

Интеллектуальный анализ данных для прогнозирования состояний устройств телекоммуникационной сети

А. К. Канаев¹, Э. В. Логин², А. В. Соколова³, С. С. Долгушин⁴

Петербургский государственный университет путей сообщения Императора Александра I

¹kanaev@pgups.ru

²elinabeneta@yandex.ru

³sokol_2001_nastia@mail.ru

⁴syomasergeevich@yandex.ru

Аннотация. Постоянное совершенствование Единой системы мониторинга и администрирования сетей связи, заключающееся в возрастании числа входящих в нее устройств, требует внедрения средств, позволяющих осуществлять не только сбор и хранение непрерывно увеличивающегося числа данных, поступающих от различных элементов телекоммуникационной сети, но также высококачественную их обработку, анализ и прогнозирование. Это, в свою очередь, позволит получить предиктивную оценку состояний входящих в телекоммуникационную сеть устройств на основе предсказания дальнейшего изменения их параметров. Таким образом, появляется возможность повышения эффективности планирования обслуживающих и ремонтных мероприятий, что поможет своевременно реагировать на происходящие в элементах сети связи изменения, устраняя их, тем самым обеспечивая более надежную работу сетевых устройств.

Ключевые слова: прогнозирование; интеллектуальный анализ данных; телекоммуникационная сеть; авторегрессия; нейронная сеть; Data Science

I. МЕТОДЫ DATA SCIENCE

Методы Data Science позволяют реализовать интеллектуальную информационную систему, которая будет включать в себя следующие этапы работы с данными: подготовка, обработка, анализ и представление. Перечисленные этапы для наглядности отображены на рис. 1 с кратким описанием.



Рис. 1. Этапы интеллектуальной обработки данных

Интеллектуальная информационная система (ИИС) представляет собой систему, основанную на знаниях, то есть комплекс разнообразных средств, позволяющих осуществлять поддержку принятия решений, наподобие поиска информации в режиме продвинутого диалога на естественном языке.

Применение методов Data Science включает в себя работу с программными алгоритмами, аналитическими инструментами, обращение к искусственному интеллекту и другим технологиям [1].

Data Science – наука о данных, занимающаяся сбором больших массивов структурированных и неструктурированных данных, их анализом и преобразованием в необходимый для потребителя формат.

Этап представления данных в требуемом виде заключается в их визуализации, работе со статистикой и такими аналитическими методами, как машинное обучение и глубокое обучение, а также анализ вероятностей, построение предиктивных моделей и нейронных сетей, используемых в решении различных задач.

Data Science также находится в тесной взаимосвязи с Big Data, что позволяет осуществлять работу с большими объемами постоянно добавляющихся и обновляющихся данных. Таким образом, Data Science используется для ускорения и автоматизации процессов исследования с применением разных подходов, гарантируя при этом математически доказанную точность при принятии решений.

К методам Data Science относятся, например, такие как:

- линейная регрессия;
- авторегрессия (как частный вид линейной регрессии);
- логистическая регрессия;
- кластеризация с обучением без учителя;
- контролируемое обучение;

- нейронные сети;
- глубокое обучение;
- прогнозное моделирование и другие [2].

II. АНАЛИЗ ПАРАМЕТРОВ УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Телекоммуникационная сеть представляет собой огромную географически распределенную техническую систему, включающую в себя множество различных сетевых устройств, число которых постоянно пополняется. При этом сами устройства непрерывно совершенствуются, происходит постепенный переход к новым более совершенным технологиям, позволяющим повысить качество связи и скорость передачи данных.

В ходе данного исследования был проведен анализ существующих на данный момент устройств телекоммуникационной сети и их особенностей.

С целью структурирования и конкретизации полученных данных были выделены параметры различных устройств, характеризующие, например, для коммутатора скорость его портов и производительность устройства, внутреннюю пропускную способность, задержку фрейма во время передачи и прочие немаловажные параметры. Для других устройств телекоммуникационной сети также были составлены списки параметров с присвоением им математических идентификаторов и кратким описанием сущности каждого параметра.

Такой подход позволил детально оценить телекоммуникационную сеть и подойти к рассмотрению параметров ее элементов в качестве объекта моделирования.

III. ВЫБОР СПОСОБА МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПАРАМЕТРОВ

В качестве возможных вариантов для построения модели, способной осуществлять прогноз значений исследуемого параметра рассматриваются такие подходы, как:

- авторегрессия;
- нейронная сеть.

Авторегрессия является частным случаем линейной регрессии, но с использованием предшествующих значений переменной.

То есть авторегрессия представляет собой регрессионную зависимость значений некоторой случайной последовательности от предшествующих значений. При этом значения какого-либо параметра, полученного в данный момент времени (к примеру, сегодня) значительно коррелируют с тем, что было вчера, менее значительно коррелируют с тем, что было получено на прошлой неделе и слабо коррелируют с теми данными, которые наблюдались в прошлом месяце и т. д.

В связи с этим предшествующие значения переменной могут быть использованы для составления

прогноза ее будущих значений, предоставляя возможность получения предиктивной оценки каких-либо исследуемых параметров.

