Санкт–Петербургский государственный университет

КОВЛАЕНКО Лев Алексеевич

Выпускная квалификационная работа

Детектирование и классификация дефектов нефтепроводов на основе данных внутритрубных роботов-дефектоскопов

Уровень образования: бакалавриат
Направление 01.03.02 «Прикладная математика и информатика»
Основная образовательная программа СВ.5005.2016 «Прикладная математика, фундаментальная информатика и программирование»
Профиль «Исследование и проектирование систем управления и обработки сигналов»

Научный руководитель:
доцент, кафедра технологии программирования,
к.т.н. Блеканов Иван Станиславович

Рецензент:
Начальник отдела математического моделирования перспективных направлений, ООО ИТСК,
Кононов Ярослав Сергеевич

Санкт-Петербург
2020 г.

Содержание

Введение

0.1 Актуальность

В промышленности на сегодняшний день появилось огромное количество программных комплексов и информационных систем позволяющих автоматизировать и оптимизировать производство. Такая тенденция подталкивает различные промышленные компании к внедрению инноваций в технологические процессы и проведение различных исследований.

Большой толчок в развитии получили многие области промышленности, например автомобильная промышленность или нефтегазовая. Примером этому может служить то, что за последние несколько лет в группе компаний Газпром Нефть были внедрены инновационные решения для оптимизации процессов [https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2may/2989334/]:

- Инструмент для планирование очистки призабойной зоны для увеличения точности прогнозирования ожидаемого эффекта.
- Проект “Умная логистика” позволяет оптимизировать логистику поставки нефти и упростить контроль.
- Информационной система «ЭКСПРЕСС-ОЦЕНКА ГЕОЛОГИЧЕСКОГО СТРОЕНИЯ И ВЕРИФИКАЦИЯ ДАННЫХ», автоматизирующая процесс оценки геологического строения и верификации данных, построение карт, таблиц, схем геологических характеристик активов компании. [http://it-sk.ru/products/]

Кроме внедрения крупных проектов, стоит отметить рост количества научно-исследовательских и опытно-конструкторских работ. С 2016 по 2018 их количество увеличилось в 24 раза. Исследования в компании Газпом Нефть проводятся по 14 различным направлениям [https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2018-june/1715829/].

Одним из важных направлений для автоматизации является оценка состояния нефтепроводов [https://cyberleninka.ru/article/n/analiz-tehnicheskogo-

sostoyaniya-obektov-lineynoy-chasti-magistralnyh-nefteprovodov-opredelenie-optimaln
sposobov-podderzhaniya/viewer]. В группе компаний Газпром Нефть эксплуатируются примерно 12 тыс. км промысловых трубопроводов. В ходе эксплуатации нефтепроводы неизбежно изнашиваются. Это приводит к прорывам и утечкам нефти, что вызывает отрицательные экономический, экологический и репутационный эффекты. Во избежание этого проводятся работы по внутритрубной диагностике - съемке магнитограмм, на основе которых эксперты могут оценить состояние нефтепровода. Следующим этапом по развитию этого направления является автоматизация процесса оценки состояния трубопровода с целью уменьшения затрат по временным ресурсам и увеличению качества процесса

0.2 Цель работы

Разработать программный комплекс, позволяющий производить автоматическую оценку состояния трубопровода, на основе анализа снимков магнитограмм.

0.3 Задачи работы

Для достижения цели были поставлены следующие задачи:

- Провести обзор существующих решений.
- Собрать и подготовить данные для дальнейших исследований: перевести магнитограммы из бинарного формата в табличный, провести точную разметку на основе представленных отчетов о состоянии нефтепроводов.
- Рассмотреть различные подходы и методы машинного обучения для детектирования швов и дефектов на различных магнитограммах.
- Разработать собственный подход анализа в виде модификации и комбинаций существующих алгоритмов.
- Выбрать стек технологий и спроектировать архитектуру программного комплекса.

- Разработать программное решение позволяющее проводить оценку состояния трубопровода.
- Провести тестирование ПО и оценку качества разработанного подхода.

0.4 Практическая значимость

Глава 1. Обзор существующих решений и подходов в области анализа состояния инженерных сооружений.

1.1 Обзор технологических решений

На сегодняшний день большое количество нефтегазовы сервисных компаний предлагает проведение диагностики состояния трубопровода с использованием магнитных и ультразвуковых устройств. Для проведения мониторинга состояния трубопроводов используются автономные зонды (Рис. 1.), оборудованные профилометрами и магнитными дефектоскопами. Такой зонд обычно имеет цилиндрическую форму, и по его бокам через равные угловые промежутки установлены измерительные блоки. Во время сервисных операций автономный зонд загружается в нефтепровод, а затем перемещается внутри вместе с нефтяным потоком, попутно замеряя магнитную индукцию вдоль стенок трубы. Полученные данные записываются в виде магнитограммы. Для ультразвуковой диагностики зон оснащается другим типом датчиков основанных на принципах ультразвукового замера толщины.

