

О. И. КУТУЗОВ, Т. М. ТАТАРНИКОВА
 Санкт-Петербургский государственный электротехнический университет
 «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург

К ПРОГНОЗИРОВАНИЮ СЕТЕВОГО ТРАФИКА

Обсуждается проблема увеличения объемов сетевого трафика, приводящая к возникновению перегрузок каналов передачи данных. Приводятся актуальные модели прогнозирования, их достоинства и недостатки. Показано, что результаты прогноза зависят от качества ретроспективных данных, на которых строится модель прогноза. Рассматриваются основные техники предобработки ретроспективных данных. Реализована компьютерная модель прогноза сетевого трафика – интегрированная модель авторегрессии и скользящего среднего. Ошибка прогноза не превышает 0,04 как для краткосрочного, так и для среднесрочного и долгосрочного периодов.

Введение. За последнее десятилетие произошло сильное изменение объемов сетевого трафика. С каждым годом к интернету присоединяются новые смартфоны, планшеты, устройства интернет вещей. Так же доступность видеохостингов и стриминговых сервисов значительно увеличивает количество передаваемого трафика. По данным сайта datareportal.com число интернет-пользователей увеличивается в среднем на 8,2 % в год. По сравнению с этим, население мира увеличивается в среднем всего на 1,05 % в год [1].

Непрерывный рост трафика приводит к возникновению перегрузки каналов передачи данных. Перегрузки приводят к снижению качества звука и изображения, к увеличению количество потерянных пакетов данных. В связи с этим основной задачей сетевых компаний становится эффективное управление трафиком, приводящее к сбалансированной нагрузке на сеть [2]. Одним из вариантов решения задачи управления трафиком инфокоммуникационных сетей является прогнозная оценка загрузки линий связи. Для прогноза используются статистические данные, которые собираются в процессе эксплуатации коммуникационного оборудования. Модель, способная с заданной точностью прогнозировать трафик как в краткосрочном, так и в долгосрочном промежутке времени позволит заранее выделять необходимые ресурсы для поддержания качества предоставляемых услуг [3].

Анализ существующих моделей прогнозирования. Существует два основных вида моделей прогнозирования: статистические и структурные. Для статистические моделей функциональная зависимость задается аналитически [4]. К таким моделям относятся: регрессионные, авторегрессионные и экспоненциального сглаживания. Структурные модели основаны на зависимости структур. К ним относятся нейронные сети, модели на базе цепей Маркова и деревья принятия решений [5]. В табл. 1 приведены достоинства и недостатки вышеперечисленных моделей прогнозирования.

Таблица 1

Модели прогнозирования		
Модель	Достоинства	Недостатки
Регрессионные	Простота в моделировании и проектировании	Сложность нахождения оптимальных коэффициентов и функциональной зависимости
Авторегрессионные	Простота в моделировании и анализа	Нельзя моделировать нелинейные процессы
Экспоненциальное сглаживание	Простота моделирования	Узкая применимость модели
Нейросетевые	Большое разнообразие архитектур; нелинейность	Сложность в выборе архитектур; размер обучающей выборки; большие временные и ресурсоемкие затраты на обучение
Цепи Маркова	Единообразие проектирования	Узкая применимость моделей; нельзя моделировать долгосрочные процессы
Классификационно-регрессионные деревья	Простота обучения модели; возможность масштабирования	Сложность построения алгоритма дерева

Оценка качества ретроспективных данных. Наличие выбросов и искажений является причиной появления ложных закономерностей и тенденций, что приводит к ошибке в прогнозировании. Для исключения таких факторов проверка ретроспективных (накопленных) данных проводится на всех этапах работы с ними.

В табл. 2 приведены существующие уровни качества ретроспективных данных.

Т а б л и ц а 2

Уровни качества ретроспективных данных		
Уровень	Решаемые проблемы	Недостатки
Технический	<ul style="list-style-type: none"> • нарушение в структуре данных; • некорректное наименование таблиц и полей; • некорректные форматы и кодировки; • нарушение полноты и целостности данных. 	Мешают выполнению консолидации и интегрированию данных, применению к ним алгоритмов обработки
Аналитический	<ul style="list-style-type: none"> • пропуски; • аномальные значения; • противоречия; • дубликаты 	Снижают достоверность данных и искажают результаты их анализа, мешают обнаружению закономерностей и тенденций
Концептуальный	Собранные и консолидированные данные в недостаточной мере отражают исследуемый процесс	Отсутствие или недостаток данных для анализа

Выбросы, пропуски и структурные нарушения можно обнаружить используя табличное представление данных, графики и диаграммы.

Предобработка данных может включать в себя два направления: очистку и оптимизацию.

Очистка данных – процесс, который исключает большинство факторов, которые снижают качество данных и мешающие дальнейшей корректной работе аналитической модели. Данный процесс включает в себя обработку дубликатов, фиктивных значений, выбросов, заполнение пропусков.

Оптимизация данных обеспечивает снижение размерности, выявление и исключение незначительных признаков, что адаптирует данные для конкретной задачи и повышает эффективность анализа.

Главное отличие очистки и оптимизации данных заключается в том, что очистка исключает факторы, которые существенно снижают точность решения задачи, в отдельных случаях делают работу аналитического алгоритма невозможным.

