

УДК 629.4.053.2

ПРИМЕНЕНИЕ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ ОТКАЗОВ ПОДВИЖНОГО СОСТАВА

Комисарчук Алексей Васильевич, главный системный аналитик, АО «ВНИИЖТ», Москва, Россия,

E-mail: komisarchuk.aleksei@vniizht.ru

Елисеев Игорь Александрович, начальник отдела, к.т.н., АО «ВНИИЖТ», Москва, Россия,

E-mail: eliseev.igor@vniizht.ru

Сидоров Алексей Викторович, руководитель проектов, АО «ВНИИЖТ», Москва, Россия,

E-mail: Sidorov.Alexey@vniizht.ru

АННОТАЦИЯ

Эта работа открывает серию исследований, объясняющих необходимость и возможность прогрессивной диагностики технического состояния железнодорожного подвижного состава, и предлагает принципиально новое решение этой проблемы на основе современного интеллектуального алгоритма глубокого обучения. Здесь сформирована абстрактная концепция аномальности данных, дана формальная математическая постановка задач поиска и прогнозирования отказов промышленного оборудования, описаны основные компоненты программной реализации решения этих задач с применением модели рекуррентной нейросети (LSTM-автоэнкодера), а также приведены оценка качества работы и аргументация в пользу выбора глубокого обучения среди актуальных методов в области искусственного интеллекта.

Ключевые слова: искусственный интеллект, нейросеть, прескриптивная аналитика, аномалия, отказ, железнодорожный подвижной состав

APPLICATION OF DEEP LEARNING FOR ROLLING STOCK FAILURE PREDICTION

Aleksey V. Komisarchuk, system architect, M.Sc., Railway Research Institute, Moscow, Russia,

E-mail: komisarchuk.aleksei@vniizht.ru

Igor A. Eliseev, department chief, cand. sci. (eng.), Railway Research Institute, Moscow, Russia,

E-mail: eliseev.igor@vniizht.ru

Aleksey V. Sidorov, project manager, Railway Research Institute, Moscow, Russia,

E-mail: Sidorov.Alexey@vniizht.ru

ABSTRACT

This paper begins a series of publications, making a case and opportunity for creation of advanced diagnostics of technical condition of the railway rolling stock and provides essentially new artificial intelligence (AI) solution based on deep learning techniques. Here, the abstraction for anomalous data is formed, statements of industrial equipment failure detection and forecasting problems are set, core components of software implementation of the solution algorithm using the recurrent network (LSTM autoencoder) are described as well as the quality assessments and the reasoning for choosing deep learning over the rest AI methods.

Keywords: artificial intelligence, neural network, long short-term memory, prescriptive analytics, anomaly, failure, railway rolling stock

Введение

Промышленное оборудование зачастую работает в режиме непрерывного цикла, поэтому каждый его критический сбой (авария, отказ) – это существенные экономические потери. Срыв сроков, резкий рост затрат на восстановление работоспособности промышленного оборудования влияют на доходность и стабильность бизнеса. Исключением не является и железнодорожный транспорт, где весь подвижной состав (тяговый и не тяговый) является сложной технической системой. Обеспечение бесперебойной работы оборудования тягового подвижного состава является одной из приоритетных задач в ОАО «РЖД». Задержка поездов на линии из-за отказа оборудования влечет за собой прямые убытки (ремонт, простой) и косвенные (упущенная выгода, штрафы, репутационный ущерб). Не вовремя обнаруженный отказ в оборудовании подвижного состава в процессе ТО или ТР ведет к перебоям в движении на всем участке. Например, в пригородном пассажирском движении – это может быть настолько существенно, что «невыход» поезда на линию может оказаться выгоднее, чем простоя в результате потенциальной аварии. Таким образом, сбой в работе подвижного состава является существенным фактором, который может поставить под угрозу всю бизнес-модель перевозок.

Большинство сбоев не случайны, а являются результатами физических процессов, поэтому их можно предсказать и предотвратить, анализируя техническое состояние системы.

Далеко не все можно выявить регулярным техническим обслуживанием, а существующая диагностика крайне ограничена: она не позволяет строить содержательные прогнозы и выявлять дефекты в системе заранее.

Возникновение и развитие неисправностей во многих случаях сопровождается изменениями поведения соответствующих сигналов и, если обнаружить это заранее, то наступление отказа можно предсказать, чтобы успеть принять решение. Долгое время область анализа сигналов ограничивалась методами ХХ века; инструмента, подходящего для анализа больших данных, не было, особенно в эксплуатации на борту.