В сервисный компании “Baker Huges” предлагаются услуги по проведению магнитной и ультразвуковой внутритрубной дефектоскопии. [https://www.b
Для провдения диагностики используются магнитный и ультрозвуковой дефектоскопы. Для интерпретации результатов дефектоскопии разработано ПО внутреннего пользования, позволяющее визуализировать магнитограмму и проводить ручную разметку.

Компания “Интрон плюс” также предлагает внутритрубную диагностику на основе данных магнитных дефектоскопов. [https://www.intron.ru/]

Рис. 1: Зонд для проведения внутритрубной диагностики.

В компании разработано внутреннее программное обеспечение для визуализации и ручной разметки магнитограммы.

Одним из лучших решений на рынке является решение Транснефти. У компании существует целый парк внутритрубных инспекционных приборов [https://diascan.transneft.ru/klientam/vnytritrybnaya-diagnostika/park-vnytritry inspekcionnih-priborov/?preview=1]. Магнитные дефектоскопы продемонстрированы на рисунке 2.

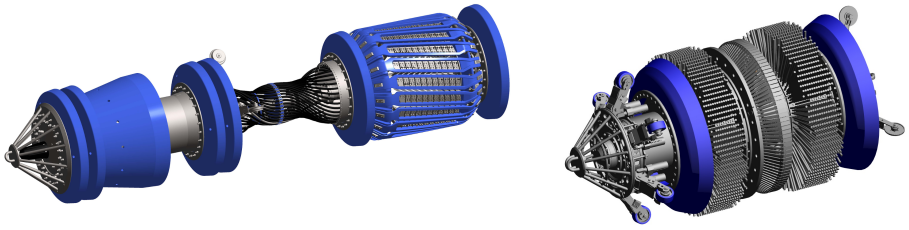


Рис. 2: Магнитные дефектоскопы.

Стоит заметить что ручной анализ магнитограммы занимает длительное время, так размер одного участка трубопровода может составлять десятки километров. Для анализа магнитограмм такого размера у экспертов уходит более одного месяца.

Информации о существующих и эксплуатируемых программных решения для автоматизации процесса интерпретации и детектирования дефектов в открытом доступе не обнаружено.

1.2 Обзор алгоритмов анализа данных о состоянии сети трубопроводов

Для анализа состояния сети трубопроводов в статье [https://www.elibrary.ru/item.asp?id=25323345] предлагается использовать статистические методы и методы машинного обучения. Сам анализ представляет собой детектирование дефектов и конструктивных элементов на магнитограмме, то есть классифкация участка трубы как дефект, конструктивный элемент или полотно трубы. Для решения этой задачи стоит в первую очередь рассмотреть алгоритмы классификации:

- Random Forest Classifier – модель машинного обучения основанная на применение деревьев решений. В основе этого ансамбля лежит подход бэкинга - обучение одинаковы моделей на разных подвыборках набора данных. Он использует усреднение для повышения точности прогнозирования и контроля соответствия. [https://cyberleninka.ru/article/n/slu lesa-obzor/viewer]

- Extra Tree Classifier – данный алгоритм имеет такой же принцип, как и описанный выше Random Forest Classifier, но в качестве базовой модели использует рандомизированное решающее дерево.

- SVM Classifier – метод опорных векторов для классификации. Исходные векторы переводятся в пространство более высокой размерности, а затем в этом пространстве ищется разделяющая гиперплоскость с максимальным зазором, т.е. находящаяся максимально далеко от точек всех представленных классов.

- Logistic Regression - статистическая модель, позволяющая прогнозировать вероятность появление некоторого события по значениям множества признаков. Вероятность вычисляется путем сравнения события с логистической кривой

В статьях [https://www.sciencedirect.com/science/article/abs/pii/S092658051
[https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-dlya-resheniya-obratnoy-zadachi-vnutritrubnoy-magnitnoy-defektoskopii-magistralnyh-truboprovodov

использование нейронных сетей для решения подобных задач. В них фигурируют сверточные нейронные сети и сети прямого распространения.

- Нейронные сети прямого распространения - многослой перцептрон, более подробно алгоритм описан в статье [https://www.sciencedirect.com/scien
- Сверточные нейронные сети - алгоритм на оснвое применение механизма свертки для выделения признаков и дальнейшей классификации на их основе [https://www.sciencedirect.com/science/article/pii/S18770509

Основываясь на природе данных - сигналы магнитометров, было принято решение рассмотреть алгоритмы обработки сигналов, в частности алгоритмы фильтрации:

- Фильтр Калмана - эффективный рекурсивный фильтр, проводит оценку вектора состояния динамической системы на основе ряда неполных и зашумленных данных. [https://cis-linux1.temple.edu/ latecki/ Courses/CIS750-03/Papers/KalmanFilterSIGGRAPH2001.pdf]
- Вейвлет фильтр - фильтр основанный на применении вейвлет преобразования к набору сигналов. [https://cyberleninka.ru/article/n/filtratsiya -signalov-s-pomoschyu-veyvlet-preobrazovaniya]

1.3 Обзор метрик качества алгоритмов по детектированию сварных швов и дефектов

Для оценки результатов были выбраны стандартные метрики: “Recall”, “Precision” и “F1-measure”. Эти метрики основываются на подсчете confusion matrix (матрицы ошибок), приведенной в Таблицах 1 и 2.