Для авторегрессии может использоваться следующее выражение:

$$y_t = c + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t$$

За указанный промежуток времени t переменная постоянна и равна c ; к ней прибавляется сумма взвешенных предшествующих значений, собранных p временных отрезков назад, и гауссовский шум (ошибка) ε_t .

Эта модель может быть обучена с помощью обычного метода наименьших квадратов, необходимо всего лишь выбрать параметр p , чтобы указать количество временных интервалов в прошлом, которое необходимо исследовать [3].

Нейронные сети способны осуществлять работу параллельно с входными переменными и, вследствие этого, достаточно просто могут обрабатывать большие объемы данных.

Основное преимущество нейронных сетей заключается в их возможности определять закономерности. Искусственные нейронные сети являются альтернативой в инструментарии специалистов, занимающихся прогнозированием.

Для решения данной задачи, имеет смысл рассмотреть метод регрессионного анализа, относящийся к числу вышеупомянутых методов Data Science. Конкретно для данной цели была построена модель авторегрессии, позволяющая получить прогноз на основе значений параметра, зафиксированных в предыдущих периодах.

Искусственные нейронные сети (ИНС) пригодны как для низкоуровневых задач распознавания, таких как распознавание речи или символов, так и для задач принятия решений, которые имеют весомое значение в настоящее время. В рамках телекоммуникационной сети внедрение систем поддержки принятия решений позволит помочь обслуживающему персоналу своевременно реагировать на изменение параметров устройств телекоммуникационной сети, тем самым поможет предотвращать некоторые возможные неисправности, путем предоставления для персонала понятных рекомендаций по осуществлению дальнейших действий.

Нейронные сети включают в себя множество обрабатывающих элементов, каждый из которых взаимодействует с другими посредством возбуждающих или тормозящих соединений. Распределённое представление поверх большого числа элементов вместе с взаимосвязью обрабатывающих элементов обеспечивает допустимую ошибку. Обучение достигается посредством правил, которые адаптируют веса связей в ответ на входные шаблоны [4].

IV. ПОСТРОЕНИЕ ПРОГНОЗА С ПРИМЕНЕНИЕМ АВТОРЕГРЕССИИ И НЕЙРОСЕТИ

Поскольку на данном этапе исследования стоит задача выбора способа моделирования для осуществления прогнозирования параметров различных устройств телекоммуникационной сети, в качестве моделируемого параметра, с целью упрощения процесса анализа, была принята уже известная совокупность значений параметра затухания в одномодовом оптическом волокне.

Такой выбор параметра обусловлен доступностью оборудования, используемого для снятия реальных значений указанного параметра. Само оптическое волокно является средой передачи сигнала, а значит частью телекоммуникационной сети, что позволяет на начальном этапе исследования допустить осуществление моделирования его параметров, как альтернативу моделированию параметров сетевых устройств. Целью же является именно проверка способности реализуемой модели выполнять прогноз значений параметров на основе уже имеющихся.

В качестве выборочной совокупности для анализа с помощью регрессионной модели был выбран диапазон измеренных значений параметра затухания за период в 12 недель.

Модель авторегрессии (рис. 2) при прогнозировании была установлена на предсказание значений исследуемого параметра относительно предыдущих измеренных значений, которые были смещены на один временной интервал относительно исходных. Результат модели подтвердил, что спрогнозированные значения максимально близки к изначальным, то есть подтверждается точность осуществляемого прогноза.

Модель нейронной сети (рис. 3) для решения задачи предсказания будущих параметров элементов была реализована автором на основе программного кода на языке Python с использованием требуемого набора тренировочных данных. После загрузки и выделения входных и выходных данных, программа выполняет процесс обучения модели. Обучение модели нейросети – это процесс настройки параметров и весов нейронной сети с помощью обучающих данных, чтобы она могла выполнять определенную задачу без явного программирования. Предсказанные выходные данные модель формирует с учётом сохранённых в результате тренировочных данных весов. При этом, если в тестовых данных есть значение, превышающее 40, программа выдает сообщение «Большие потери в одномодовом волокне».

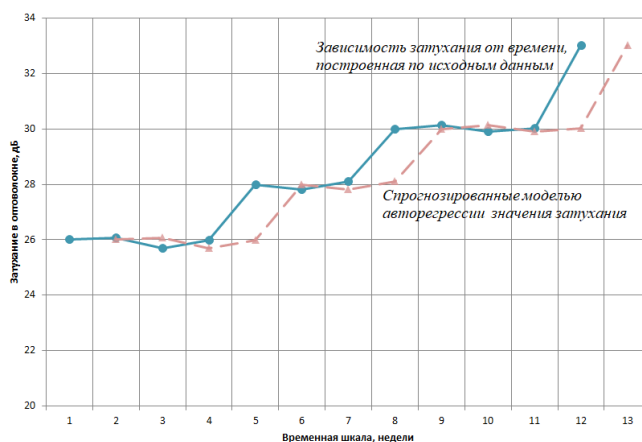


Рис. 2. Спрогнозированная моделью авторегрессии графическая зависимость затухания в оптическом волокне в сравнении с исходными данными

