

Сбор данных о работе оборудования в сети мобильного оператора связи

С. Б. Жанаева

Операторы мобильной сети передачи данных с ростом количества оборудования сталкиваются с увеличивающейся сложностью эксплуатации и растущими затратами на обслуживание. При увеличении количества базовых станций растет и количество сбоев. Современные технологические решения, основанные на алгоритмах нейронных сетей, способны заблаговременно с определенной вероятностью предсказать возникновение сбоев в работе оборудования. Для обучения модели нейронной сети требуются данные о работе и сбоях на оборудовании мобильной сети передачи данных. В данной статье рассказывается о выполненном сборе данных в сети мобильного оператора 4G+, об особенностях и ограничениях, которые в дальнейшем могут повлиять на обучение модели.

Ключевые слова: сбор данных, мобильные сети, сбои в работе оборудования, центр управления сетью.

1. Введение

С развитием технологии виртуализации сетей и внедрением 5G/IoT мобильные сети предлагают более разнообразные услуги и становятся более сложными. Большинство телекоммуникационных мобильных операторов контролируют сотни тысяч микроволновых каналов различной конфигурации. Как следствие, центры управления сетью (Network Operational Center, NOC) часто бывают перегружены поступающими данными о состоянии базовых станций и мгновенными сигналами аварий или сбоев в работе оборудования.

Несмотря на то, что центры управления сетью NOC могут выполнять множество задач, увеличивающийся размер мобильных сетей и большое количество разнообразных аварийных сигналов затрудняют обслуживание сети и делают его более ресурсозатратным [1].

К тому же нередко случается, что на сети происходят тысячи аварийных сигналов в минуту, и перед сотрудниками центра управления сетью появляется несколько задач, например, им необходимо вручную проанализировать и определить приоритетные базовые станции, на которые следует в первую очередь отправить инженерную бригаду, и распознать/предположить вероятную причину проблемы на станциях.

Вследствие этого центрам управления сетью зачастую приходится отправлять сервисные группы на места для замены оборудования, не зная точной причины проблемы. Это часто приводит к ненужным дорогостоящим посещениям объекта и замене оборудования, возможно, без обнаружения неисправности.

Специалисты, работающие в центрах управления сетью NOC, собирая и анализируя большой объем данных, охватывающий определённый промежуток времени, с высокой долей вероятности могут быстро и легко найти первопричину неисправности в микроволновой линии связи. Но постоянная ручная работа по перебору большого объема данных в сетях с десятками тысяч микроволновых линий связи с целью найти и определить небольшую вы-

борку проблемных станций является сложной задачей, требующей постоянной большой внимательности и определенной нагрузки на зрение.

Задачи по работе с большими объемами данных как раз являются той областью, где алгоритмы глубокого обучения могут принести немалую пользу и выполнить утомительную работу [2].

Можно обучить алгоритмы решать как сложные, так и более простые задачи, выполняемые экспертами, а также возложить на них задачу непрерывно просматривать огромные объемы данных в поисках первопричины сбоев. Это возможность использования алгоритмов машинного обучения с целью помочь операторам связи управлять сетевой инфраструктурой более эффективно по времени и с меньшими эксплуатационными расходами.

2. Обзор литературы

Со времени увеличения мощностей вычислительных машин и последовавшего затем возрождения и успехов алгоритмов машинного и глубокого обучения многие разработчики заинтересовались темой применения искусственного интеллекта в телекоммуникационной сфере, так как именно эта область деятельности человека сопровождается производством огромного количества структурированных данных. В том числе многих исследователей заинтересовал вопрос прогнозирования сбоев в сетях мобильного оператора, и ими был выполнен сбор необходимых для исследования данных.

Например, авторы в [3–5] собирали данные об авариях на базовых станциях, получаемых ежедневно, объединяли их с диспетчерской информацией о неисправностях на базовых станциях с целью установления связи между временем возникновения неисправности и предупреждающей информацией об аварии за K дней до возникновения неисправности. В качестве источника входных данных они использовали тревожную информацию с базовых станций (Alarms), генерируемую модулем анализа обработки тревожных сигналов, а в качестве источника данных маркировки использовали диспетчерскую информацию о неисправностях базовых станций (Tickets). Авторы в [3–5] описали интеллектуальную систему управления мобильной сетью, созданную с помощью методов контролируемого машинного обучения в целях создания системы интеллектуального обслуживания.

