

Выбор метода анализа и обработки данных в Big Data при моделировании трафика в телекоммуникационных сетях

Е. А. Мазина¹, Э. В. Логин²

Петербургский государственный университет путей сообщения Императора Александра I

¹e07.mazina@gmail.com, ²elinabeneta@yandex.ru

Аннотация. В статье рассмотрены технологии Big Data, которые уже нашли применение на железных дорогах. Также в работе предлагается оценка этих технологий на предмет перспективного моделирования процессов анализа данных в телекоммуникационных сетях. В частности – для анализа трафика сетевых узлов с целью повышения качества обслуживания, масштабирования и оптимизации работы.

Ключевые слова: Big Data; анализ данных; телекоммуникация; сети связи; моделирование трафика; маршрутизация

I. ВВЕДЕНИЕ

В настоящее время цифровые технологии нашли широкое применение практически во всех сферах деятельности, появляется огромное количество источников информации, которую необходимо передавать и обрабатывать. Одной из значимых частей информационной системы являются телекоммуникационные сети, в узлах которых происходит агрегация и маршрутизация трафика посредством осуществления процессов управления информационными потоками.

Термин Big Data в узком смысле – большие объемы данных, которые настолько велики, что не поддаются традиционным средствам обработки [1].

Опираясь на базовые понятия [1], сформулируем уникальные свойства больших данных применительно к телекоммуникационным сетям (рис. 1).

Объем. Количество данных – в сетях связи источником таких данных является трафик, информация о физическом состоянии коммутационных приборов в узлах и пр. Самоподобный трафик носит случайный характер и характеризуется изменением интенсивности потоков данных в разное время суток, а также непредсказуемостью характера и темпа работы пользователей в сети передачи данных. Существенное изменение перечисленных характеристик трафика может приводить к возникновению перегрузок и снижению качества обслуживания [2]. Наличие больших объемов неструктурированных данных и их дальнейший анализ вызывает трудности. Ценность таких данных не всегда известна.

Скорость. Это скорость приема данных, обработки и передачи. Обычно высокоскоростные потоки данных поступают прямо в оперативную память, а не

записываются на диск. Некоторые "умные" продукты, функционирующие на основе Интернета, работают в режиме реального или практически реального времени. Соответственно, такие данные требуют оценки и действий в реальном времени [1].

Разнообразие. Текстовые, аудио и видео сообщения, передаваемые в телекоммуникационных сетях, требуют дополнительной обработки для определения их важности, структурирования и последующей работы над ними.

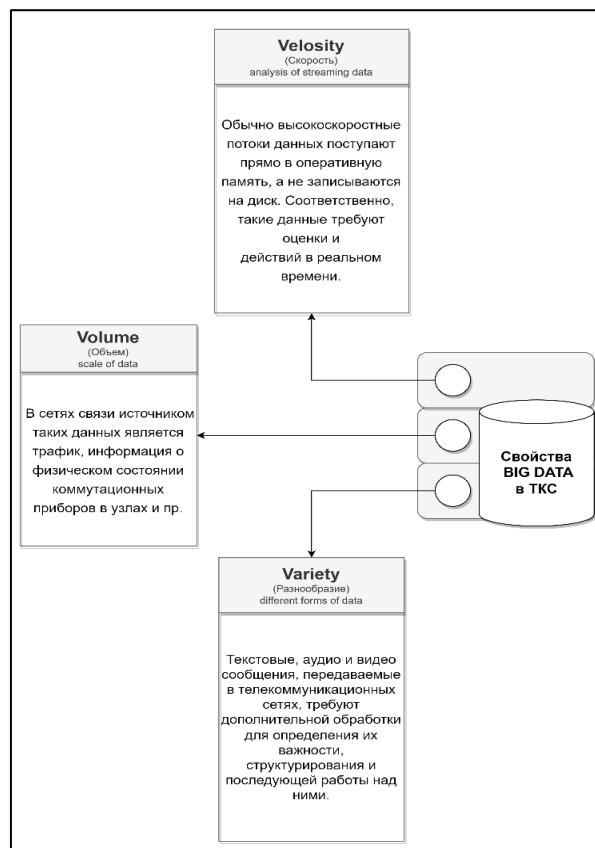


Рис. 1. Свойства больших данных применительно телекоммуникационным системам

Итак, телекоммуникационные сети являются одним из источников Big Data. Анализ трафика сети с применением методов анализа Big Data позволит

решить задачу по обслуживанию и мониторингу телекоммуникационных сетей связи.