Таблица 1: Матрица ошибок для сварных швов

	Сварной шов по разметке	Дефекты и структурные элементы по разметке
Предсказание класса сварного шва	True positive	False positive

Предсказание класса дефектов и структурных элементов	False negative	True negative

Таблица 2: Матрица ошибок для дефектов

	Дефекты по разметке	Сварные швы и структурные элементы по разметке
Предсказание класса дефекта	True positive	False positive
Предсказание класса сварных швов и структурных элементов	False negative	True negative

True positive – области, которые алгоритм определил как швы/дефекты, на самом деле являющиеся ими в размеченной выборке.

False positive – области, которые алгоритм определил как швы/дефекты, НЕ являющиеся ими в размеченной выборке.

False negative – области, которые алгоритм определил как НЕ швы/дефекты, на самом деле являющиеся ими в размеченной выборке.

True negative – области, которые алгоритм определил как НЕ швы/дефекты, НЕ являющиеся ими в размеченной выборке.

На основе представленной матрицы можно подсчитать выбранные метрики по следующим формулам:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

Стоит отметить, что recall считается наиболее важной метрикой, поскольку большую ценность имеет выделение всех сварных швов и дефектов, которые присутствуют в экспертной разметке.

Глава 2. Разработка программного комплекса

2.1 Проектирование архитектуры решения и выбор стека технологий

Для реализации программного комплекса был выбран следующий стек:

- Python 3 - основной язык для реализации модели машинного обучения и веб сервиса. [https://www.python.org/]
- C++ - низкоуровневый язык для реализации парсера. [https://en.cppreferenc
- React JS - фреймворк для реализации пользовательского интерфейса. [https://ru.reactjs.org/]
- Sqlite3 - база данных, используемая в проекте. [https://www.sqlite.org/index.h
- Scikit-Learn - библиотека для реализации моделей машинного обучения. [https://scikit-learn.org/]
- Pandas - библиотека для работы с табличными данными. [https://pandas.pyda
- SciPy - библиотека реализующая методы оптимизации и фильтрации. [https://www.scipy.org/]
- Flask - фреймворк для реализации веб сервиса. [https://flask.palletsprojects.co

Также для проекта была спроектирована архитектура. Ее схема продемонстрирована на рисунке 3. Архитектура системы разделена на 5 компонент: интерфейс пользователя, API, загрузчик данных, модель машинного

обучения и база данных. Данная архитектура поддерживает горизонтальное масштабирование приложения.

- Интерфейс пользователя позволяет визуализировать данные магнитограммы, проводить ручную и автоматическую разметку, оценку и коррекцию автоматической разметки. Логически его можно разбить на 5 частей: компоненту визуализации данных, таблицы для швов и дефектов, интерфейс загрузчика данных и взаимодействия с моделью. Компонента визуализации магнитограммы позволяет демонстрировать участок магнитограммы с помощью двухмерной heatmap. Таблицы для швов и дефектов отображают результаты разметки и позволяют корректировать ее результаты, а также дают возможность демонстрировать конкретные швы или дефекты. Интерфейс загрузчика данных реализует возможность загружать данные магнитограмм в бинарном формате. Интерфейс модели необходим для задания гиперпараметров фильтрации и запуска моделей.

- Компонента API необходима для связи пользовательского интерфейса со всей остальной системой, сам модуль реализован в виде restful web-сервиса с помощью фреймворка Flask. Внутри API находится три интерфейса для работы с базой данных, моделью и загрузчиком данных.

- Загрузчик данных занимается расшифровкой бинарных файлов магнитограмм в соответствии с заданным форматом, преобразование данных из формата WideData в TidyData формат и загрузке данных в базу данных. Загрузчик был реализован на языке C++, так как была необходимость в расшифровке бинарных файлов и непосредственной работой с битами.

- Модуль модели необходим для автоматической разметки магнитограммы по швам и дефектам с дальнейшей загрузкой их в базу данных. Для реализации модели машинного обучения применялся язык Python 3 и набор библиотек: Pandas, Scikit-Learn и SciPy.

- База данных - система для хранения данных программного комплекса в табличном виде. В реализации прототипа использовалась sql база данных Sqlite3.