Модель ARIMA. Авторегрессионные модели на данный момент являются лучшим инструментом прогнозирования временных рядов [6]. В работе применяется модель $ARIMA(p,d,q)$ – интегрированная модель авторегрессии и скользящего среднего, где p – порядок авторегрессии, d – порядок интегрированного ряда, q – порядок скользящего среднего. Блок-схема алгоритма ARIMA приведена на рисунке.

В блоках 2–4 рисунка происходит получение стационарного ряда и коэффициента d ; блок 5 подразумевает тестирование автокорреляционной функции и частной автокорреляционной функции для получения коэффициентов p и q ; в блоке 6 оцениваются параметры модели; в 7 блоке происходит предобработка временного ряда, включая очистку и оптимизацию; блок 8 включает использование модели для прогнозирования [7].

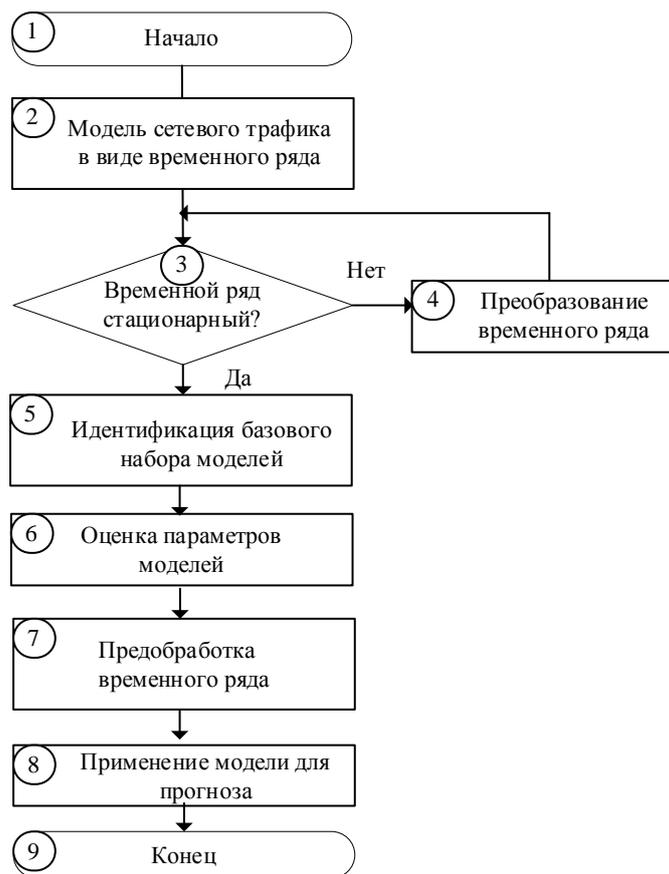


Рисунок. Блок-схема модели ARIMA

В работе получена ошибка прогноза 0,04, которая остается стабильной и на длинных трассах трафика не меняется при среднесрочном и долгосрочном прогнозах.

Заключение. Модель ARIMA конкурирует с методами машинного обучения по точности прогноза, но проигрывает по времени отклика – получении результатов прогноза. Это связано с необходимостью применения автокорреляционного анализа для выбора параметров модели ARIMA.

ЛИТЕРАТУРА

1. **Пойманова Е.Д., Татарникова Т.М., Краева Е.В.** Модель управления хранением трафика данных. *Известия высших учебных заведений. Приборостроение*. 2021. Т. 64. № 5. С. 370-375.
2. **Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V.** Model and Interaction Efficiency of Computer Nodes Based on Transfer Reservation at Multipath Routing. *2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECNF)*. 2019. P. 8840647. doi: 10.1109/WECNF.2019.8840647
3. **Советов Б.Я., Татарникова Т.М., Пойманова Е.Д.** Организация многоуровневого хранения данных. *Информационно-управляющие системы*, 2019, № 2, с. 68–75. doi:10.31799/1684-8853-2019-2-68-75.
4. **Татарникова Т.М.** Статистические методы исследования сетевого трафика. *Информационно-управляющие системы*. 2018. № 5 (96). С. 35-43.
5. **Tatarnikova T.M., Roymanova E.D.** Differentiated capacity extension method for system of data storage with multilevel structure. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2020. Vol. 20. No.1. P. 66–73. Doi: 10.17586/2226-1494-2020-20-1-66-73.
6. **Татарникова Т.М., Вольский А.В.** Оценка вероятностно-временных характеристик сетевых узлов с дифференциацией трафика. *Информационно-управляющие системы*. 2018. № 3 (94). С. 54-60.
7. **Lorido-Botran Tania, Miguel-Alonso Jose, Lozano Jose A.** A review of auto-scaling techniques for elastic applications in cloud environments. *Journal of grid computing*. 2014. Vol. 12. № 4. P. 559–592.

O.I.Kutuzov, T.M.Tatarnikova (Saint Petersburg Electrotechnical University “LETI”, St. Petersburg)
Network traffic prediction

The problem of increasing the volume of network traffic, which leads to the occurrence of overloads of data transmission channels, is discussed. Actual forecasting models, their advantages and disadvantages are given. It is shown that the forecast results depend on the quality of retrospective data on which the forecast model is built. The main techniques of pre-processing of retrospective data are considered. A computer model for predicting network traffic is implemented - an integrated model of autoregression and moving average. Forecast error does not exceed 0.04 for both short-term and medium-term and long-term periods.

Авторы готовы представить текст на английском языке для сборника материалов мультиконференции, который будет подан для индексирования в Scopus