

УДК 004.032.26:658.581

Применение больших языковых моделей для предсказания отказов оборудования

Долгачев Даниил Владимирович

daniil2012dol@yandex.ru

МГТУ им. Н.Э. Баумана, Москва, Россия

В статье рассматривается применение больших языковых моделей (Large Language Models, LLM) в задачах предсказания отказов промышленного оборудования. Особое внимание уделено анализу существующих методов прогнозирования, а также роли языковых моделей в обработке и интерпретации текстовых данных, содержащихся в технической документации, отчетах и журналах обслуживания. Описываются проблемы интеграции неструктурированных источников информации с сенсорными и эксплуатационными данными, а также сложности, возникающие при анализе больших объемов разнородных данных в производственной среде. Показана актуальность использования современных моделей обработки естественного языка для повышения эффективности мониторинга состояния оборудования и поддержки принятия решений в системах предиктивного обслуживания.

Ключевые слова: большие языковые модели, предсказание отказов, промышленное оборудование, машинное обучение, нейронные сети, анализ текстовых данных, предиктивное обслуживание

Введение. В современном производстве повышение надежности оборудования является ключевой задачей. Отказы машин приводят к значительным экономическим потерям и простоям.

Отказы относятся к малоизученным явлениям. Это объясняется, прежде всего, тем, что время возникновения отказа зависит от большого числа случайных факторов, его трудно исследовать и еще труднее измерить. Время возникновения отказа или время работы системы между отказами представляют собой явления случайные [1].

В данной статье рассматривается возможность использования языковых моделей, для предупреждения аварийных ситуаций на производстве.

Материалы и методы. Для решения данной проблемы активно внедряются методы предиктивного обслуживания (Predictive Maintenance), основанные на анализе потоков данных и прогнозировании возможных отказов.

Предиктивное обслуживание или PdM (Predictive maintenance) является одним из передовых подходов к управлению технологическим процессом на производстве. Это часть развития индустрии 4.0, больших данных (BigData) и Интернета вещей (IoT), в нем используются новейшие приложения искусственного интеллекта (AI) и машинного обучения (ML) [2].

Нижний уровень — это сбор данных. Данные собираются при помощи датчиков и передаются на уровень выше — ПЛК. В свою очередь ПЛК обрабатывает данные и передает на сервер, который уже складывает их в БД. Да-

лее в работу включается система прогнозирования, основанная на предиктивной аналитике. Система мониторинга — является инструментом отображения всех параметров в онлайн режиме и дает возможность вести контроль человеку.



Рис. 1. Структурная схема системы с предиктивным обслуживанием

Предиктивная аналитика — это прежде всего множество методов статистики, анализа данных и теории игр, которые используются для анализа текущих и исторических данных/событий для прогноза данных и событий в будущем [3].

В данной статье рассмотрим применение больших языковых моделей. LLM (Large Language Model, большая языковая модель) — это тип модели машинного обучения (или тип модели глубокого обучения), основанный на больших объемах данных, который может выполнять любые задачи по обработке естественного языка (Natural Language Processing, NLP) [4].

Большие языковые модели предоставляют возможность интеграции текстовой и числовой информации, что делает их эффективным инструментом в промышленной аналитике. Они показали свои возможности в понимании естественного языка, обработке текстовых данных, извлечении ключевой информации формировании ответов на сложные вопросы.

Большая языковая модель — это разновидность искусственной нейронной сети, которая обучается на больших объемах данных для генерации и понимания естественного языка [5].

Это нейронная сеть с крайне большим количеством изменяемых параметров, которая позволяет решать задачи по обработке и генерации текста. Чаще всего реализована в виде диалогового агента, с которым можно общаться в разговорной форме. LLM обучают на огромных наборах данных, они построены на машинном обучении: в частности, на базе нейронной сети, называемой моделью-трансформером. Она способна выделять упоминания факторов, распознавать причинно-следственные связи и сопоставлять их с регламентами эксплуатации. Это позволяет переводить технические журналы и отчеты в понятные описания состояния оборудования. Тем самым повышается полнота картины и уменьшается время от появления скрытого дефекта до его формализованного диагностического заключения.

