

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.383

Борисюк
Фёдор Александрович

Алгоритмы и программные средства анализа данных
в предизолированных системах тепло- и водоснабжения

АВТОРЕФЕРАТ
на соискание степени магистра
по специальности 1-40 80 04 – Информатика и технологии
программирования

Научный
руководитель
Батура М.П.
д.т.н., профессор

Минск 2024

КРАТКОЕ ВВЕДЕНИЕ

В современном мире концепция Интернета вещей (IoT), охватывающая умные устройства и сенсоры, играет ключевую роль в мониторинге различных систем, включая те, которые обеспечивают тепло и водоснабжение в городах. Особенно важно следить за состоянием предизолированных трубопроводов, так как они несут ответственность за передачу тепла и воды в различные точки города.

Выбор данной темы обусловлен стремлением улучшить качество жизни в городах и повысить экономическую эффективность систем жизнеобеспечения. Прогнозирование состояния трубопроводов является важной задачей, поскольку от этого зависит надежность и бесперебойность снабжения теплом и водой. Эффективное управление этими системами позволяет не только своевременно выявлять и устранять потенциальные проблемы, но и оптимизировать расходы на обслуживание и ремонт инфраструктуры.

Диссертационная работа направлена на разработку и анализ методов прогнозирования состояния трубопроводов на основе данных операционного контроля. В рамках исследования изучаются различные подходы к прогнозированию, включая традиционные методы, такие как модели ARIMA, а также современные методы, использующие нейронные сети, такие как рекуррентные нейронные сети (RNN). Основные задачи исследования включают анализ текущих методов мониторинга и выбор подходящей архитектуры нейронной сети для прогнозирования данных.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка и анализ алгоритма прогнозирования данных, получаемых от системы оперативного дистанционного контроля предизолированных систем тепло- и водоснабжения, с использованием нейронных сетей. Основная задача заключается в создании эффективного и точного метода, способного автоматически предсказывать параметры состояния теплотрассы.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести исследование и изучение различных методов прогнозирования данных, включая как классические подходы (например, модель ARIMA), так и современные методы, основанные на нейронных сетях (например, рекуррентные нейронные сети). Это позволит определить преимущества и недостатки существующих решений и выявить наиболее перспективные подходы.

2. Создать архитектуру нейронной сети, которая будет способна прогнозировать параметры состояния теплотрассы на основе временных рядов данных, получаемых от СОДК. Это может включать комбинацию GRU и LSTM слоев для учета как краткосрочных, так и долгосрочных зависимостей в данных.

3. Предварительно обработать набор данных, содержащий записи с интервалом в 15 минут. Может потребоваться очистка данных, метод которой необходимо выбрать, а также их нормализация для улучшения качества моделирования.

4. Обучить разработанную архитектуру нейронной сети на подготовленном наборе данных, применяя методы оптимизации и регуляризации для достижения высокой точности прогнозирования параметров состояния теплотрассы.

5. Спроектировать, реализовать и развернуть систему, включающую базу данных для хранения временных рядов и средство визуализации для мониторинга.

Объектом исследования являются алгоритмы прогнозирования данных с помощью нейронных сетей.

Предметом исследования является разработка и оптимизация алгоритмов прогнозирования данных от системы оперативного дистанционного контроля предизолированных систем тепло- и водоснабжения с использованием нейронных сетей.

Основной *гипотезой*, положенной в основу диссертационной работы, является предположение, что внедрение глубоких нейронных сетей с архитектурами GRU и LSTM в алгоритмы прогнозирования данных повысит точность и эффективность процесса прогнозирования по сравнению с традиционными методами. Глубокие нейронные сети, такие как GRU и LSTM, показали высокую эффективность в задачах анализа временных рядов и прогнозирования, что делает их перспективными для использования в данной области. Использование GRU позволяет модели улавливать краткосрочные зависимости, в то время как LSTM слои эффективно моделируют долгосрочные зависимости в данных.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя М. П. Батуры заключается в формулировке целей и задач исследования

Публикации результатов диссертации

По теме диссертации опубликовано 3 печатные работы, 3 работы в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырёх глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования и показаны направления их решения. Во второй главе рассмотрены алгоритмы обработки данных от системы операционно-дистанционного контроля, включая типы данных, предоставляемых системой, методы предобработки данных и прогнозирования — как классические, так и современные методы на основе нейронных сетей. Третья глава посвящена созданию нейросетевого алгоритма прогнозирования параметров состояния теплотрассы, включая описание процесса подготовки датасета, создание и компиляцию модели, а также обучение и анализ результатов модели. В четвертой главе изложены принципы проектирования и реализации системы обработки данных, включая выбор баз данных, средства визуализации и брокера сообщений, описание архитектуры системы, процесс её работы, настройку и

развертывание приложения на облачной платформе, а также взаимодействие с Grafana для настройки уведомлений и мониторинга состояния теплотрассы в реальном времени.

