи. Предлагается использование нейронной сети Кохонена в качестве инструмента классификации трафика провайдера телекоммуникационных услуг, представленного в виде временных рядов. Описана соответствующая программистская модель на основании кроссплатформенных алгоритмических библиотек с открытыми исходными кодами. Разработано программное решение, реализующее описанную программистскую модель, и выполнен анализ полученных результатов.

Введение

В области телекоммуникаций актуальной задачей является защита от фрода (несанкционированного или мошеннического использования сотовой связи). По оценкам экспертов потери в мире от фрода составляют 20-25 миллиардов долларов в год. Учитывая многообразие телекоммуникационных технологий, широкий спектр методов мошенничества и способов нелегитимного использования ресурсов операторов, невозможно создать универсальную систему, способную выявлять все возможные существующие методы мошенничества и методы, которые могут появиться в будущем. Соответственно, невозможно пресекать злоупотребления новыми способами, появляющимися как реакция на рекламные акции, проводимые операторами [1]. В виду этого возникает необходимость разработки моделей, методов и алгоритмов анализа использования коммуникационных ресурсов абонентами мобильной связи для создания автоматизированных средств выявления случаев нелегитимного использования оборудования и каналов связи. Эти модели могут применяться ИТ-персоналом, занимающимся вопросами сохранения доходности бизнеса оператора мобильной связи.

Анализ проблемы

Классификация фродов в телекоммуникациях чрезвычайно затруднена, поскольку мошенники постоянно совершенствуют свои навыки и способы отъема денег у операторов и клиентов связи. Оператору необходим некий способ классификации видов мошенничества, который помог бы ему упорядочить деятельность по борьбе с мошенничеством в своих сетях. Для контроля объемных и динамических процессов, которые имеют место в сетях операторов связи, необходимо использование быстродействующих алгоритмов и методов мониторинга и анализа трафика. Представление трафика в виде временного ряда позволяет использовать инструментарий прикладного анализа данных. При этом целесообразно отдельно рассматривать и анализировать ряды с различными аргументами, параметрами или показателями. Например, при рассмотрении трафика операторов сотовой связи, это могут быть ряды, отображающие загруженность каналов связи в фиксированные моменты времени, активность отдельных абонентов во времени, количество и продолжительность звонков, проходящих через конкретную базовую станцию, активность абонентов отдельного тарифного плана, трафик роуминга и прочее.

Постановка задачи

В качестве объекта классификации рассмотрен трафик оператора мобильной связи, где исходными данными является профиль абонента, состоящий из наиболее значимых характеристик пребывания в сети.

Несанкционированный доступ, в таком случае, математически может быть представлен как резкое изменение определенного участка временного ряда, которому присущ определенный характер, например линейный или константный.

Таким образом, перед разработчиком стоит задача классификации временного ряда путем поиска в нем определенных зависимостей, задаваемых оператором. Это типичная задача классификации, для решения которой могут использоваться как классические, так и нейросетевые методы.

Целью данной статьи является разработка инструмента для анализа трафика оператора на основе нейронных сетей, который мог бы выполнять классификацию аномалий во временных рядах, что в дальнейшем может быть использовано для повышения достоверности решения задач обнаружения фродов.

Предлагаемый подход к выявлению нелегитимного использования каналов связи

Трафик провайдера представляет собой одномерный временной ряд, где временными отметками являются относительные отсчеты времени, например, в общепринятом формате timestamp [2], а значениями – величины определенной количественной характеристики сети передачи данных.

Для решения задач такого рода широко используются нейронные сети Кохонена [3], специализированные для решений задач классификации и кластеризации.

Обучающей выборкой для таких нейронных сетей являются задаваемые оператором предположительные законы изменения временного ряда, по которым можно с высокой степенью вероятности определить факт несанкционированного доступа к сети передачи данных. Исходными данными являются векторы, представляющие временные ряды описывающие трафик провайдера.

На основе обучающих векторов нейронная сеть Кохонена имеет возможность классифицировать входные вектора, относя их к одному из классов, которым она была обучена. Таким образом, выходом нейронной сети является класс обучающего вектора, к которому нейронная сеть отнесла входной вектор. Если нейронная сеть не смогла точно выполнить классификацию входного вектора, то она его аппроксимирует в сторону наиболее похожего обучающего вектора.