Внедрение цифровых систем управления и микропроцессорной техники на подвижном составе позволило получить необходимые вычислительные мощности, на которых теперь можно в реальном времени получать и обрабатывать большие объемы информации о техническом состоянии подвижного состава.

Создание программного решения для его диагностирования, реализованного на вычислительных мощностях информационно-управляющих систем подвижного состава, позволит перейти от решения задач класса «что произошло и каковы причины» к задачам класса «что делать». Для успешного решения последнего класса задач необходимо научиться прогнозировать поведение системы. Это позволит повысить живучесть системы или, другими словами, коэффициент технической готовности подвижного состава, а, следовательно, экономическую эффективность его эксплуатации.

Основной целью данной работы является первичное теоретическое, а также практическое обоснование применимости методов машинного обучения (и нейросетей, в частности) для решения ряда подобных задач, а также описание архитектуры общего решения.

В первых двух частях статьи рассматривается природа данных и недостатки существующих методов диагностики, затем (в части 3) задача диагностики сводится к нескольким задачам машинного обучения, далее предлагается программное решение одной из них (структура описана в части 4). В последней части мы оцениваем результаты и предлагаем альтернативный способ решения оставшейся задачи с помощью уже достигнутого, а также проводим полезную аналогию с другими областями глубокого обучения.

1. Техническая постановка задачи

Динамическое состояние сложной мультифизической системы (такой как поезд, реакторная установка) описывается множеством сигналов различной природы, которые проходя через аналогово-цифровые преобразователи и концентраторы регистрируются

в виде данных в информационно-управляющей системе. Эти данные могут быть использованы для функционирования различных алгоритмов диагностирования локомотива или моторвагонного подвижного состава.

2. Прямые алгоритмы

Простейшие алгоритмы самодиагностирования некоторых подсистем обеспечивают проверку соблюдения некоторых правил в каждый момент времени: соответствие сигнала диапазону значений или теоретическому остаточному ресурсу до плановой замены детали. Более сложные алгоритмы включают математические модели, описывающие физические законы в виде систем алгебраических и дифференциальных уравнений и оценивают степень отклонения результатов измерений реальных параметров с выходными параметрами такой модели. Так, например, модели некоторых узлов поезда ЭС2Г создавались на специальном языке Modelica [10] в проприетарной среде разработки Amesim. Это множество существующих алгоритмов диагностирования назовем прямыми алгоритмами.

У них есть ряд недостатков:

1. Ограничность. Каждый из таких алгоритмов учитывает в работе крайне небольшое число параметров в отрыве от всех остальных, имеющихся в наличии в системе управления. В целом, это проявление более фундаментальной проблемы, заключающейся в том, что данные телеметрии собираются, но почти никак не анализируются.
2. Ad-hoc разработка. Отсутствие универсального подхода к разработке алгоритмов. Каждый алгоритм создается вручную для конкретного узла, что трудозатратно, требует узкой специализации и координации усилий нескольких специалистов, поэтому такие алгоритмы зачастую трудно поддерживать в эксплуатации и масштабировать в будущем.
3. Невозможность прогнозирования состояния по многомерным сигналам. Они позволяют в лучшем случае моделировать линейный тренд отдельного сигнала, что дает условный прогноз его поведения в краткосрочной перспективе, но подходит лишь для очень простых сигналов, а не всего процесса.

Экстенсивный путь развития имеющейся системы (нагромождения разрозненных по подсистемам алгоритмов диагностирования) является тупиковым. Для выхода на другой уровень развития необходима иная стратегия диагностирования технического состояния системы. Она должна базироваться на принципиально новом инструменте, который обладал бы такими качествами:

1. Выявление отклонений в техническом состоянии работы оборудования для предсказания сбоев или отказов на ранней стадии развития дефекта.
2. Универсальность (масштабируемость): разрабатываемый алгоритм должен быть применим не только для одного конкретного поезда, а в целом для любой подвижной железнодорожной единицы с информационно-управляющей системой.
3. Полнота – учет всех измеряемых сигналов, которые прямо или косвенно отражают состояние системы или окружающей среды.
4. Автоматизация – минимизация настроек алгоритма.
5. Возможность локализации неисправности.
6. Наглядность причинно-следственных связей для обнаружения неочевидных конструктивных дефектов для учета в будущих разработках.

Наделенные такими качествами алгоритмы позволяют решать задачи прескриптивной (предписывающей) аналитики – задачи класса «что делать»?