Общий объем работы составляет 65 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ

Введение является вступительным разделом работы, в котором осуществляется обоснование выбора темы исследования, описание мотивации исследователя, а также постановка целей и задач работы. Введение создает контекст и подчеркивает значимость выбранной темы, объясняя причины, по которым она является актуальной и важной для исследования. Кроме того, во введении определяются цели исследования, а также формулируются конкретные задачи, направленные на достижение данных целей.

Первая глава представляет собой аналитический обзор методов получения данных в предизолированных системах. Глава начинается с рассмотрения понятия предизолированных систем тепло- и водоснабжения. Описаны основные характеристики и преимущества предизолированных труб, включая надежность, устойчивость к внешним воздействиям, эффективность теплоизоляции, экономичность, долговечность, простоту монтажа, минимальное обслуживание, гибкость и адаптивность, а также экологическую безопасность.

Затем проводится анализ существующих методов мониторинга и диагностики состояния изоляции трубопроводов. Рассматриваются традиционные методы, такие как визуальный осмотр, измерение сопротивления изоляции, ультразвуковая и тепловизионная диагностика. Каждый метод описывается с точки зрения его преимуществ и ограничений.

Далее в главе описана система операционно-дистанционного контроля (СОДК), применяющаяся для мониторинга сопротивления изоляции трубопроводов. Описаны два типа датчиков: переносной и стационарный, а также их функции и принцип работы. Рассматриваются задачи и методы, используемые датчиками для контроля влажности и целостности теплоизоляционного покрытия. Обсуждаются этапы и методы ремонта поврежденных участков трубопровода.

Целью первой главы является предоставление обзора существующих методов и подходов к мониторингу и диагностике состояния теплоизоляции в предизолированных системах, а также анализ преимуществ и ограничений различных методов контроля и систем автоматического мониторинга.

Вторая глава посвящена технологии СОДК (системы оперативного дистанционного контроля), играющей ключевую роль в мониторинге и обслуживании теплотрассы. Она предоставляет обширные данные о различных параметрах теплотрассы, что позволяет оперативно выявлять потенциальные проблемы и предотвращать аварийные ситуации.

В главе описываются функции и основные типы данных, которые СОДК предоставляет для мониторинга состояния теплотрассы. Эти данные включают

уровень заряда батарей датчиков, сопротивление сигнального проводника и изоляции, а также идентификационные номера датчиков и временные метки замеров, необходимые для динамического анализа параметров.

Важным аспектом главы является использование медианного сглаживания данных для предобработки информации от СОДК. Этот метод позволяет устранять шумы и выбросы в данных, что важно для точной и надежной диагностики состояния теплотрассы и прогнозирования её работы.

Цель главы состоит в том, чтобы продемонстрировать эффективность и значимость технологии СОДК в обеспечении надежности и безопасности работы теплотрассы. Анализ и прогнозирование данных, основанные на методах медианного сглаживания и ARIMA, направлены на оптимизацию обслуживания и управления тепловыми сетями, что способствует повышению их эффективности и долговечности.

Третья глава посвящена реализации нейросетевого алгоритма прогнозирования временных рядов, специально разработанного для мониторинга и поддержки состояния теплотрассы. Она начинается с подробного описания процесса подготовки датасета, который является критическим этапом в разработке и обучении модели.

В разделе "Подготовка датасета" представлены ключевые шаги: от выбора исходного датасета, состоящего из записей с интервалом в 15 минут, до очистки данных методом медианного сглаживания и разделения на обучающую, тестовую и валидационную выборки.

Выбор исходного датасета осуществлялся с учетом необходимости включения разнообразных параметров, таких как заряд батареи датчика, данные о сопротивлении сигнального проводника и изоляции теплотрассы, что важно для полноценного мониторинга и оперативного выявления потенциальных проблем. Очистка данных с использованием медианного сглаживания с размером окна 96 направлена на устранение шумов и выбросов, при этом сохраняя основные тренды и паттерны временных рядов.

После этапа подготовки данные преобразовываются для обучения модели временных рядов с использованием алгоритмов GRU и LSTM. В разделе "Создание и компиляция модели" описана структура нейронной сети, начиная с добавления слоев GRU и LSTM с применением регуляризации L2 для улучшения обобщающей способности модели. Dropout слои использованы для предотвращения переобучения, что повышает устойчивость и надежность модели в обучении на данных различных временных интервалов.

