

Санкт-Петербургский государственный университет

Программная инженерия

Группа 21.М07-мм

Разработка библиотеки методов обработки данных акустической эмиссии на языке PYTHON

МАКАРИХИН Павел Андреевич

Отчёт по преддипломной практике

Научный руководитель:
доцент кафедры информатики, к. т. н., А. В. Григорьева

Санкт-Петербург
2023

Оглавление

Введение	3
1. Постановка задачи	7
2. Обзор	8
2.1. Существующие методы анализа данных АЭ	8
2.1.1. Методы локации	8
2.1.2. Методы кластеризации	9
2.1.3. Методы классификации	9
2.2. Программное обеспечение	10
3. Архитектура	12
4. Реализация	15
4.1. Первичная обработка	15
4.2. Локация	15
4.3. Кластеризация	16
4.4. Классификация	18
4.5. Вспомогательные методы	18
5. Эксперимент	19
Заключение	29
Список литературы	30

Введение

В повседневной жизни человеком используется множество объектов и конструкций, которые подвергаются постоянным нагрузкам. Такие объекты, как мостовые колонны, резервуары для перевозки жидкостей, различные конструкции в промышленности и многие другие, в процессе функционирования испытывают нагружение [4, 19, 10]. Поэтому при эксплуатации очень важно знать их состояние во избежание аварийных ситуаций. Для контроля надежности объектов существуют методы разрушающего и неразрушающего контроля (НК). Преимуществом НК является то, что для оценки состояния объекта нет необходимости его разрушать или демонтировать, что позволяет сохранить объект после проведения контроля.

Существует множество методов НК: акустический, радиационный, оптический, электрический и многие другие. Каждый из них обладает своими преимуществами и недостатками. Одним из наиболее популярных подходов является метод акустической эмиссии (АЭ), благодаря его точности и экологичности. Суть метода АЭ заключается в обнаружении акустических волн, возникающих в объекте при возникновении и развитии дефектов в процессе эксплуатации объекта. Для выполнения этого метода на исследуемый участок объекта устанавливаются датчики АЭ - преобразователи акустической эмиссии, после чего при непрерывном нагружении объекта в нем начинают появляться различные дефекты: трещины, сколы, выкрашивание материала и др. Нагружение объекта происходит по-разному в зависимости от его конструкции. Так, например, для цистерн и сосудов высокого давления часто применяют гидравлический тест, в ходе которого объект наполняется водой под давлением. После заполнения полостей возникающая нагрузка на конструкцию стимулирует рост дефектов, который необходим для их регистрации датчиками АЭ. Для таких объектов же как колонны, балки и пластины применяют растяжение-сжатие. Для этого исследуемый образец фиксируется в нескольких точках, а затем с помощью специальных прессов на определенную область объекта подается нагрузка, в

ходе которой он деформируется растягиваясь или сжимаясь, из-за чего в нем также возникают дефекты и повреждения. При возникновении и в процессе своего роста дефекты излучают распространяющиеся по объекту акустические волны, которые регистрируются датчиками (рис.1). Фиксируются различные параметры акустической волны: амплитуда, энергия, время прихода на датчик, время роста волны, пиковая амплитуда и многие другие. Это похоже на сложные системы с непериодической динамикой [3].

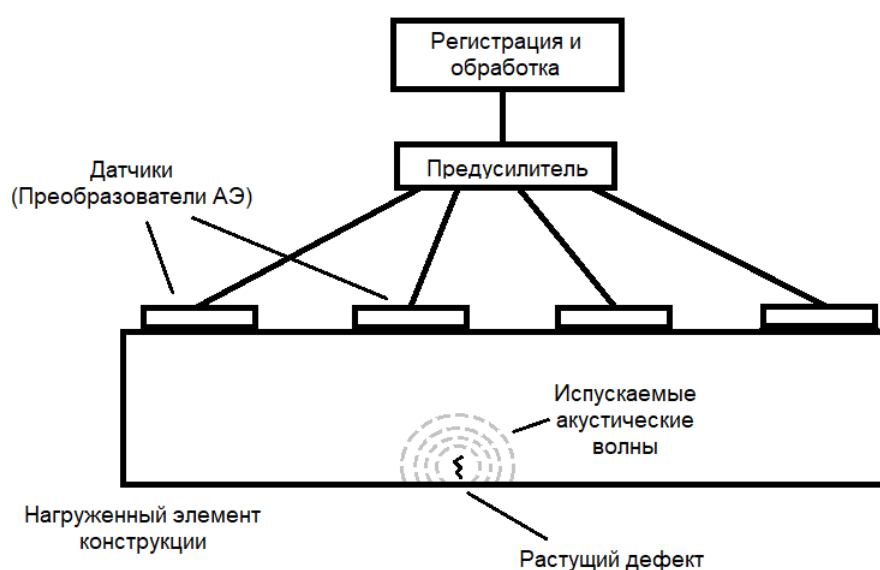


Рис. 1: Пример системы для АЭ контроля

Существует три основных применения метода АЭ: определение местоположения источника волн (локация), мониторинг состояния конструкции и оценка характеристик материалов и конструкций. Размеры исследуемой конструкции могут быть весьма значительными и специалистам необходимо провести анализ всей площади, так как любой незначительный дефект со временем может достичь критического уровня, в результате чего возможен выход объекта из строя или даже возникновение ЧС. Оценка состояния объекта может занимать очень много времени из-за размера обследуемого объекта, поэтому крайне важно

уменьшить участки, которые необходимо проверить специалисту. Достичь этого можно с помощью комбинации методов анализа данных АЭ и эмпирических знаний, полученных за время существования НК и АЭ в частности.