Интеграция LLM в производственную среду осуществляется в виде дополнительного интеллектуального уровня анализа, взаимодействующего с существующими системами мониторинга и управления (SCADA, MES, ERP).



Рис. 2. Структурная схема системы с использованием большой языковой модели

Общая архитектура такая же, как при применении системы прогнозирования. Меняем систему прогнозирования на LLM модель.

Для лучшего прогнозирования состояния необходимо использовать исторические данные о ремонтах, времени работы узлов и параметрах эксплуатации. На их основе можно определить закономерности, предшествующие сбо-

ям. Система может заранее предупредить персонал о вероятном износе детали, снизив риск внезапной остановки оборудования. LLM формирует рекомендации по устранению неисправностей, опираясь на базы данных инструкций и предыдущие случаи. Модель адаптирует свои ответы под конкретный контекст — тип оборудования, место установки, историю обслуживания. Модель способна составлять отчеты, заявки на ремонт, объяснительные записки и сводные аналитические документы, тем самым сокращая время на подготовку документации.

В рамках производственного процесса LLM может быть интегрирован в корпоративный чат или веб-интерфейс, где сотрудники формулируют запросы в естественном языке, а модель предоставляет точные и релевантные ответы, включая ссылки на документы. Для этого используется хранилище RAG (Retrieval Augmented Generation). Это может повысить уровень правильности выполнения работ.

Для примера проанализируем 36 центробежных насосов. Источники данных представленные в табл. 1 включали анализ: вибрации, температуры, давления и силы тока, журналы событий и регламентную документацию. Данные собирались за период 24 месяца.

Таблица 1

Данные для анализа

Источник	Показатели	Частота сбора
Датчики	Вибрация, Температура, давление, сила тока	1 раз в минуту
Журналы событий	Ошибки, предупреждения, перезапуски	По событию
Документация	Глоссарий	Единоразово (при обновлении)

Перед загрузкой числовых данных, в LLM был загружен глоссарий, что бы исключить путаницу с определениями в рамках задачи.

Числовые данные с датчиков были очищены от шумов и нормализованы. Текстовые данные из журналов были загружены в векторную базу данных для использования в механизме RAG (Retrieval Augmented Generation).

Языковая модель выполняла функции интерпретации сигналов и текстов в режиме Retrieval Augmented Generation. Для каждого инцидента формировалась сводка последних значений телеметрии и критическое событие. Модель генерировала объяснение причин возможной аномалии, перечень проверок и конкретные действия, а также приоритет происшествия. Ответ записывался базу и отправлялся для подтверждения ситуации.

Для тестирования работоспособности системы в условиях, приближенных к реальным, были сгенерированы синтетические датасеты, имитирующие поведение 36 центробежных насосов. Данные включали в себя все парамет-

ры: вибрация, температура, давление, сила тока. В данных были заданы аномалии для проверки того, сможет ли определить система будущую остановку оборудования. Эти данные передавались в систему не единым пакетом, а потоково, через REST API, что имитировало работу с реальными источниками в режиме, близком к реальному времени. В табл. 2 наглядно представлены данные сколько раз система смогла предсказать поломку, сколько не смогла и сколько раз галлюцинировала.

По представленным данным можно сделать вывод, что LLM способна предсказывать более аварийных 85 % ситуаций при постепенном приближении к ним. В случае с неожиданными остановками, она показала себя значительно хуже. Для резких остановок не хватает контекста и тестовых данных, если на настоящем производстве будет вестись журнал состояний и проверок оборудования, система справится с анализом ситуаций значительно лучше, так как будет большее количество релевантных данных. И в рамках тестирования было зафиксировано 4 галлюцинации, когда оборудование по данным работало, но система говорила об их будущей остановке. Показатель галлюцинаций не сильно высокий. Над понижением процента галлюцинаций нужно работать, дообучая модель на разных данных, что в рамках данной работы нельзя было сделать.