В разделе "Обучение модели" представлен процесс обучения, включая выбор функции потерь MSE и оптимизатора Adam. Модель настроена на минимизацию ошибки прогнозирования на обучающем наборе данных, с

последующей проверкой обобщающей способности на валидационном наборе. Использование механизма ранней остановки помогает предотвратить переобучение модели, что подтверждается стабильными результатами на тестовых данных.

Цель третьей главы заключается в разработке и реализации нейросетевой модели для прогнозирования временных рядов с целью мониторинга и поддержки состояния теплотрассы. Основные задачи включают подготовку качественного датасета, создание и обучение модели с использованием алгоритмов GRU и LSTM, а также оценку точности и обобщающей способности модели на различных наборах данных. Глава стремится представить полный цикл разработки и тестирования модели с упором на её эффективность в реальных условиях мониторинга тепловых сетей.

Четвёртая глава посвящена проектированию архитектуры и реализации системы обработки данных, необходимой для эффективного управления данными, полученными с датчиков. Целью главы является представление комплексной архитектуры, способной обеспечить надежную и масштабируемую обработку данных в режиме реального времени, а также их визуализацию для пользователей системы.

В разделе "Проектирование архитектуры" описывается структура системы, которая включает несколько ключевых компонентов. Система состоит из слоя ввода данных, инфраструктурного слоя с брокером сообщений RabbitMQ, слоя бизнес-логики с API-сервисом и анализатором данных, а также слоя данных с использованием реляционной базы данных PostgreSQL и специализированной базы данных временных рядов InfluxDB. Каждый компонент выполняет определенные функции, направленные на обеспечение надежности, масштабируемости и высокой производительности системы.

Далее в разделе "Выбор баз данных, средства визуализации и брокера сообщений" рассматривается выбор технологий для каждого компонента системы. В качестве реляционной базы данных выбран PostgreSQL, благодаря его поддержке ACID-транзакций и высокой производительности при OLTP. Для хранения временных рядов выбрана InfluxDB, специализированная база данных, оптимизированная для работы с временными данными и метриками, что соответствует потребностям проекта. Для визуализации данных выбран Grafana, предоставляющий широкие возможности по созданию дашбордов и мониторингу временных рядов. В качестве брокера сообщений выбран RabbitMQ, обеспечивающий надежную и асинхронную передачу данных между компонентами системы.

Глава завершается разделом "API-сервис", где описывается роль и функции сервиса, обеспечивающего стандартизированный прием данных от

СОДК по HTTP-запросам. Для обработки запросов выбран язык программирования Go, который обеспечивает эффективную многопоточность и высокую производительность, необходимые для работы с множеством одновременных запросов. Для взаимодействия с реляционной базой данных используется ORM библиотека GORM, упрощающая работу с данными и поддерживающая миграции схемы. В целом, API-сервис играет ключевую роль в обеспечении надежной передачи данных в систему обработки и анализа.

Целью четвертой главы является проектирование архитектуры и реализация системы обработки данных для эффективного сбора, хранения, обработки и визуализации данных, полученных от датчиков. Глава охватывает несколько ключевых аспектов, начиная с описания архитектуры системы, включающей слои ввода данных, инфраструктуры, бизнес-логики и вывода данных. Далее рассматривается выбор баз данных (PostgreSQL для реляционных данных и InfluxDB для временных рядов), средства визуализации (Grafana) и брокера сообщений (RabbitMQ), на основании их спецификаций и преимуществ для данного проекта. В заключение главы представлены API-сервис и сервис обработки данных, описывающие работу системы от приёма данных от датчиков до анализа и прогнозирования временных рядов, с подробным описанием используемых технологий и инструментов.

ЗАКЛЮЧЕНИЕ

В данной работе проведено исследование различных методов и алгоритмов, применяемых для обработки и прогнозирования данных, получаемых от системы оперативного дистанционного контроля в предизолированных системах тепло- и водоснабжения. Были проанализированы существующие методы мониторинга и диагностики, а также подробно рассмотрены алгоритмы обработки данных, предоставляемых датчиками СОДК. В рамках исследования были выделены и описаны типы данных, предоставляемых системой, и предложены методы их предобработки, включая медианное сглаживание и нормализацию. Особое внимание уделено методам прогнозирования, таким как экспоненциальное сглаживание, модель ARIMA и рекуррентные нейронные сети (RNN). На основе анализа различных подходов был выбран комбинированный метод с использованием GRU и LSTM слоев для создания нейросетевой модели, обеспечивающей высокую точность и стабильность прогнозов.