В основу решения задач такого класса, обычно закладываются алгоритмы, основанные на прескриптивной аналитике. Такие алгоритмы учитывают в себе модели искусственного интеллекта и машинного обучения. Они позволяют анализировать разнообразные сценарии

работы оборудования и выбирать наилучшие решения по критерию безопасности и эффективности эксплуатации подвижного состава.

3. Формальная задача машинного обучения

Пусть заданы два множества: объекты X (вопросы) и целевая переменная Y (ответы).

Пусть между ними есть неизвестная зависимость $f: X \rightarrow Y$, которую нужно найти, зная лишь конечное множество пар, называемое обучающей выборкой D (или датасетом):

$$D := \{(x_i, y_i) \mid x_i \in X, y_i \in Y\}_{i=1}^N$$

Каждый алгоритм машинного обучения предлагает свое семейство допустимых отображений, называемое пространством гипотез,

$$H := \{\hat{f}(x): X \rightarrow Y\}$$

Процесс выбора «лучшего» кандидата f^* в таком семействе H и есть машинное обучение.

Это удобно представлять, как аппроксимацию функций, обобщением которой оно и является. Впрочем, теперь отображение \hat{f} может вообще не иметь функциональной формы. Это стандартная постановка задачи машинного обучения, все прочие так или иначе сводятся к ней. Иногда она называется обучением с учителем (supervised learning).

Задачи без предзаданного Y иногда называются обучением без учителя (unsupervised learning) и легко приводятся к стандартному виду (выше) выбором $Y = X$. К примеру, такой является задача сжатия (снижения размерности) данных, рассматриваемая в этой статье.

Для оценки кандидатов на H вводится функционал качества,

$$Q(\hat{f}): H \rightarrow R$$

и машинное обучение сводится к его оптимизации на H по некоторому критерию. Без ограничения общности, можно ввести функционал ошибки,

$$\mathcal{L}(\hat{f}) := Q(\hat{f}): H \rightarrow R^+ \cup \{0\}$$

и взять его минимизацию в качестве критерия оптимальности,

$$\min_{\hat{f} \in H} \mathcal{L}(\hat{f})$$

Две популярные в алгоритмах функции потерь — это реализации функционала ошибки,

$$L_1(\hat{f}_\theta) := \sum_{(x,y) \in D} |\hat{f}(x, \theta) - y|$$

$$L_2(\hat{f}_\theta) := \sum_{(x,y) \in D} (\hat{f}(x, \theta) - y)^2$$

Первая, L_1 также называется средней абсолютной ошибкой (*mean average error, MAE*) и мы именно ее будем использовать далее в алгоритме.

В случае, если речь идет о параметрическом семействе (функций),

$$H = \{\hat{f}_\theta(x): X \rightarrow Y\}_{\theta \in \Theta}$$

задача оптимизации на H сводится к задаче оптимизации на пространстве параметров Θ ,

$$\min_{\hat{f} \in H} \mathcal{L}(\hat{f}) = \min_{\theta \in \Theta} \mathcal{L}(\hat{f}(x, \theta))$$

Например, нейросеть определяет некоторую функцию параметрически и её обучение – это оптимальная настройка набора параметров $\theta \in \Theta$ на обучающей выборке D .

Действительно, задача машинного обучения для нейросети может быть сформулирована следующим образом:

$$H = \{\hat{f}_\theta(x): X \rightarrow Y = X\}_{\theta \in \Theta}, \quad X \subset R^n$$

сводится к минимизации функции потерь ($L1$) на пространстве параметров нейросети \hat{f} ,

$$\min_{\theta \in \Theta} \mathcal{L}(\hat{f}(x, \theta)) = \min_{\theta \in \Theta} \sum_{(x, y) \in D} |\hat{f}(x, \theta) - y|$$

4. Аномалии и решение задачи их поиска

На словах отказ понимается как необратимое критическое состояние системы, но это наше субъективное представление нужно формализовать и передать в алгоритм. Это можно сделать, создав размеченный (с примерами отказов) датасет для обучения,

$$X \subset R^n, \quad Y = \{0, 1\} \cong \{\text{норма, отказ}\} \quad (1)$$

Итак, с теоретической точки зрения, детекция отказов – это задача бинарной классификации (1). Данные телеметрии – это просто объекты X без пар, они не позволяют использовать методы машинного обучения с учителем, но вполне подходят для алгоритмов обучения без учителя и можно рассмотреть другой подход. Прежде всего, абстрагируемся от отказов, которые невозможно объективно формализовать и рассмотрим более широкое понятие – аномальность.