II. ОБЗОР ЛИТЕРАТУРЫ

В транспортной области технологии Bid Data применяются для анализа данных об обслуживании железнодорожных вокзалов [3], управления рисками и безопасностью движения поездов [4], а также для облегчения технического обслуживания участков пути [5].

В статье [3] было изучено географическое распределение поездок с пересадкой из железнодорожного транспорта на автобусный, и их характеристики с использованием данных транзитных смарт-карт. Совокупное распределение высаживающихся пассажиров по автобусным маршрутам помогло определить форму и размер зоны обслуживания городского железнодорожного вокзала в каждом направлении, а также зависимости от времени суток. В данном методе используется достаточно простая и понятная модель анализа данных. При моделировании трафика такой способ анализа данных может быть использован для гибкой маршрутизации сети связи с учетом быстрого реагирования системы на изменения в таблице маршрутизации, исключая перерыв в работе сервиса.

Для управления рисками для безопасности железных дорог в Великобритании уже используются технологии Big Data и оценивается целесообразность их применения на железных дорогах [4]. Программное обеспечение RAATS (аббревиатура от регрессионного анализа временных рядов, представляет собой статистический пакет для анализа временных рядов и эконометрики, применяемое для анализа) рисков SPAD (событие, когда поезд проезжает сигнал, показывающий красный сигнальный свет без авторизации известен как «Signal Passed At Danger», SPAD) предоставляет сложные сведения о количестве поездов, приближающихся к светофору. RAATS повышает качество обеспечения безопасности движения поездов на железных дорогах Великобритании путем анализа большого потока данных в реальном времени. Данное программное обеспечение уже было опробовано на железных дорогах Великобритании, что может быть использовано в дальнейшем.

Применимость методов больших данных для облегчения принятия решений по техническому обслуживанию железнодорожных путей рассмотрено в статье [5]. Данный способ позволяет работать с разными источниками данных и типом получаемых данных. Может быть заимствован как сам аппарат, так и концепция сбора и обработки неструктурированных и разнородных данных о состоянии сетевых элементов с последующей программной реализацией для достижения более высоких показателей эффективности функционирования.

III. ТЕОРЕТИЧЕСКИЙ ПОДХОД И МЕТОДОЛОГИЯ

Для решения задачи, связанной с повышением качества обслуживания сетей связи в условиях роста контролируемых диагностических данных, могут быть

использованы методы аналитики Big Data, приведенные выше.

Для определения модели обработки данных проанализируем существующие методы по признакам, которые могут быть применены непосредственно при оценке трафика телекоммуникационной сети (табл. 1).

ТАБЛИЦА I. СРАВНЕНИЕ МЕТОДОВ

Методы	Собираемые данные	Особенности реализации
Анализ зоны обслуживания железнодорожного вокзала на основе анализа транзитных смарт-карт	ID карты Время отправления ID транзакции Тип метро Тип автобуса ID автобусного маршрута Тип пассажира Время посадки ID места посадки Время высадки ID места высадки Тариф Дополнительный тариф Дистанция поездки	Каждый автобусный маршрут извлекается от исходной автобусной остановки, которая является ближайшей на пути к станции метро. На графике отображается приблизительное направление от станции метро. Однако некоторые автобусы ходят по одним и тем же маршрутам менее чем в 5 км от начальной автобусной остановки, поэтому каждая прямая линия представляет приблизительное направление от начальной автобусной остановки, чтобы уточнить распределение совокупного процента высаживающихся пассажиров по каждому автобусному маршруту.
Анализ данных при техническом обслуживании железнодорожного пути	GPS данные Частота определения местоположения Информация об ошибках Ультразвуковые и вихретоковые измерения Видеоизображения	Так как деградация пути является медленным процессом, то не все данные требуют ежедневного мониторинга. В зависимости от среднего количества поездов, использующих определенный путь, интервал обработки для полного анализа может быть скорректирован. В среднем ежемесячной обработки каждого пути может быть достаточно. С помощью принципа разложения прямая и двойственная задача линейного динамического программирования сводятся к эквивалентным задачам линейного программирования с неявно заданными ограничениями [6]. Постоянный контроль в реальном времени и анализ получаемых данных на участках с критическими объектами усложняет процесс обработки. (САМА) Используется метод сокращения: данные делятся на множество меньших наборов. Затем, несколько различных наборов обрабатываются либо с

Методы	Собираемые данные	Особенности реализации
		параллельной обработкой, либо на отдельных компьютерах, что делает загрузку данных своевременной в текущих пределах обработки.
ПО RAATS как метод обработки Big Data на железных дорогах	Сообщения с TD ¹	Пользователь может выбрать анализ всех сигналов в определенной области или всех сигналов в базе данных. Графический интерфейс демонстрирует состояние TD и позволяет преподнести проанализированную информацию в доступном виде.