Реализация нейросетевого алгоритма прогнозирования включала несколько ключевых этапов: подготовку и очистку датасета, разделение данных на обучающую, тестовую и валидационную выборки, нормализацию данных и создание модели. Модель была обучена и продемонстрировала низкие значения потерь на тестовой (0.55%) и валидационной (0.62%) выборках, что подтверждает её высокую точность и эффективность.

Проектирование архитектуры системы обработки данных включало выбор баз данных, средств визуализации и брокера сообщений. В процессе разработки были рассмотрены различные программные средства, включая реляционные базы данных и базы данных временных рядов, инструменты визуализации, такие как Grafana, и брокеры сообщений для обеспечения надежного и масштабируемого обмена данными.

Для дальнейшего улучшения системы рекомендуется рассмотреть возможность её развёртывания в облаке, например, на платформе AWS. Это позволит обеспечить более высокую масштабируемость и доступность, а также улучшить управление ресурсами и безопасность данных. Облачные решения предоставляют гибкость и возможность автоматического масштабирования в зависимости от нагрузки, что является критически важным для систем мониторинга и анализа больших объемов данных. Также целесообразно настроить систему уведомлений из Grafana. Это позволит оперативно информировать ответственных сотрудников о любых отклонениях или критических изменениях в данных, получаемых от СОДК. Настройка

уведомлений поможет своевременно реагировать на возникающие проблемы и поддерживать стабильную работу системы. Достижения:

1 Низкие потери данных: В ходе исследования и разработки нейросетевой модели прогнозирования были достигнуты низкие значения потерь на тестовой (0.55%) и валидационной (0.62%) выборках. Это свидетельствует о высокой эффективности модели в точном воспроизведении данных и минимизации ошибок в прогнозах.

2 Высокая точность прогнозирования: Разработанная комбинированная нейросетевая модель с использованием GRU и LSTM слоев обеспечивает высокую точность прогнозов. Это достижение подтверждает возможность применения сложных алгоритмов обработки данных для достижения целей прогнозирования в системе оперативного дистанционного контроля.

3 Спроектирована архитектура системы обработки данных: В процессе проектирования были выбраны и интегрированы различные технологии, включая базы данных (PostgreSQL, InfluxDB), средства визуализации (Grafana) и брокеры сообщений (RabbitMQ). Это обеспечивает надежную основу для сбора, хранения, обработки и визуализации данных от системы оперативного дистанционного контроля.

Проблемы:

1 Высокое загрязнение данных в большинстве датасетов: Одной из ключевых проблем является высокая степень загрязнения данных, получаемых от датчиков системы оперативного дистанционного контроля. Это требует дополнительной работы по очистке и предобработке данных для обеспечения их качества и достоверности.

2 Высокая вероятность воздействия на датчики извне: Система подвержена высокому риску воздействия извне на датчики, что может привести к искажению или потере данных. Это требует усиленных мер по защите системы от внешних воздействий и обеспечения целостности собираемых данных.

3 Высокое потребление ресурсов системой: Анализатор в Kubernetes требует значительных ресурсов для обучения модели каждый раз, что может привести к неэффективному использованию вычислительных мощностей. Необходимо оптимизировать процесс обучения и развертывания моделей для снижения нагрузки на систему и повышения её эффективности.

Эти достижения и проблемы являются важными аспектами диссертационной работы и требуют дальнейших исследований и разработок для улучшения системы оперативного дистанционного контроля в предизолированных системах тепло- и водоснабжения.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1 Борисюк Ф.А. ПРИМЕНЕНИЕ КРОСС-ВАЛИДАЦИИ В ОПТИМИЗАЦИИ МОДЕЛИ ХОЛЬТА-ВИНТЕРСА ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ / Ф.А. Борисюк // Технические науки: проблемы и решения: сб. ст. по материалам LXXII Международной научно-практической конференции «Технические науки: проблемы и решения». – № 5(67). – М., Изд. «Интернаука», 2023.

2 Борисюк Ф.А. ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ НЕЙРОННОЙ СЕТИ GRU / Ф.А. Борисюк // Инновационные подходы в современной науке: сб. ст. по материалам CLXVIII Международной научно-практической конференции «Инновационные подходы в современной науке». – № 12(168). – М., Изд. «Интернаука», 2024.

3 Борисюк Ф.А. СТРУКТУРА LSTM-СЛОЯ НЕЙРОННОЙ СЕТИ / Ф.А. Борисюк // Технические науки: проблемы и решения: сб. ст. по материалам LXXXVI Международной научно-практической конференции «Технические науки: проблемы и решения». – № 7(80). – М., Изд. «Интернаука», 2024.