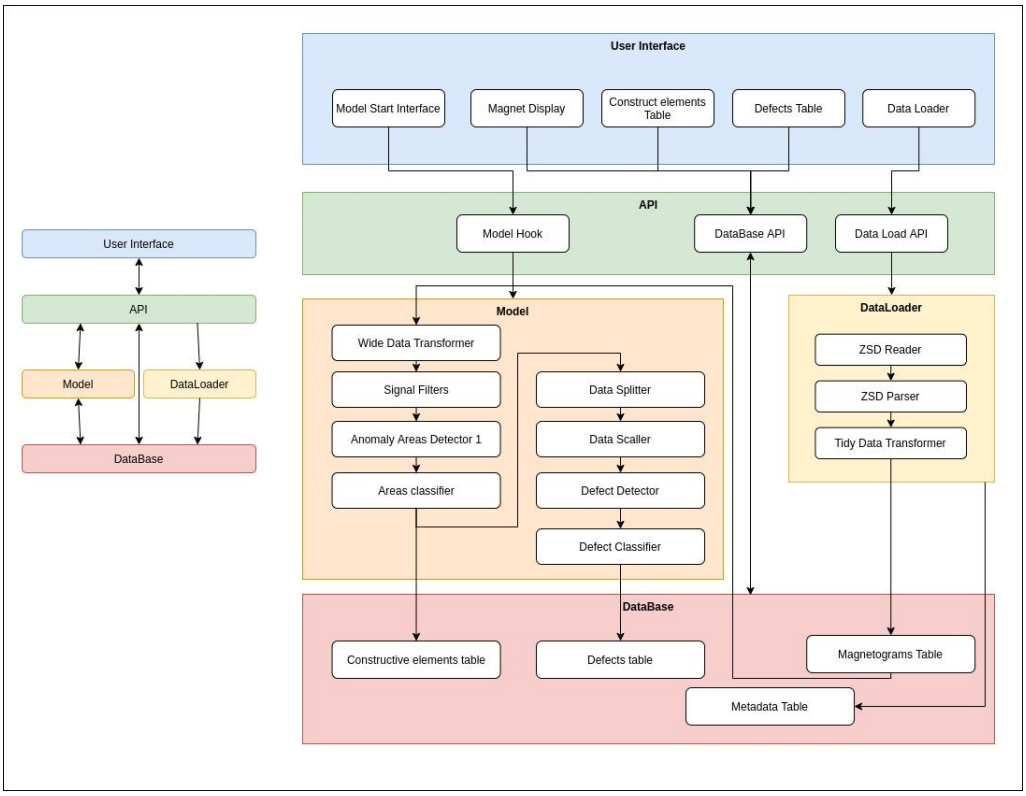


Рис. 3: Архитектура программного комплекса.

2.2 Структура и особенности данных

Исходные данные зондов находились в бинарном формате. Для анализа и построения математической модели эти данные были расшифрованы и преобразованы в табличный формат “.csv” с помощью специально разработанного скрипта.

Для каждой магнитограммы был следующий набор файлов: исходный файл магнитограммы в бинарном виде, преобразованный табличный файл магнитограммы “.csv” и экспертный отчет “.pdf” или “.doc”. В свою очередь в отчёте находилась информация об условиях съемки, а также давалась экспертная разметка магнитограммы с указанием конструктивных элементов и дефектов (включая изображения соответствующих примеров).

Всего в изначальном датасете было 36 магнитограмм, из которых 33 оказались уникальными (3 оставшиеся соответствовали повторному обследованию одних и тех же труб и не содержали дополнительной информации). Список, полученных после расшифровки магнитограмм параметров, приведён в Таблице 3.

Таблица 3: Исходные данные

Параметр	Описание	Единица измерения
Tag	Начало новой записи данных (технический параметр)	-
Size	Размер поля для записи данных (технический параметр)	-
Time	Момент записи	мкс
Dist	Значение дистанции	10 мкм
Index	Порядковый номер измерения	-
Flags	Флаги сканирования (технический параметр))	-
Status X	Статус для блока из четырех датчиков, описывающий возможные ошибки в них: 0 – отсутствие ошибок; X – номер блока	-
X.Y	Значение датчика Y из блока X; например: “1.3”	-

Термин “Технический параметр” означает, что данная переменная используется либо для расшифровки, либо для конверсии из исходного бинарного файла. Из приведённого списка для анализа магнитограмм наиболее важными факторами являются Dist и X.Y, так как именно они указывают на каком расстоянии находится дефект/шов/структурный элемент, и каким измеренным значениям он соответствует.

Первичный анализ данных, показал необходимость перепроверки и переразметки данных. Параметры, использованные для разметки сварных швов, дефектов и структурных элементов, приведены в Таблицах 4 и 5.

Таблица 4: Параметры конструктивных элементов

Параметр	Описание	Единица измерения
id	Уникальный идентификатор конструктивного элемента – строка	-
contructive type	Тип конструктивного элемента на основе отчетов	-
start dist	Начало конструктивного элемента	10 мкм
end dist	Конец конструктивного элемента	10 мкм
wall width	Толщина стенки трубы на основе отчета	мм

Таблица 5: Параметры дефектов

Параметр	Описание	Единица измерения
id	Уникальный идентификатор дефекта – строка	-
defect type	Тип дефекта на основе отчетов	-
defect place	Местоположение дефекта: "Стенка трубы"или "Сварной шов"	-
defect depth	Глубина дефекта, на основе отчета	%
start dist	Начало дефекта	10 мкм
end dist	Конец дефекта	10 мкм

После переразметки, в соответствующие папки магнитограмм были добавлены файлы конструктивных элементов "$construct\_element.csv$" и файл разметки дефектов "$defect.csv$".