Аномалия – это последовательность точек, которые настолько выделяются [1] среди остальных данных, что возможен иной механизм их возникновения (ненормальный ход процесса, например, возникшая неисправность или внешнее воздействие).

Отсюда прямо следует, что:

- любое определение аномалии всегда относительно (нормы), ведь неявно определяет норму как некое господствующее поведение системы;
- для выявления аномалий крайне важно окружение точек (порядок и контекст); точки сами по себе (изолированные) ничего не значат.

На рисунке 1 наряду с нормальными формами сигналов показаны некоторые примеры различных аномалий одномерного сигнала: цветом выделены выбросы (фиолетовый), аperiодичность (красный), нарушение волатильности (зеленый).

В многомерном случае, точке в пространстве соответствует вектор значений нескольких переменных. Аномалия, связывающая несколько сигналов сразу, называется многомерной (многоканальной). Необходимо отметить, что это не то же самое, что множество одновременных одномерных аномалий и вполне может существовать, даже если по отдельности все сигналы в норме. В общем случае, она устроена сложно, ее непротиворечивая визуализация невозможна и можно лишь определить ее границы. Ее практически невозможно найти вручную, но для этого можно использовать нейронные сети.

Большинство аномалий бесследно исчезают и вообще не приводят к сиюминутному отказу. Отказ – это просто одна из множества аномалий, особая лишь для диагностирования.

В силу субъективности, выделить именно отказы среди остальных аномалий практически невозможно. В данной публикации ограничимся рассмотрением задачи поиска аномалий на неразмеченных данных, методы машинного обучения позволяют это сделать.

Суть предлагаемого алгоритма в том, что, считая, что норма – это основное состояние системы, мы будем моделировать нормальное поведение физических процессов, а аномалии будут найдены как существенные отклонения от него.

В результате рассмотрения преимуществ и недостатков различных подходов к моделированию, нейросеть была выбрана как лучший и единственный на данный момент универсальный подход к автоматизированному моделированию сложного многомерного процесса.

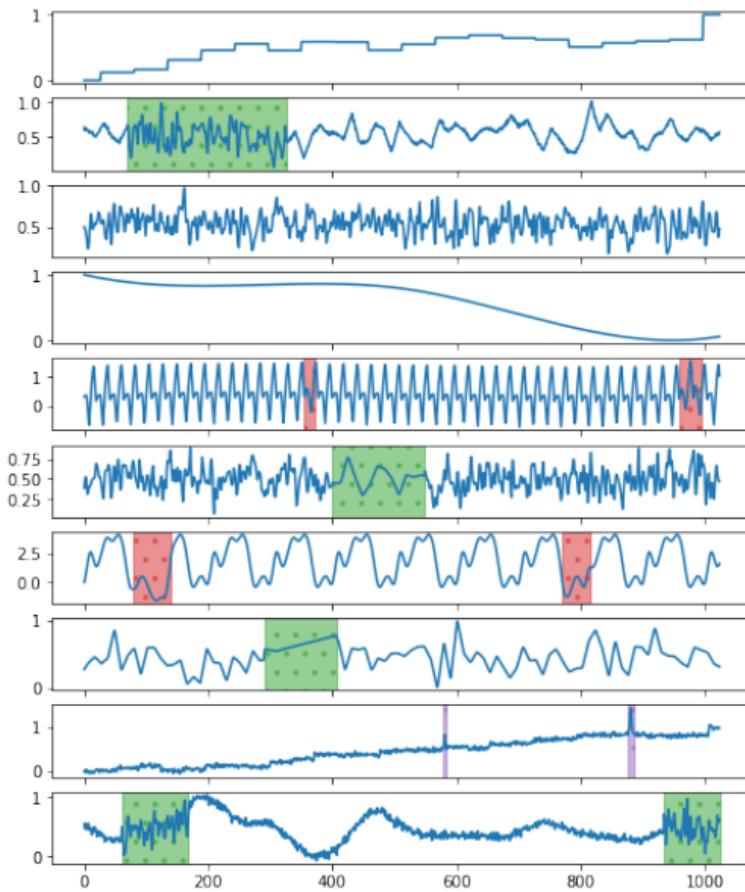


Рис. 1 – Цветом выделены примеры различных аномалий одномерных сигналов

Для того, чтобы оценить аномальность i -параметра, мы вводим метрику, которая сравнивает MAE на интересующем нас (*control*) периоде со всем предшествующим периодом (*past*) и коэффициент аномальности рассчитывается так:

$$AC(i) = \frac{MAE_{control}(i)}{MAE_{past}(i)}$$

Усреднение здесь происходит по каждому параметру независимо. Если модель была хорошо обучена, высокое значение коэффициента свидетельствует о ненормальном развитии процесса. Выбирая порог для каждого сигнала, разработчик настраивает чувствительность модели. Кроме этого, с помощью базовых ошибок этой модели можно автоматически создать предварительную разметку датасета для решения задачи обучения с учителем.