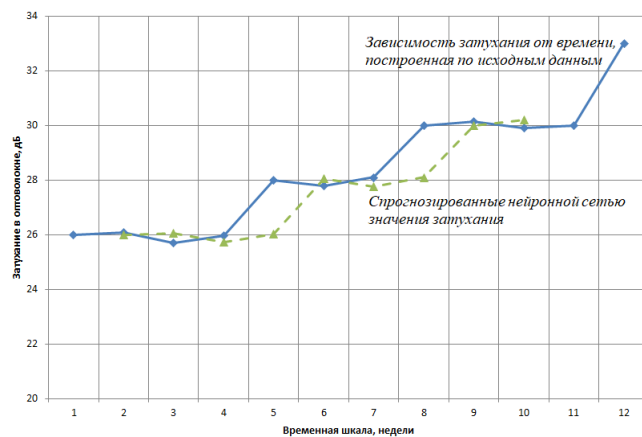


Рис. 3. Спрогнозированная нейронной сетью графическая зависимость затухания в оптическом волокне в сравнении с исходными данными

V. ЗАКЛЮЧЕНИЕ

В ходе проведенного исследования были изучены особенности и параметры некоторых устройств, применяемых в настоящее время в телекоммуникационной сети. В результате было выявлено, что следует рассмотреть направления совершенствования Единой системы мониторинга и администрирования сетей связи с помощью внедрения в нее средств, позволяющих осуществлять наряду со сбором и хранением данных, также их прогнозирование.

Для осуществления цели получения предиктивной оценки параметров устройств телекоммуникационной сети посредством процессов прогнозирования были предложены для рассмотрения два способа: авторегрессионная модель и нейронная сеть.

Оба способа были исследованы экспериментальным путем. Построение модели авторегрессии позволило получить прогноз значений на основе данных, полученных в предыдущие периоды наблюдений.

Графическая зависимость прогноза исследуемого параметра затухания в оптическом волокне показала максимальную приближенность данных, полученных в результате прогноза, к исходным данным. Этот факт свидетельствует о точности построенного прогноза.

При сравнении графической зависимости, отражающей предиктивную оценку значений исследуемого параметра, полученной в результате применения нейронной сети с результатом модели авторегрессии были выявлены следующие особенности:

- оба метода показывают приблизительно одинаковое запаздывание при прогнозировании резких изменений параметра;
- нейронная сеть с увеличением объёма данных (анализируемого временного ряда) показывает увеличение точности прогноза за счёт свойств обучения.

Соответственно, следует вывод о том, что данный метод также подходит для построения прогноза.

Важно отметить, что авторегрессионная модель в целом уступает нейронной сети вследствие того, что она нуждается в постоянном контроле, в том числе в плане интерпретации получаемых результатов.

Для того чтобы добиться существенного преимущества прогноза, получаемого с помощью нейронной сети, необходимо вводить как можно большее количество обучающих данных. В рамках данной работы, в связи с ограниченностью диапазона имеющихся исходных данных, полноценное обучение нейронной сети затруднительно, по этой причине результаты прогноза, полученного с помощью модели авторегрессии и с помощью нейронной сети приблизительно схожи.

Тем не менее, применение нейронных сетей несет в себе очевидное преимущество в случае, когда обучающих данных достаточно для корректного определения нейросетью взаимосвязей между исследуемыми параметрами. В рамках генерируемых устройствами телекоммуникационной сети данных, это условие соблюдается. Таким образом, имеет смысл рассмотреть возможности внедрения в состав Единой системы мониторинга и администрирования сетей связи средств, позволяющих дать предиктивную оценку значений параметров различных устройств.

Предлагаемый метод позволит не только прогнозировать изменения параметров и, соответственно, возможные сбои в работе оборудования сети связи, но также и перейти к внедрению систем поддержки принятия решений, которые на основе предсказанных параметров будут способны оценивать состояние элементов телекоммуникационной сети и выдавать рекомендации по осуществлению дальнейших действий персоналу по обслуживанию и ремонту устройств связи.

СПИСОК ЛИТЕРАТУРЫ

- [1] Что такое Data Science – База знаний Timeweb Community// timeweb.com [Электронный ресурс]: <https://timeweb.com/ru/community/articles/data-science>
- [2] Топ-45 методов data science// datascientist.one [Электронный ресурс]: <http://datascientist.one/top-45-data-science-methods/>
- [3] Что такое «анализ временных рядов»?// www.megaputer.com [Электронный ресурс]: <https://www.megaputer.com/ru/what-is-time-series-analysis/>
- [4] Нейронные сети и регрессионный анализ как метод прогнозирования временных рядов// cyberleninka.ru [Электронный ресурс]: <https://cyberleninka.ru/article/n/neyronnye-seti-i-regressionnyy-analiz-kak-metod-prognozirovaniya-vremennyh-ryadov?ysclid=lf48s4qinc501744262>