При сопоставлении исходных данных магнитограмм с экспертной раз-

меткой, приведённой в отчётах, было обнаружено расхождение длины пути зонда на расстояние вплоть до 150 см, что в свою очередь привело к отличию координат дефектов и конструктивных элементов между датасетом и сопутствующим отчётом. Предположительно, причиной этого расхождения является ошибка в ПО, используемом для генерации графиков к отчёту. Для экспертного отчёта подобное расхождение является нежелательным, но не критичным, так как ремонт повреждённого участка происходит не точечно, а в рамках целой секции трубы. В то же время, такое несоответствие между отчётом и магнитограммой весьма негативно сказывается на точности математической модели, которая не может обучиться на противоречащих данных. По этой причине силами сотрудников Дирекции Инновационного Развития была выполнена ручная проверка и переразметка датасета каждой магнитограммы (“.csv”), а обновлённые файлы были добавлены в соответствующие папки.

Важно заметить, что число записей (строк с показаниями датчиков) в магнитограмме зависят от пройденного зондом расстояния, которое может достигать 40 км. В представленном ниже датасете длина составляла 1.5-10 км. Участки свыше 10 км были исключены из рассмотрения в связи с трудоёмкостью уточнения изначальной разметки. В результате итоговый датасет составил 22 магнитограммы для анализа швов и 18 для анализа дефектов (некоторые трубы не имели дефектов согласно отчётам). Итоговый список магнитограмм для построения математической модели приведён в Таблице 6.

Таблица 6: Итоговый список магнитограмм

Магнитограмма	Количество записей
Магнитограмма №1	2125866
Магнитограмма №2	1470101
Магнитограмма №3	3815418
Магнитограмма №4	2482089
Магнитограмма №5	3761227
Магнитограмма №6	1072643

Магнитограмма №7	449613
Магнитограмма №8	711720
Магнитограмма №9	412681
Магнитограмма №10	1165283
Магнитограмма №11	3047114
Магнитограмма №12	575441
Магнитограмма №13	1498927
Магнитограмма №14	683275
Магнитограмма №15	1108527
Магнитограмма №16	1630936
Магнитограмма №17	338127
Магнитограмма №18	233018
Магнитограмма №19	291116
Магнитограмма №20	222167
Магнитограмма №21	860107
Магнитограмма №22	1772941

Для выявления особенностей данных были визуализированы данные для всех магнитограмм. Поскольку данные писались одновременно 64 независимыми датчиками, то для отображения графиков использовалось усреднённое значение по всем датчикам, записанное на данной координате (для сварных швов и структурных объектов). Для дефектов строились индивидуальные графики для каждого датчика. В итоге были зафиксированы следующие результаты (изображения ниже приводятся только для структурных элементов и сварных швов).

Определены паттерны конструктивного элемента "Сварной шов"и их средние показатели (ширина 5 сантиметров и высота 250 у.е.) (Рис. 4).

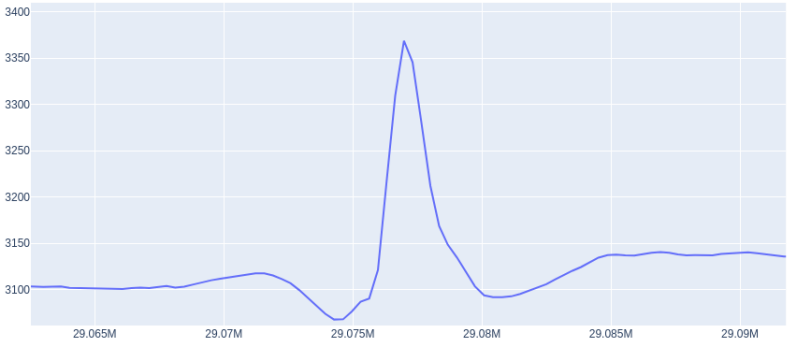
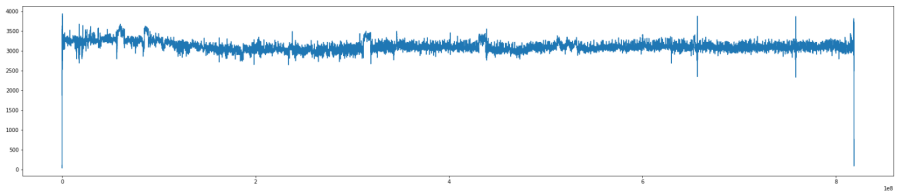
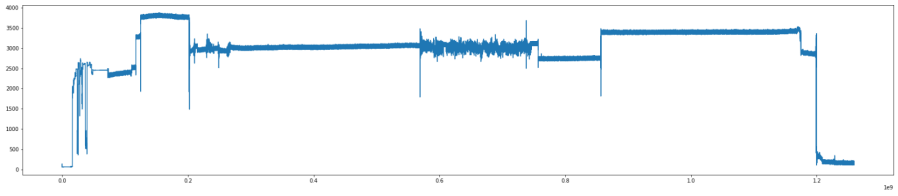