Есть два ключевых ограничения, определяющих архитектуру нейросети:

1. Для того, чтобы заставить нейросеть эффективно выявлять закономерности, а не запоминать и повторять данные, нужна структура слоев с «бутылочным горлышком», называемая автоэнкодером. Другими словами, перед нами встает задача сжатия данных X (с потерями) в латентное пространство меньшей размерности с качественными (отражающими физику процесса) признаками.
2. Для того, чтобы определять сложные аномалии (в контексте), нейросеть должна уметь работать с последовательностями. Для начала, можно взять рекуррентные [2] слои с *LSTM*-ячейками [3], обладающими кратко- и долгосрочной памятью. Кроме

всего прочего, у них есть уникальная особенность – умение работать с (условно) бесконечными последовательностями.

Итак, в качестве отправной точки была взята архитектура рекуррентного автоэнкодера (*RAE*). Отметим, что существует [4-6] сразу несколько его разновидностей; конкретная реализация, описанная здесь, похожа на них лишь отдаленно; ее структурная схема представлена на рисунке 2 ниже:

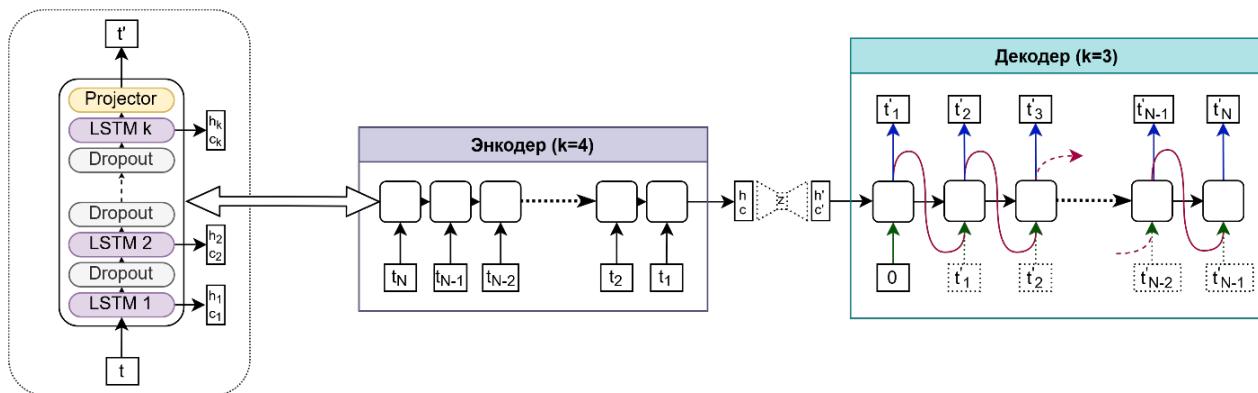


Рис. 2 – Структурная схема модели НС (RAE) и ее характеристики

Если представить, что последовательность состояний системы на входе – текст, можно увидеть аналогию нашей задачи с задачей машинного перевода текстов в области обработки естественных языков (*NLP*). Это не случайность и наша модель напоминает *Seq2Seq*-модели [4] для анализа текстов. Можно ускорить разработку, перенеся лучшие решения оттуда на наш домен, например, архитектуру трансформеров [9] в семействе больших языковых моделей *LLM*.

Исходные данные представляют собой массив данных сигналов телеметрии произвольного размера и в таком «сыром» виде не подходят для обучения нейросети, требуется их существенная предобработка: из них формируется специально структурированный двумерный массив данных, подаваемых на вход для обучения нейросетевой модели парами (датасет). Создание собственного датасета всегда представляет значительную сложность, наша задача не является исключением.

В Таблице 1 видно, что на вход наша модель принимает двумерный тензор размерности (100, 1807), а на выходе ожидает тензор размерности (100, 1337):

Таблица 1

Характеристики используемой модели

Наименование характеристики	Тензор размеоности
Длина входной последовательности	100
Длина выходной последовательности	100
Количество входных признаков	1807
Количество выходных признаков	1337
Латентная размерность	333
Число (рекуррентных) слоев	7
Число параметров	13M

В результате, решение нашей задачи было разделено на пять этапов:

1. Получение данных.
2. Анализ и предобработка данных.
3. Моделирование с помощью нейросети (*RAE*).