Авторы в [6, 7] предлагают реализовать систему прогнозирования сбоев и мониторинга сервиса, способную прогнозировать и оценивать качество услуг, близкое к тому, что на самом деле испытывают клиенты, оценивая его по сетевым данным состояния базовых станций и данным по авариям. Для достижения цели они предлагают собирать различные данные из сети, включая информацию об аварийных сигналах оборудования, и объединять их в большие кластеры данных для многовекторного анализа с помощью алгоритмов глубокого обучения. При реализации сервисного мониторинга авторы предлагают преобразовать процесс технического обслуживания из технического обслуживания объекта, сосредоточенного на аварийных сигналах оборудования, в техническое обслуживание сети, основанное на опыте клиентов.

Авторы в работе [8] для улучшения обслуживания и прогнозирования сбоев предлагают использовать собранную информацию об авариях на базовых станциях и данные о сроках эксплуатации оборудования, увеличивая расчетную вероятность отказа оборудования с возрастанием срока эксплуатации.

Большинство авторов [3–10] сходятся во мнении о необходимости использования нескольких типов источников данных. Так, например, авторы [10, 11] советуют, помимо данных мониторинга состояния, также собирать и использовать в прогнозировании сбоев влияние переменных внешней среды, таких как данные о рабочем состоянии, а также влиянии незначительных действий по техническому обслуживанию.

3. Описание собранных данных

Для того чтобы обучить алгоритмы машинного обучения и помочь операторам связи в управлении сетевой инфраструктурой, необходимо было собрать требуемые данные о работе сети. На данном этапе проведения исследовательской работы был выполнен сбор данных по качеству работы сетевого оборудования, базовых станций, по количеству и времени вынужденных простоев. Сбор данных проводился в течение нескольких месяцев, в разные дни недели и в разное время суток. Дни недели и время для сбора статистических данных определялись прогнозируемым объемом трафика и зависели от степени загруженности сети. Все данные сохранялись специальным программным обеспечением в формате csv-файлов.

3.1. Данные по авариям на оборудовании

В первую очередь необходимо было собрать данные, выдаваемые системой о различных аварийных событиях в сети, с учетом определенного специальным программным обеспечением по наблюдению за станциями распределения аварий по уровню критичности. При эксплуатации беспроводной сети такого типа данные об аварийных ситуациях собираются на постоянной основе и сохраняются в дневных, недельных, месячных отчетах (Alarms Daily Report, Alarms Weekly Report, Alarms Monthly Report) и в дальнейшем анализируются несколькими отделами компании оператора связи. Всего были собраны данные по аварийным ситуациям на 2700 станциях, данные по каждой станции и сектору были сохранены под определенным идентификатором. Сбор аварийных данных проводился в течение 10 месяцев, был собран большой объем данных на основе ежедневного отчета об аварийных сигналах Alarms Daily Report, который содержит такие параметры, как дата, время, станция, сектор, тип аварии, наименование аварии. Данные сохранены в формате csv-файлов и содержат около 3 млн записей (в дальнейшем часть записей была удалена из базы данных как дублированная).

3.2. KPI-показатели оборудования

Были собраны и сохранены данные о текущем состоянии базовых станций для того, чтобы в дальнейшем проанализировать их и определить наиболее вероятную первопричину сбоев на сети. Эти данные характеризуют производительность сетевого оборудования и отображают основные показатели KPI (Key Performance Indicator). Было собрано два типа подобных отчетов:

1. Показатели Average KPI (среднее значение показателей производительности) базовых станций: собрана информация усредненных за день технических показателей качества сервисов для каждой соты, на которой предоставляли сервис абонентам.

2. Показатели Channel KPI (канальное значение показателей производительности) базовых станций: собрана информация об изменениях (в час наибольшей нагрузки) технических показателей качества сервисов для каждой соты, на которой предоставляли сервис абонентам.