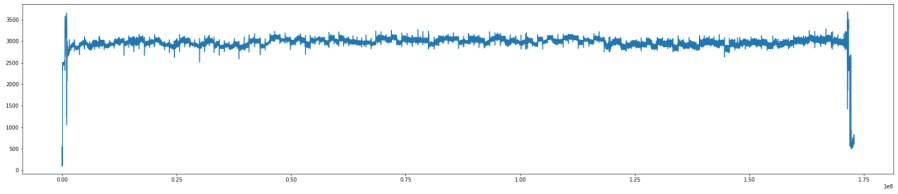
Рис. 4: Выделенный паттерн шва.



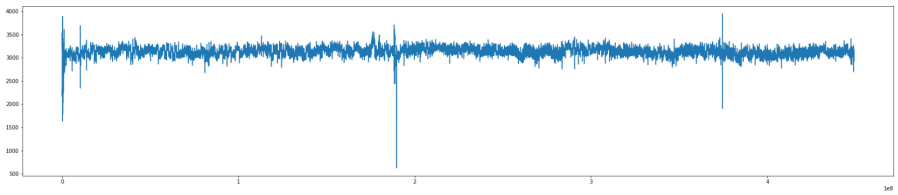
Магнитограмма №1.



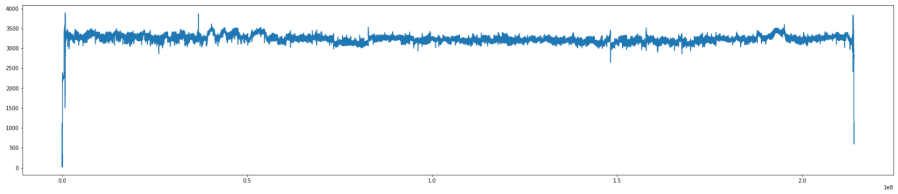
Магнитограмма №3.



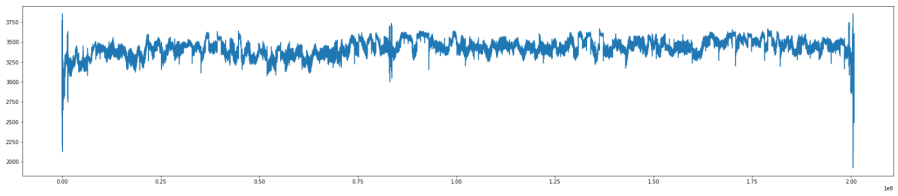
Магнитограмма №9.



Магнитограмма №10.

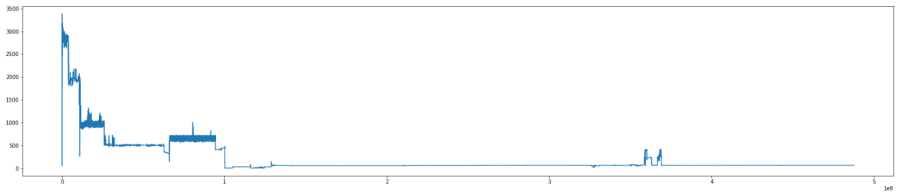


Магнитограмма №12.

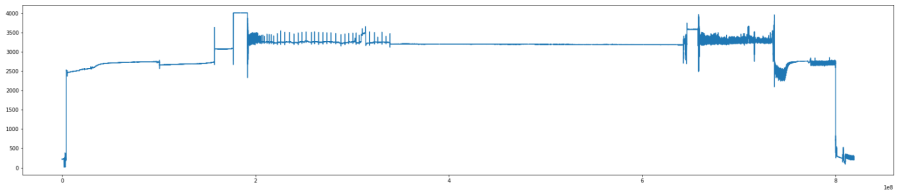


Магнитограмма №14.

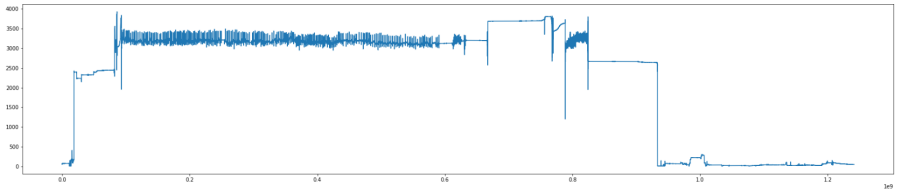
Рис. 5: Магнитограммы с сильными шумами.



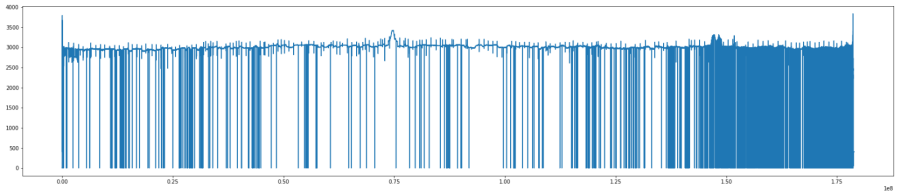
Магнитограмма №2.