4. Оценка аномальности (метрика).
5. Визуализация результатов.

Они реализованы независимо и исполняются последовательно, напоминая конвейер, изображенный на рисунке 3 ниже:

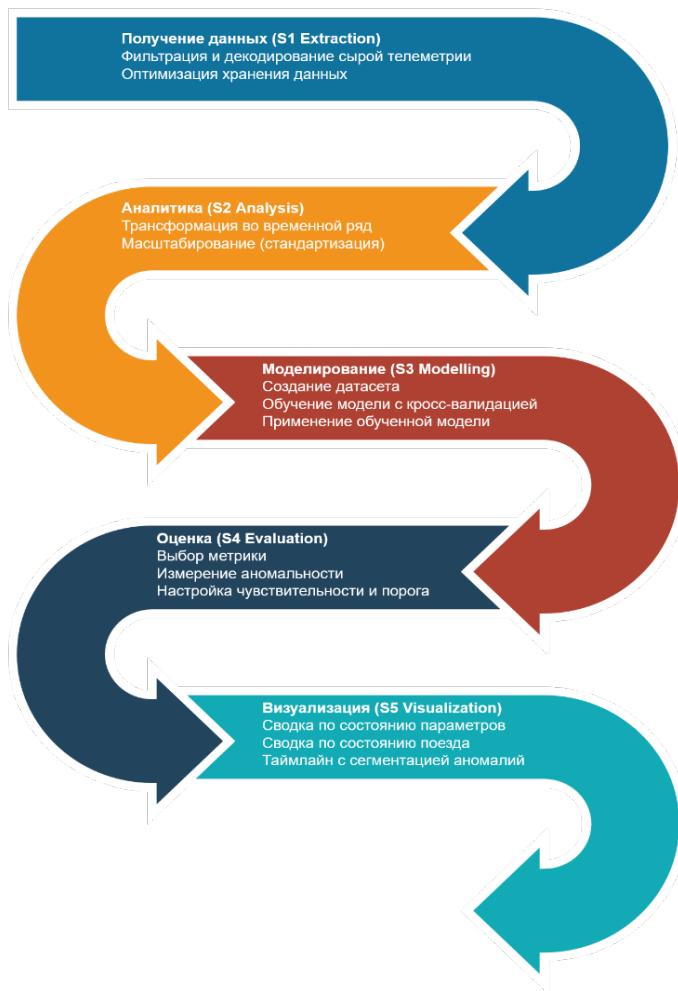


Рис. 3 – Иллюстрация «конвейера данных» с этапами обработки

Процесс обработки данных на борту подвижного состава полностью автоматизирован с помощью протокола информационного обмена. Для промежуточного хранения данных между этапами был выбран иерархический файловый формат *HDF5*, а также разработаны средства для упаковки (архивирования) в него различных метаданных, сопровождающих вычисления. В результате, каждый такой файл содержит информацию, необходимую для следующего этапа обработки.

Это – одно из правил нашего системного дизайна, которое позволяет вести параллельную разработку каждого этапа, а, следовательно, сокращает сроки реализации проекта.

5. Алгоритм обучения и оценка результатов

В качестве функции потерь выбрана *LI (MAE)*, алгоритм ее оптимизации при обучении – *Adam* с начальным шагом *LR*=0.001 и стратегией снижения шага обучения при отсутствии улучшений на валидации (*ReduceLROnPlateau* в *Pytorch*).

Модель решает задачу регрессии, поэтому ключевой метрикой эффективности обучения является ошибка *MAE*, измеряемая на небольшой (10% от всей) отложенной (валидационной)

выборке, которая исключается из данных для обучения.

За 60 эпох обучения эта метрика уменьшилась в 5 раз (Таблица 2) и этот результат достаточно устойчив, что позволяет говорить о том, что модель в первом приближении решает задачу. Дальнейшее совершенствование требует аккуратного подбора гиперпараметров и длительного обучения на подготовленных качественных данных.