Каждый из двух вышеперечисленных отчетов по состоянию 2700 базовых станций содержит 42 показателя, которые в дальнейшем планируется использовать в качестве характеристик данных и атрибутов при обучении алгоритмов машинного обучения. Сбор данных проводился в течение 10 месяцев, и каждый из отчетов Average KPI и Channel KPI содержит более 2 миллионов записей.

3.3. Причины сбоев в работе оборудования

Системы, отображающие состояние сети и базовых станций, по умолчанию выдают предположительную причину аварийного состояния, но бывает так, что предложенная системой причина не соответствует действительности. Например, выдаваемая системой авария Rectifier Failure (отказ выпрямителя) зачастую на практике означает выход из строя кондиционера и перегрев на станции. Поэтому одновременно со сбором данных по аварийным ситуациям на сети также был проведен сбор данных по выяснившимся первопричинам аварий. Первопричины аварий и сбоев на сети были выявлены и проверены выехавшими на место сетевыми инженерами и инженерами связи. Инженерный персонал ежедневно созванивался и высылал в центр управления сетью отчеты по устраненным авариям и сбоям на сети с выявленной и подтвержденной причиной и перечнем предпринятых ими действий для устранения аварии. Таким образом, была сформирована база данных по первопричинам аварийных ситуаций на сети.

После завершения сбора данных из общего объема трех вышеперечисленных типов данных была сделана выборка статистически значимого объема данных по анализируемым станциям.

3.4. Сложности в сборе данных

По признаку «видимости» сбои на оборудовании базовых станций мобильного оператора условно можно разделить на два основных типа: видимые явные и невидимые скрытые. Видимые или явные сбои в работе оборудования отображаются приходящими с системы управления и мониторинга сигналами тревоги, аварии на оборудовании и высвечиваются в соответствующих окнах GUI. Увидев их, сотрудники центра мониторинга могут немедленно связаться с инженерной бригадой и отправить ее на место для решения проблемы в работе оборудования.

Иногда кроме видимых аварий, явно отображаемых в системе мониторинга, возникают так называемые невидимые скрытые аварии в сети. Например, иногда на некоторых базовых станциях можно заметить полное пропадание трафика и возникновение молчания, на станцию как будто не поступают запросы на соединение от абонентов, кажется, что на данной станции или в секторе абоненты отсутствуют. Такая ситуация иногда возникает и бывает вызвана «зависанием» программного обеспечения станции [3]. Сотрудники службы мониторинга могут заметить отсутствие трафика и зависание программного обеспечения, для решения проблемы они обычно выполняют операцию Reset станции, и через минуту-две трафик на станции восстанавливается. При этом в системе мониторинга возникает авария отключения станции, падения радиоканала. Как следствие, данная авария сохраняется в базе данных в виде аварии отключения базовой станции и в дальнейшем при обучении нейронной модели прогнозированию сбоев может быть неверно интерпретирована и ошибочно воспринята как авария отключения питания или помехи в радиоканале. Подобную аварию можно распознать, одновременно собирая данные по показателям производительности базовых станций, и выявить продолжительное отсутствие трафика перед отключением станции.

4. Результаты сбора данных

4.1. Данные по авариям на оборудовании

Всего было собрано три комплекта данных: база данных по авариям на оборудовании, база данных по состоянию базовых станций с показателями производительности KPI и база данных по первопричинам сбоев на станциях. Все данные были сохранены в формате csv-

файлов с разделительными запятыми. Данные были проанализированы с помощью библиотек Python: Numpy и Pandas, были определены количество собранных данных, размерность.

```
import numpy as np
import pandas as pd
url = '.../alarm_daily/data'
data_alm = pd.read_csv(url, header=None)
data_alm.shape ()

(3847653, 5)
```

Так, база данных по авариям состоит из 3847653 строк, каждая строка отображает аварийное состояние оборудования мобильной сети. Столбцы содержат следующие данные: уровень критичности аварии, наименование и ID базовой станции, время фиксирования аварии, предположительная причина аварии, наименование трансмиссии. В табл. 1 показана выборка и пример содержащейся информации в базе данных по авариям.