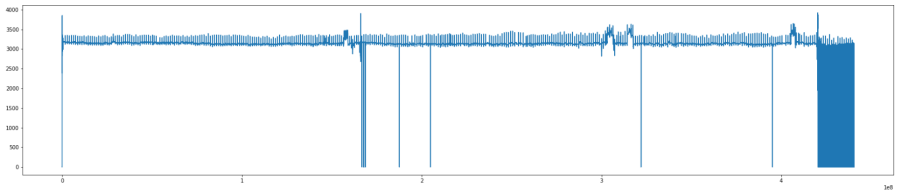
Магнитограмма №4.



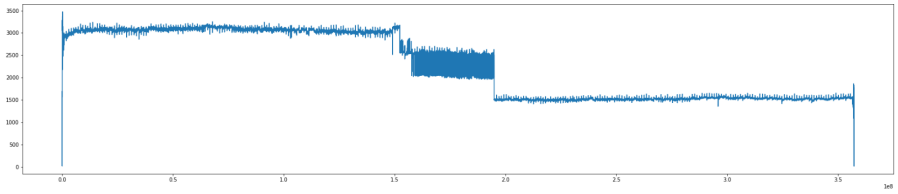
Магнитограмма №5.



Магнитограмма №7.



Магнитограмма №15.



Магнитограмма №16.

Рис. 6: Магнитограммы с битыми или недостоверными данными.

Зашумлённые и "битые"магнитограммы также использовались для

построения математической модели, но с заведомым ожиданием ухудшения качества детектирования на них.

2.3 Процесс предобработки данных

При анализе статусов блоков датчиков было замечено, что почти для половины магнитограмм датчики выдавали сообщения об ошибках. Поскольку эксперты всё же смогли разметить подобные "битые"данные, было принято решение не исключать проблемные магнитограммы и попробовать восстановить их двумя разными способами:

- Анализ показаний статусов блоков датчиков. Данный подход оказался неудачным, ввиду того, что для некоторых магнитограмм количество ненулевых статусов соответствовало количеству выполненных измерений. Соответственно, во всей магнитограмме были отражены недостоверные данные (например, см магнитограмма №7 в Таблице 4).

- Проведение порогового анализа. На основе магнитограмм, визуально не имеющих битых данных, была рассчитана нижняя граница доверительного интервала показания датчиков. Определение "битых"/ недостоверных данных проводилось с помощью рассчитанного порога: в среднем значения в магнитограмме по оси ординат колебались от 0 до 4000 у.е. (условных единиц); значения ниже 1000 у.е. экспертно были признаны недостоверными и заменены на арифметическое среднее по достоверным данным этой магнитограммы. Наглядные примеры различных "битых"и зашумлённых магнитограмм продемонстрированы на рисунках 5 и 6.

Таким образом, метод порогового анализа позволил использовать для работы весь датасет, включая магнитограммы, состоящие только из "битых"данных.

2.4 Построение математической модели

Для решения задач был выбран следующий подход:

- Разработать аналитическое решение, способное определять области, схожие с паттерном сварного шва;
- Уточнить решение с помощью модели машинного обучения, обучив ее на подготовленной выборке;
- Провести разделение магнитограммы на секции по детектированным швам;
- Для каждой выделенной секции с помощью аналитического решения найти области, являющиеся локальными минимумами значений сигналов;
- Уточнить решение с помощью фильтрации и модели машинного обучения, обучив ее на подготовленной выборке.

2.5 Выбор признаков

Для обучения модели по детектированию швов на сформированном датасете в качестве признака было выбрано окно размером в 17 показаний на усредненном по всем датчикам сигнале: 8 ближайших показаний левее заданной точки, показание на данной точке, 8 показаний правее. Размер окна был выбран эмпирическим образом после проведения экспериментов для максимизации значений метрик recall и precision.

Для обучения модели по детектированию дефектов на сформированном датасете в качестве признака было выбрано окно размером в 20 показаний на всех датчиках: 9 ближайших показаний левее заданной точки, показание для данной точки, 10 показаний правее. В результате выбора такого окна была получена матрица признаков размером $n \times 20$, где n – количество датчиков. Для увеличения количества сэмплов с дефектами, производился циклический сдвиг полученной матрицы построчно: изначально строка с дефектом размещалась "внизу"матрицы, а затем на каждой итерации смещалась “вверх”, пока не оказывалась на первом месте. Подобная манипуляция была нужна чтобы проиллюстрировать принципиальную возможность нахождения дефекта в любом угловом секторе трубы. В про-

тивном случае, в процессе обучения модель могла бы решить, что дефекты всегда располагаются исключительно в определенном секторе, а похожие показания в других секторах – статистические выбросы. После подготовки всех семплов, полученные матрицы подверглась преобразованию в массив путем последовательной записи столбцов. Аналогично сварным швам, размер окна был выбран эмпирически, исходя из экспериментов для максимизации значения метрик recall и precision у модели машинного обучения.