Таблица 2

Характерное снижение ошибки (MAE) моделирования на обучающей и валидационной выборках

Epoch	MAE(Train)	MAE(Val)
1	1.18	1.22
10	1.05	1.09
20	0.92	0.94
30	0.74	0.79
40	0.48	0.55
50	0.20	0.32
60	0.13	0.21
70	0.09	0.23

Заметим, что низкая ошибка *MAE* означает, что разработанный алгоритм позволяет достаточно хорошо моделировать 1337 параметров на выходе на протяжении 100 секунд (в каждую секунду). На рисунке 4 представлены один из них (бирюзовый) и предсказание его поведения на 100 секунд вперед (малиновый) на валидационной выборке моделью, а также построен график ошибки *MAE*.

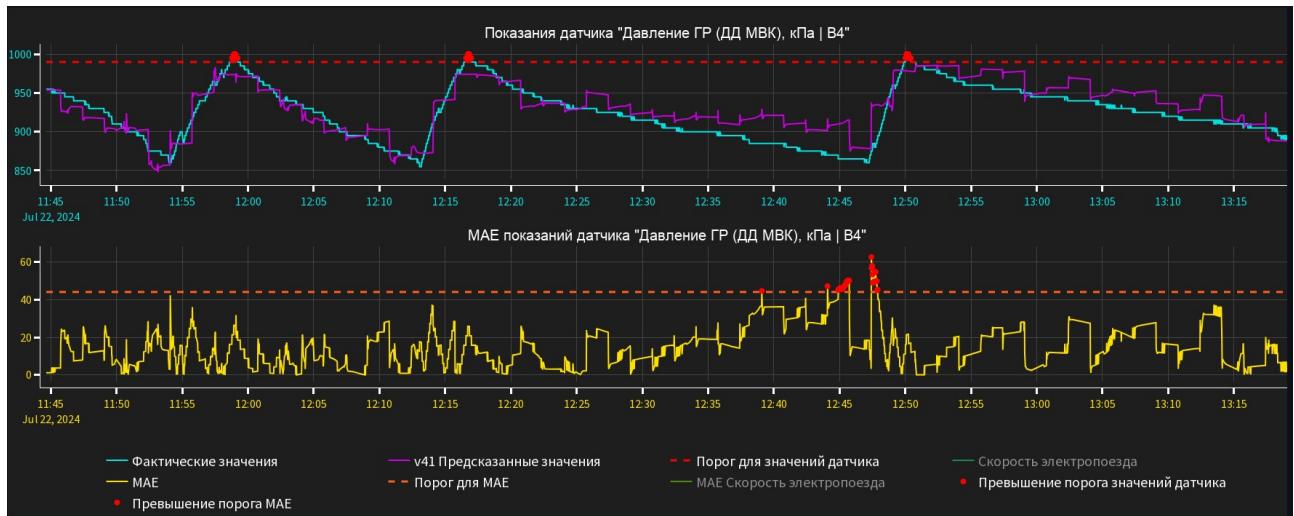


Рис. 4 – Результаты моделирования (v41-идентификатор модели) давления в главном резервуаре в прототипе интерфейса для визуализации (система АИСТ)

Кроме того, отсюда следует, что 1337×100 чисел на входе описываются всего 333 значениями (в Таблице 1 – размерность латентного пространства модели), что означает наличие более простой структуры в реальных данных. На самом деле, это – проявление общей закономерности [8], связанной с тем, что данные получены в ходе реального физического процесса.

Используя авторскую модель в системе ПСД (прескриптивная система диагностики, разработка АО «ВНИИЖТ») был найден сбой в работе системы ПСН (преобразователь

собственных нужд – это инвертор, который преобразует ток высокого напряжения в низкие напряжения для питания различных вспомогательных систем поезда ЭС2Г), который для проверки работы модели был нагружен искусственно (в ходе испытаний). На рисунке 5 ниже заметно, как инвертор был загружен, но эта мощность не питала штатные системы поезда (оборудованные датчиками) и потому проявилась как аномалия.