Таблица 1. Выборка из базы данных по аварийным ситуациям в сети

№	Уровень	Наименование	Время	Причина	Трансмиссия
1	Minor	37323_AKT-Beyneu	2021-02-15 11:32:28	Abis control link broken	Aktau-BSC-2
2	Minor	51247_KZL-Merke	2021-02-15 11:32:28	Abis control link broken	Kzl-BSC-1
3	Major	71422_ALM-Issyk	2021-02-15 11:32:28	Rectifier failure	Alm-BSC-2
4	Critical	37375_AKT-Karasu	2021-02-15 11:32:29	POWER OFF	Aktau-BSC-2
5	Minor	71458_ALM-Issyk	2021-02-15 11:32:29	Abis control link broken	Alm-BSC-1

База данных по авариям обладает большой избыточностью, так как зачастую на одно аварийное состояние оборудования система мониторинга и записи аварий выдает несколько тревожных сообщений с разницей в секунды, в основном такое отображение аварийных состояний характерно при помехах в радиоканале. На рис. 1 показано избыточное отображение одной и той же аварии в системе мониторинга с разницей в несколько секунд.

	Major	54836_SMP-Ayaguz	2021-03-18 16:21:42	Rectifier failure	Semey-BSC-1
	Critical	54836_SMP-Ayaguz	2021-03-18 16:21:43	POWER OFF	Semey-BSC-1
	Minor	72423_ALM-Issyk-3	2021-03-18 16:21:43	Abis control link broken	Alm-BSC-2
	Minor	72423_ALM-Issyk-3	2021-03-18 16:21:45	Abis control link broken	Alm-BSC-2
	Minor	72423_ALM-Issyk-3	2021-03-18 16:21:46	Abis control link broken	Alm-BSC-2
	Minor	72423_ALM-Issyk-3	2021-03-18 16:22:10	Abis control link broken	Alm-BSC-2
	Minor	72423_ALM-Issyk-3	2021-03-18 16:22:12	Abis control link broken	Alm-BSC-2
	Minor	72423_ALM-Issyk-3	2021-03-18 16:22:14	Abis control link broken	Alm-BSC-2

Рис. 1. Дублирующее отображение аварий в системе мониторинга

В дальнейшем при подготовке данных из базы данных по авариям избыточность была удалена.

4.2. Данные по состоянию базовых станций

Был проведен сбор данных по состоянию базовых станций. Как было описано в пункте 3.2, состояние базовых станций характеризуется двумя наборами данных: показателями Average KPI и показателями Channel KPI.

Информация, содержащаяся в этих комплектах данных, необходима для анализа состояния базовых станций до момента появления аварии, чтобы в дальнейшем проанализировать и найти возможную корреляционную зависимость между определенными авариями и изменениями показателей состояния оборудования базовых станций с целью прогнозирования первопричин аварий.

В каждой из собранных баз данных Average KPI и Channel KPI содержится информация по 2700 базовым станциям и 42 показателям. С помощью библиотек Python Numpy и Pandas были определены количество собранных данных и размерность.

```
import numpy as np
import pandas as pd
url_a = '.../average_kpi/data'
url_c = '.../channel_kpi/data'

data_average_kpi = pd.read_csv(url_a, header=None)
data_channel_kpi = pd.read_csv(url_c, header=None)

data_average_kpi.shape ()
data_channel_kpi.shape ()

(896727, 42)
(896560, 42)
```

Всего было собрано около 900 тысяч данных по 42 показателями, по каждой базе данных Average KPI и Channel KPI. В целях прогнозирования сбоев и вероятностного определения первопричин аварий будут необходимы данные только по базовым станциям, на которых по тем или иным причинам возникали аварийные сообщения. Поэтому из всего объема собранных данных по показателям производительности базовых станций будут выбраны только данные с номерами базовых станций, совпадающими с номерами станций из базы данных по аварийным сообщениям. Таким образом, базы данных Average KPI и Channel KPI будут уменьшены.