Реализация математической модели

Для программной реализации поставленной задачи был выбран язык программирования C++, поскольку он предоставляет широкий спектр возможностей для реализации различных алгоритмов за счет своей стандартной библиотеки шаблонов STL и дополнительных библиотек с открытыми исходными кодами, такими как boost и Qt.

Кроссплатформенная библиотека Qt предоставляет набор средств и инструментов для создания графического интерфейса пользователя. Кроссплатформенная алгоритмическая библиотека boost содержит множество реализаций базовых алгоритмов и решений типовых задач, при этом абстрагирует пользователя от способа хранения данных, используя подход, применяемый в STL.

Средой разработки была выбрана Microsoft Visual Studio 2010.

Реализация математической модели базируется на инструментах библиотеки boost. В виду этого использовался следующий набор решений.

В качестве структуры данных использовался стандартный контейнер STL vector, каждым элементом которого является vector, в свою очередь состоящий из объектов,

реализующих математическую модель нейронов сети Кохонена, так как такая комбинация стандартных контейнеров наиболее точно отображает внутреннюю структуру этой сети.

Каждый объект нейрона сети Кохонена инкапсулирует в себе алгоритмы обучения, подстройки весов и активационную функцию. Поскольку все эти алгоритмы являются базовыми алгоритмами boost, то класс нейрона Кохонена представляет собой специализированный шаблонный класс C++, шаблонные параметры которого – активационная функция и алгоритм подстройки весов.

Активационная функция и алгоритм подстройки весов, в свою очередь являются специализированными шаблонными классами базового шаблонного класса basic_function.

В качестве активационной функции использовалась упрощенная активационная функции Коши:

$$y(x) = \frac{1}{1 + \left(\frac{x}{\sigma}\right)^p}, \quad (1)$$

где x – это внутреннее значение нейрона, y – выходной сигнал нейрона, а параметры σ и p выбираются оператором в зависимости от конкретной задачи, которую реализует нейронная сеть. Например, для значений $\sigma = 1$ и $p = 1$ график данной активационной функции имеет вид (рис. 1):

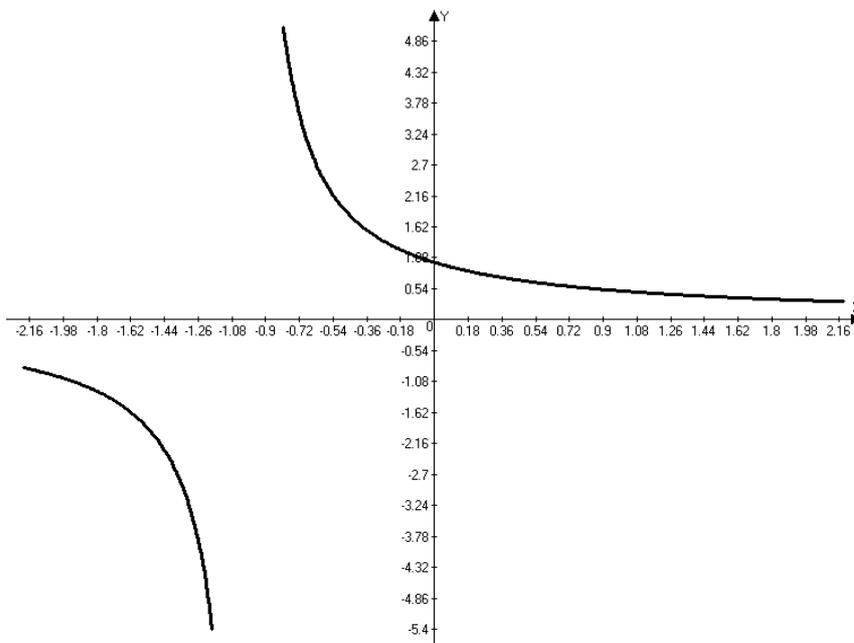


Рис. 1. График активационной функции при значениях параметров $\sigma = 1$, $p = 1$

Полученные результаты и выводы

Разработанное программное решение Qohonen имеет наглядный графический интерфейс и способно выполнять классификацию временных рядов, при этом предоставляя визуальное изображение в виде графиков исходного временного ряда и обучающего временного ряда, к классу которого он был отнесен или аппроксимирован.