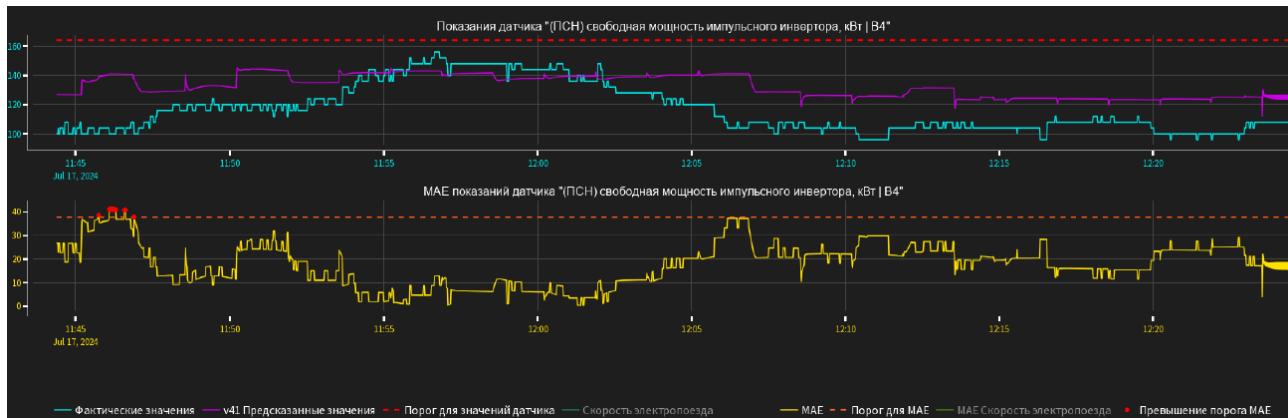


Рис. 5 – Иллюстрация аномалии в показаниях свободной мощности ПСН, найденной в период с 11:45 до 11:47, свидетельствующей об ограничении доступной мощности ПСН

Для решения следующей задачи (предсказания отказов) эффективнее сразу работать с новым, компактным признаковым описанием (z) вместо оригинальных данных. На рисунке 6 показана связь этих двух задач.



Рис. 6 – Иллюстрация связи задачи поиска аномалий с задачей прогнозирования отказов при помощи трансферного обучения

Возьмем только обученный энкодер (E) модели и зафиксируем его веса, а остальное заменим на нейросеть-классификатор (C), выдающую вероятность отказа с последующим её обучением (см. рисунок 5). Это называется [7] трансферным обучением (*Transfer learning*).

Именно так и надо решать задачу предсказания отказов, потому что, если данные для нее размечены, но их мало, обучать модель для предсказания отказов напрямую нерационально: качество будет низким. Об этом будет подробно изложено в одной из следующих публикаций.

Заключение

В ходе данного исследования, реальная проблема предсказания отказов и поиска аномалий обрела формализованную математическую постановку в виде задачи оптимизации на примере оценки технического состояния электропоезда ЭС2Г.

Среди различных алгоритмов машинного обучения было найдено и теоретически обосновано уникальное решение на основе нейросети для поиска аномалий в больших данных. С его помощью были исследованы данные с электропоездом, которые, несмотря на существенные проблемы с качеством, были хорошо описаны предложенной архитектурой модели RAE.

Реализованный подход в системе ПСД (прескриптивная система диагностики, разработка АО «ВНИИЖТ») помог найти сбои в системе ПСН моторвагонного подвижного состава, созданные искусственно в ходе испытаний.

Найденное решение поиска аномалий в поведении сигналов, характеризующих работу подвижного состава, положено в основу решения задачи предсказания отказов.

Разработанный алгоритм и модель RAE обеспечивают автоматизацию разметки предотказного состояния и формирование классификатора отказов.

В будущем, предлагаются усовершенствовать математическую модель НС и алгоритм аналитики, используя подходы [9] из области NLP, продолжить поиск лучших архитектур нейросети, оптимизацию гиперпараметров и совершенствование статистических метрик аномальности.

Список использованной литературы

1. D. Hawkins, *Identification of Outliers*. London: Chapman and Hall, 1980
2. J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proc. Nat. Acad. Sci. U.S.A.*, vol. 79, no. 8, pp. 2554–2558, 1982.
3. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
4. N. Srivastava, E. Mansimov, and R. Salakhutdinov, “Unsupervised Learning of Video Representations using LSTMs,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 843–852.
5. J. F. Kolen and S. C. Kremer, *A Field Guide to Dynamical Recurrent Networks*. New York: Wiley-IEEE Press, 2001.
6. I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” in *Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3104–3112.
7. A. Sagheer and M. Kotb, “Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems,” *Sci. Rep.*, vol. 9, p. 19038, 2019.
8. N. Whiteley, A. Gray, and P. Rubin-Delanchy, “Statistical exploration of the Manifold Hypothesis,” *J. Amer. Stat. Assoc.*, 2025. (Early Access).
9. A. Vaswani et al., “Attention is all you need,” in *Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 5998–6008.
10. K. Besbes et al., “Sizing of an electric powertrain using a hybrid model combining Simcenter Amesim and Modelica components,” in *Proc. IEEE Veh. Power Propuls. Conf. (VPPC)*, 2020, pp. 1–6.