4.3. Данные по первопричинам сбоев

В результате совместной работы сотрудников центра мониторинга сети и инженерного состава, обслуживающего оборудование базовых станций, была сформирована база данных по первопричинам аварийных ситуаций на сети. Эти данные необходимы для обучения модели прогнозирования и определения первопричин аварийных ситуаций в работе оборудования. Собранные данные были скомпонованы сотрудниками отдела эксплуатации базовых станций и сохранены в формате csv-файлов. С помощью библиотек Python Numpy и Pandas были определены количество собранных данных и размерность.

```
import numpy as np
import pandas as pd
url_pa = '.../alarm_reasons/data'

data_alarm_reasons = pd.read_csv(url_alarm_reasons, header=None)
data_alarm_reasons.shape ()

(68596, 4)
```

Было подсчитано, что база данных с первопричинами аварий состоит из 68596 строк. Столбцы базы данных содержат следующую информацию: наименование и ID базовой станции, время фиксирования аварии, выявленная причина аварии, трансляция.

5. Заключение

На данном этапе исследовательской работы был проведен сбор данных по работе оборудования базовых станций мобильной сети телекоммуникационного оператора 4G+. Сбор данных проводился с целью дальнейшего использования для разработки нейронной модели прогнозирования сбоев в работе оборудования мобильной сети. Разработанная модель призвана послужить в качестве решающего средства для предотвращения непредсказуемых потерь в работе сети, оптимизации расходов на обслуживание оборудования, способствуя тем самым повышению качества предоставляемых услуг.

Литература

1. *Тухвинский В. О., Терентьев С. В., Коваль В. А.* Сети мобильной связи 5G: технологии, архитектура и услуги. М.: Медиа Паблишер, 2019. 375 с.
2. *Brink H., Richards J. W., Fetherolf M.* Real world Machine Learning. US Manning Publications Co., 2017. 266 p.
3. *Yin-Hsin Liu, Yao-Chung Tu, Chang-Yu Hsu, Hsin-Chieh Chao.* Predicting malfunction of mobile network base station using machine learning approach // Proc. 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), 18–20 September, Matsue, Japan. 2019. P. 1–4. DOI: 10.23919/APNOMS.2019.8892894.
4. *Kumar Y., Farooq H., Imran A.* Fault Prediction and Reliability Analysis in a Real Cellular Network // Proc. 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 2017. P. 1090–1095. DOI:10.1109/IWCMC.2017.7986437.
5. *Corazza A., Isgro F., Longobardo L., Prevete R.* A machine learning approach for predictive maintenance for mobile phones service providers // Advances on P2P, Parallel, Grid, Cloud and Internet Computing / Lecture Notes on Data Engineering and Communications Technologies. / F. Khafa (Ed.). Springer International Publishing, 2017. DOI: 10.1007/978-3-319-49109-7_69.
6. *Otani T., Toubé H., Kimura T., Furutani M.* Application of AI to mobile network operation // ITU Journal: ICT Discoveries. 2018. Special Issue. №. 1. P. 43–48.
7. *Hu C., Youn B. D., Kim T.* Semi-supervised learning with co-training for data-driven prognostics // Proc. IEEE Int. Conf. on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, Denver, CO, USA. 18–21 June 2012. P. 1297–1306. DOI: 10.1109/ICPHM.2012.6299526.
8. *Yousefia N., Tsianikasa S., Zhoua J. Coit D. W.* Inspection plan prediction for multi-repairable component systems using neural network // Proc. IISE Annual Conference. / L. Cromarty, R. Shirwaiker, P. Wang (Eds.). Institute of Industrial and Systems Engineers, IISE, Virtual, Online, United States. 2020.
9. *Schmidt B., Wang L.* Cloud-enhanced predictive maintenance // Int J Adv Manuf Technol. 2018. № 99. P. 5–13. DOI 10.1007/s00170-016-8983-8.
10. *Wei Wu, Feng Zhang, Min Liu, Weiming Shen.* A Multi-agent Based Failure Prediction Method Using Neural Network Algorithm // Proc. IEEE International Conference on Systems, Man, and Cybernetics. San Diego, CA, 2014. P. 2268–2272. DOI: 10.1109/SMC.2014.6974263.
11. *Karballaezadeh N., Zaremotekhas F., Shamshirband S., Mosavi A., Nabipour N., Csiba P., Várkonyi-Kóczy A.* Intelligent Road Inspection with Advanced Machine Learning; Hybrid Prediction Models for Smart Mobility and Transportation Maintenance Systems // Energies. 2020. № 13. P. 1718. DOI: 10.3390/en13071718.