Таблица 2

Отчет по работе LLM

Всего аварийных остановок		Предсказанных точно		Не предсказанных		Галлюцинации
Плавных	Резких	Плавных	Резких	Плавных	Резких	
100	20	86	2	14	18	4

Пример с центробежными насосами наглядно демонстрирует, что LLM не только способны генерировать текстовые отчеты, но и выполнять сложную аналитическую работу, выявляя скрытые корреляции и формируя обоснованные прогнозы. Это открывает новые возможности для создания интеллектуальных, объяснимых и адаптивных систем управления техническим состоянием, что в конечном итоге ведет к повышению надежности, снижению затрат на ремонт и минимизации простоев в промышленном производстве.

Дальнейшие исследования в этой области могут быть направлены на оптимизацию архитектуры RAG для промышленных нужд, разработку более эффективных методов дообучения моделей на узкоспециализированных данных и создание стандартизированных интерфейсов для интеграции LLM с существующими MES и ERP-системами.

Заключение. В статье представлен подход интеграции языковой модели в производственные процессы для предотвращения аварийных ситуаций. Показано, что совместная обработка текстовых источников повышает полноту

анализа и сокращает время реакции. Формальная постановка задачи и набор метрик образуют основу для объективной оценки эффекта. Встроенная в контур SCADA языковая модель обеспечивает интерпретацию телеметрии и текстовых источников, что ускоряет принятие решений и снижает аварийность. Представленная архитектура и порядок подготовки данных позволяют быстро запустить пилот на ограниченном участке и масштабировать решение на весь парк оборудования. Полученные результаты позволяют рекомендовать поэтапное внедрение подхода с началом на ограниченном участке, с последующим расширением и совершенствованием процедур контроля состояния оборудования.

Литература

- [1] Алексеев А.В. *Диагностика и надежность автоматизированных систем*. URL: <https://repo.ssau.ru/bitstream/Uchebnye-izdaniya/Diagnostika-i-nadezhnost-avtomatizirovannyh-sistem-Elektronnyi-resurs-elektron-obrazovat-kontent-v-sisteme-distanc-obucheniya-Moodle-71193/1/Алексеев%20А.%20В.%20Диагностика.pdf> (дата обращения 15.09.2025).
- [2] Фомин А.И., Игнатъева Е.А. *Предиктивное обслуживание и интернет вещей в системах охраны труда*. URL: <https://science.kuzstu.ru/wp-content/Events/Conference/RM/2023/RM23/pages/Articles/010107.pdf> (дата обращения 15.09.2025).
- [3] Веревкин А.П. *Предиктивная аналитика*. URL: https://rusoil.net/files/sveden_rpd/7242/mm/Предиктивная%20аналитика/УП%20Веревкин%20Пред%20аналит%20лек.pdf (дата обращения 01.07.2024).
- [4] Абраменко Г. *Большой обзор больших языковых моделей*. URL: <https://habr.com/ru/companies/gaz-is/articles/884410/> (дата обращения 01.07.2024).
- [5] ГОСТ Р 71657–2024. *Национальный стандарт Российской Федерации. Технологии искусственного интеллекта в образовании. Функциональная подсистема создания научных публикаций. Общие положения*. Москва, Российский институт стандартизации, 2024.

Using large language models to predict hardware failures

Dolgachev Daniil Vladimirovich

daniil2012dol@yandex.ru

BMSTU, Moscow, Russia

The article examines the application of Large Language Models (LLM) in the task of predicting failures of industrial processing. Particular attention is paid to the analysis of existing forecasting techniques, as well as the role of language models in the processing and interpretation of text data contained in technical documentation, reports and maintenance logs. The problems of integration of unstructured data sources with sensory and operational data are described, as well as the difficulties that arise when analyzing large volumes of heterogeneous data in a production environment. The relevance of using modern models of natural language processing to improve the efficiency of monitoring the condition of equipment and support decision-making in systems of preventive service is shown.

Keywords: *large language models, failure prediction, industrial upgrading, machine learning, neural networks, text data analysis, predictive maintenance*