Программа имеет возможность задания параметров нейронной сети и входных векторов для её обучения как с помощью автоматически генерируемых шаблонных функций,

параметры которых может задать пользователь, так и с помощью ручного ввода из текстового файла. Также программа обладает возможностью тестирования нейронной сети на эталонных образцах временных рядов с целью определения точности обучения.

На рис. 2 изображены факты обнаружения нейронной сетью константой и линейной, а на рис.3 – квадратичной и синусоидальной зависимостей в исходном временном ряду соответственно.

Тонкой линией дано графическое представление исходного вектора, жирной – графическое представление обучающего временного ряда того класса, к которому сеть отнесла входной вектор.

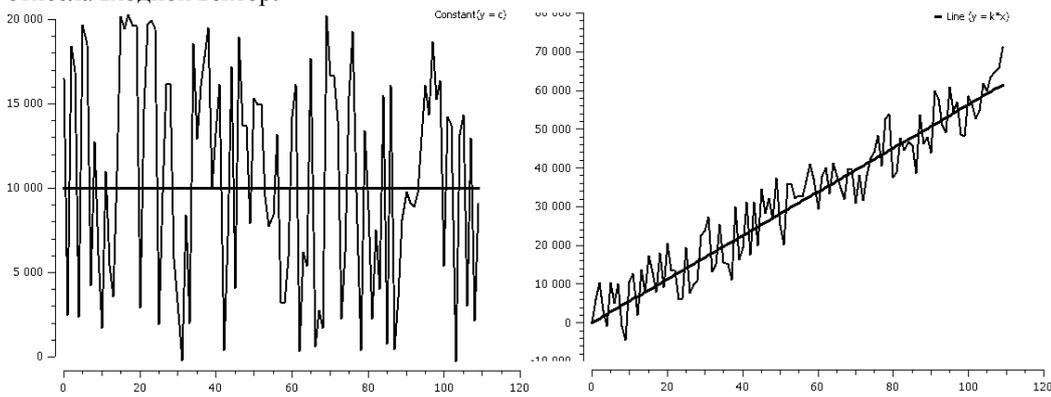


Рис. 2. Факты обнаружения нейронной сетью константной и линейной зависимостей в исходном временном ряду

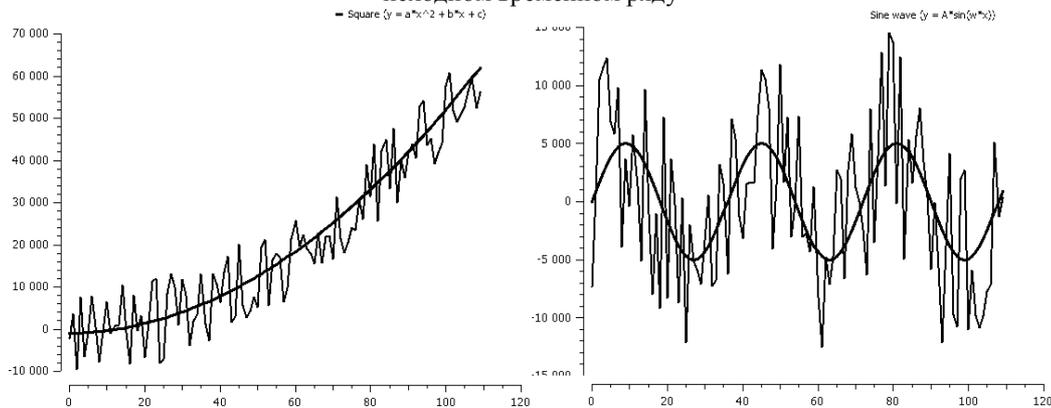


Рис. 3. Факты обнаружения нейронной сетью квадратичной и синусоидальной зависимостей в исходном временном ряду

Литература

1. Ратынский М. В. Основы сотовой связи / Под ред. Д. Б. Зимины – М.: Радио и связь, 1998. - 248 с.
2. Timestamp – Wikipedia, the free encyclopedia [Electronic resource] / Интернет-ресурс. - Режим доступа : www/ URL: <http://en.wikipedia.org/wiki/Timestamp>, свободный – Загл. с экрана.
3. Аксенов С.В., Новосельцев В.Б. Организация и использование нейронных сетей (методы и технологии) / Под общ. ред. В.Б. Новосельцева. – Томск: Изд-во НТЛ, 2006 – 128 с.