Статья поступила в редакцию 03.05.2022;
переработанный вариант – 17.05.2022.

Жанаева Сауле Бактыкереевна

аспирант, кафедра прикладной математики и кибернетики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: szhanayeva@gmail.com.

Data collection concerning equipment operation in the network of a mobile operator**Saule B. Zhanayeva**

Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), szhanayeva@gmail.com.

Mobile data network operators face increasing operational complexity and rising maintenance costs as the number of equipment increases. As the number of base stations increases, so does the number of failures. Modern technological solutions based on neural network algorithms are able to predict in advance with a certain probability the occurrence of equipment failures. Data of operation and failures on the mobile network equipment is required to train a neural network model. The article considers the performed data collection on the 4G mobile operator network, features and limitations that may further affect the model training.

Keywords: data collection, mobile networks, equipment failures, network operational center.

References

1. Tihvinskiy V. O., Terentev S. V., Koval V. A. *Seti mobilnoy svyazi 5G: tehnologii, arhitektura i uslugi* [Mobile Networks 5G: technologies, architecture and services]. Moscow, Media Publishers, 2021, 375 p.
2. Brink H., Richards J. W., Fetherolf M. *Real world Machine Learning*. US, Manning Publications Co., 2017, 266 p.
3. Yin-Hsin Liu, Yao-Chung Tu, Chang-Yu Hsu, Hsin-Chieh Chao. Predicting malfunction of mobile network base station using machine learning approach. *Proceedings of the 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 18-20 September, Matsue, Japan, 2019, pp. 1-4. DOI: 10.23919/APNOMS.2019.8892894.
4. Kumar Y., Farooq H., Imran A. Fault Prediction and Reliability Analysis in a Real Cellular Network. *Proceedings of 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 26-30 June, pp. 1090-1095, Valencia, Spain, 2017. DOI: 10.1109/IWCMC.2017.7986437.
5. Corazza A., Isgro F., Longobardo L., Prevete R. A machine learning approach for predictive maintenance for mobile phones service providers. *Advances on P2P, Parallel, Grid, Cloud and Internet Computing. Lecture Notes on Data Engineering and Communications Technologies*, Springer International Publishing, 2017. DOI: 10.1007/978-3-319-49109-7_69.
6. Otani T., Toubé H., Kimura T., Furutani M. Application of AI to mobile network operation. *ITU Journal: ICT Discoveries*, 2018, Special Issue, no. 1, pp. 43-48.
7. Hu C., Youn B. D., Kim T. Semi-supervised learning with co-training for data-driven prognostics. *Proceedings of IEEE Int. Conf. on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application*, 18-21 June, Denver, CO, USA, 2012, pp. 1297-1306. DOI: 10.1109/ICPHM.2012.6299526.
8. Yousefia N., Tsianikasa S., Zhoua J., Coit D. W. Inspection plan prediction for multi-repairable component systems using neural network. *Proceedings of the 2020 IISE Annual Conference*, Institute of Industrial and Systems Engineers, IISE, United States, 1-3 Nov, 2020.
9. Schmidt B., Wang L. Cloud-enhanced predictive maintenance. *Int J Adv Manuf Technol*, 2018, no. 99, pp. 5–13. DOI 10.1007/s00170-016-8983-8.
10. Wei Wu, Feng Zhang, Min Liu, Weiming Shen. A Multi-agent Based Failure Prediction Method Using Neural Network Algorithm. *Proceedings of 2014 IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA., 2014, pp. 2268-2272, DOI: 10.1109/SMC.2014.6974263.
11. Karballaezadeh N., Zaremotekhas F., Shamshirband S., Mosavi A., Nabipour N., Csiba P., Várkonyi-Kóczy A. Intelligent Road Inspection with Advanced Machine Learning; Hybrid Prediction Models for Smart Mobility and Transportation Maintenance Systems. *Energies*, 2020, no. 13, pp. 1718. DOI: 10.3390/en13071718.