1TD (Train Describer) – это электронное устройство, подключенное к панели сигнализации, которое предоставляет описание каждого поезда (его код маршрута) и какой участок пути (или секции пути) он занимает в данный момент.

Опираясь на модель, описанную в статье [3], может быть предложен метод анализа трафика (табл.1) в узлах сети связи с дальнейшим определением загруженности сети во времени.

Это позволит определить в дальнейшем возможность масштабирования сети, в зависимости от полученной максимальной пропускной способности сети, и определить оптимальное время для проведения технических работ, при минимальной загруженности. То есть при использовании модели, подобной той, что предлагается в статье [3], в сети связи будет оцениваться временной промежуток работы сети, аналогично дистанции поездки в существующей модели (табл. 2), количество подключений в узле (аналогично ID карты в модели (табл. 2)), в какие периоды и какая нагрузка ложится на узлы (аналогично времени посадки/высадки (табл. 2)). На основе анализа данных и трафика, получаемых в режиме реального времени, могут быть также разработаны модели прогнозирования пропускной способности, что в дальнейшем наглядно продемонстрирует возможности по модернизации сети или вычислению времени его обслуживания.

В [5] приведен более сложный метод мониторинга данных. Его отличие от описанного выше заключается в сборе для анализа нескольких параметров с разных источников, однако алгоритм остается также простым и понятным для пользователя. Используется метод сокращения, скорость обработки значительно увеличивается, однако при использовании данной модели не все полученные данные могут быть обработаны вовремя, т. е. параметр своевременности и актуальности будет утерян.

Также для этих целей может применяться приложение RAATS, используемое для анализа сигналов в статье [4]. Программное обеспечение сохраняет передаваемую информацию в базе данных, представляет данные в графическом интерфейсе или создает файл

Excel для дальнейшего анализа. Результаты от одного источника могут быть проанализированы в течение периода от одного дня до года. Альтернативно пользователь может выбрать анализ всех сигналов в область или даже все сигналы в базе данных. Таким образом, применительно для сети телекоммуникации, можно получить результат анализа в понятном для оператора виде и принять решение о состоянии сети связи и дальнейшей ее модернизации и оптимизации.

IV. ЗАКЛЮЧЕНИЕ И ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЯ

Таким образом, рассмотренные технологии могут быть применены для анализа трафика телекоммуникационных сетей. В дальнейшем при модернизации системы, с учетом всех критериев управления и сложности рассматриваемого объекта ТКС, может более предметно выбран способ анализа, а также подобрана модель трафика, соответствующей сети телекоммуникации. Также могут быть подобраны инструменты для более точной аналитики получаемых результатов.

Дальнейшие исследования в направлении моделирования трафика и процессов маршрутизации телекоммуникационных сетей позволят сформулировать требования как для перспективного аппарата сбора и обработки данных о функционировании сетей связи с учетом технологий Big Data и Data Science, а также раскрыть требования для системы управления, способной реализовать предложенные программные методы. В работе предложен концептуальный подход к формированию уникальных критериев, позволяющих оценивать объект управления с точки зрения характеристики диагностируемых данных о нем. Полученный результат является актуальным для обоснования выбора необходимой технологии Big Data и ее соответствия телекоммуникационной сети связи как объекта управления.

СПИСОК ЛИТЕРАТУРЫ

- [1] OCL: сайт. – URL: <https://www.oracle.com/cis/big-data/what-is-big-data/> (дата обращения: 27.02.2023)
- [2] Канаев А.К., Сахарова М.А. Моделирование процессов обработки потоков IP-пакетов с различными типами информационных составляющих на основе глубокого анализа трафика // Бюллетень результатов научных исследований. 2014. №. 3 (12). С. 85-93.
- [3] Van Gulijk C. et al. Big data risk analysis for rail safety? // Proceedings of ESREL 2015. CRC/Balkema, 2015.
- [4] Núñez A. et al. Facilitating maintenance decisions on the Dutch railways using big data: The ABA case study // 2014 IEEE International Conference on Big Data (Big Data). IEEE, 2014. С. 48-53.
- [5] Eom J.K. et al. Exploring the catchment area of an urban railway station by using transit card data: Case study in Seoul // Cities. 2019. Т. 95. С. 102364.
- [6] Белухин В.П. Принцип разложения и метод сокращения невязок для задач линейного динамического программирования // Журнал вычислительной математики и математической физики. 1975. Т. 15. №. 6. С. 1424-1435.