2.6 Решение по детектированию швов

Аналитическое решение представляет собой алгоритм по поиску характерных пиков, явно выбивающихся из основного сигнала, и согласно картам разметки, соответствующих сварным швам. Основной проблемой такого подхода является наличие битых данных и сильная зашумленность некоторых магнитограмм. Для решения проблемы зашумленных данных были предприняты следующие шаги:

- Использование среднего значения показаний датчиков для анализа;
- Подсчет скользящего среднего с окном размером в 100 показаний и вычет его из сигнала;
- Подсчет скользящего среднего с окном размером в 2 показания и вычет его из сигнала.

Также в ходе решения проблемы было рассмотрено применение фильтра Калмана и фильтра на основе вейвлет-преобразования. В финальную версию решения они не вошли, поскольку их применение к магнитограмме занимало значительно больше времени, чем весь остальной алгоритм аналитического решения. В частности, для одной магнитограммы длиной 1 км время работы фильтров составило:

- Вейвлет фильтр – 123 минуты;
- Фильтр Калмана – 36 минут.

После применения описанных преобразований к усредненному сигналу, производился поиск пиков с помощью функции find peaks, реализованной в библиотеке SciPy. Для отсечения шумов использовались пороговые значения по относительной и абсолютной высоте пика. Для работы было определено пороговое значение, равное 1000.

Важно заметить, что данные фильтры использовались исключительно для аналитического решения, чтобы эффективнее выделять пики и находить их максимальные значения. При этом профиль самих пиков искажался, что делало невозможным их дальнейшее использование для обучения математической модели (разрабатываемая модель машинного обучения по-прежнему обучалась на зашумлённых данных, впрочем, теперь с уже точно заданными координатами пиков).

Далее после выделения областей интереса происходила фильтрация с помощью модели машинного обучения. Обучение модели проводилось на выделенных ранее признаках, соответсвующих двум классам: швам и полотну трубы. Баланс обучающей выборки был 1:1.

2.7 Решение по детектированию дефектов

Для аналитического решения был выбран следующий подход:

- Магнитограмма разбивалась на секции по швам;
- Для каждой секции производился поиск локальных минимумов значений сигналов датчиков;
- Производилась фильтрация выбранных минимумов по глубине.

Для поиска локальных минимумов производилась предобработка данных. Сначала данные интерполировались по дистанции для достижения сетки с равномерным шагом (т.е. пространственно измерения были равноудалены друг от друга). Также применялся гауссовский фильтр с целью уменьшения зашумленности. Минимумы искались с помощью функции find peaks, реализованной в SciPy. Данная функция применялась к

среднему сигналу с каждого блока датчиков, умноженному на -1 для отражения сигнала относительно оси абсцисс (Ox) на Рисунках 6a-7e. Также производилась фильтрация пиков по их метапараметрам – ширине и относительной высоте. Для каждого пика сохранялись данные об абсолютной и относительной высоте, ширине, координатах левой и правой границ пика. Также производился подсчет процента глубины дефекта от общего сигнала. Наконец, для фильтрации дефектов применялся пороговый фильтр по проценту глубины потери металла.

После выбора зон возможных дефектов, производилось уточнение с помощью модели машинного обучения. Её обучение проводилось подобным образом, как описано в пункте выше.

Ссылка на статью: [1], [2]

Выводы

Жизнь — тлен.

Заключение

В рамках проведенных работ были выполнены следующие этапы:

- Разработано программное обеспечение для преобразования файлов формата “.zsd” к табличному формату “.csv”;
- Проведена корректировка экспертной разметки с целью уточнения дистанций дефектов и конструктивных элементов для 22 магнитограмм;
- Реализован прототип цифрового продукта, производящий автоматическую разметку магнитограммы с дальнейшей возможностью её корректировки в ручном режиме;
- Разработаны две модели для детектирования сварных швов и дефектов; результаты качественной оценки решений зафиксированы в разделах и ;

При дальнейшем развитие проекта и доработке решения следует обогатить выборку за счет добавления новых магнитограмм с экспертной разметкой. Так же для улучшения работы самого решения существуют следующие возможности:

- Применить фильтрацию к данным;
- Провести более подробный анализ паттерна дефекта;
- Обсудить с экспертами выявленные аномальные ситуации (в том числе их возможную физическую природу);
- Исследовать применимость дополнительных дата-саенс подходов (бустинговые деревья в текущей постановке задачи, нейросети для обработки изображений и выявления на них паттернов швов/дефектов);
- Провести детальный подбор гиперпараметров для моделей;
- Добавить такие метапризнаки, как ширина, высота, глубина и т.п. для детектируемого объекта.

Список литературы

[1] Griffin D.W., Lim J.S. «Multiband excitation vocoder». IEEE ASSP-36 (8), 1988, pp. 1223-1235.

[2] Griffin D.W., Lim J.S. «Multiband excitation vocoder». IEEE ASSP-36 (8), 1988, pp. 1223-1235.