Локация источников АЭ позволяет специалистам уменьшить область, необходимую для анализа, а также построить соответствие между сигналами и источником, их испускающим. Сложность локации заключается в том, что датчики, зафиксировавшие сигнал, лишь ограничивают область, в которой потенциально располагается источник. Для более точного вычисления координат необходимо проводить дополнительный анализ данных: учитывать различные параметры сигналов, особенности структуры области, в которой они возникают (различные полости, люк-лазы, штуцеры и тому подобное) и даже особенности материала, из которого состоит объект. Однако, даже при использовании такого большого количества показателей на практике невозможно избавиться от погрешности при расчете координат источников.

Основной целью мониторинга состояния конструкции является своевременное нахождение сигналов, испускаемых дефектами, и приостановка эксплуатации до возникновения критической ситуации. Для этого применяют методы кластеризации и классификации сигналов по их параметрам. В ходе своего развития дефекты испускают множество волн, которые фиксируются датчиками. Однако, определить, что все они принадлежат одному и тому же источнику достаточно проблематично, так как параметры волн могут не только варьироваться с ростом дефекта, но и различаться по типу (например, продольному или поперечному). Разбиение на кластеры позволяет сгруппировать схожие по параметрам сигналы, а полученные области сгущения элементов кластеров отражают активность дефектов. По полученным кластерам можно сформировать основные типы сигналов, возникающие в объекте. Разделение дефектов и шумов по типам сигналов дает возможность соотносить новые сигналы, к полученным классам, что ускоряет анализ. Определение классов шумов также позволяет выделить полезные сигналы из всего объема данных. Таким образом методы кластеризации и

классификации имеют множество полезных применений в АЭ.

Одной из основных проблем методов анализа данных АЭ является вариативность анализируемых конструкций и материалов, из которых они состоят. Исследования в данной области продолжаются уже более 50 лет, но до сих пор нет общих решений, так как для разных материалов и разных конструкций распространение упругих волн происходит по-разному [28]. Существует не так много инструментов для работы с данными АЭ. Часто методы, используемые в них, скрыты коммерческой тайной или же доступны в усеченном варианте, поэтому разработка библиотеки методов обработки данных АЭ с явным указанием, какие алгоритмы в ней используются, является актуальной задачей и по сей день. Для реализации библиотеки был выбран язык программирования PYTHON, так как он хорошо подходит для анализа данных и используется во многих вспомогательных библиотеках машинного обучения и обработки данных, а также он достаточно популярен среди ученых в области АЭ. За время обучения было разработано и реализовано множество методов анализа данных АЭ на разных языках программирования (R, C#), поэтому их необходимо переписать на PYTHON для объединения в единую библиотеку.

1. Постановка задачи

Целью данной выпускной квалификационной работы является разработка библиотеки методов обработки данных АЭ для автоматизации процесса их анализа. Для достижения цели были поставлены следующие задачи:

- Переписать разработанные ранее алгоритмы на язык программирования PYTHON;
- Разработать новые методы для анализа данных;
- Апробировать новые методы на реальных данных;
- Сформировать библиотеку методов.

2. Обзор

2.1. Существующие методы анализа данных АЭ

За последние 50 лет было разработано множество методов анализа данных АЭ. Применялись различные подходы к локации, кластеризации и классификации. Подходы отличаются не только используемыми алгоритмами, но и параметрами сигналов, к которым они применяются. Данные АЭ, зафиксированные датчиками, могут быть объемными как по количеству учитываемых параметров сигнал, так и по количеству самих сигналов. Причем для разных конструкций и используемых материалов значения параметров сигналов могут сильно различаться, что и объясняет такое разнообразие подходов.

2.1.1. Методы локации

Для локации источников используются различные подходы, такие как модальный анализ и анализ частот [15], специальное расположение датчиков для упрощения расчетов [23], вейвлет анализ [17], аналитические решения [8, 27] и множество других. Такое разнообразие способов обусловлено сложностью проблемы локации и вариативностью структур конструкций и материалов, из которых они построены.

Частым недостатком традиционных методов является зависимость от инвариантности скорости распространения сигнала, изотропности материала, однородности конструкции (отсутствие люк-лазов, полостей, штуцеров и тому подобного) и типов дефектов. Трещины, расслоения, смещения, крошение и многие другие по-разному испускают волны, а следовательно и обрабатывать их необходимо по-разному.

В современных исследованиях пытаются избавиться от этих недостатков и разработать универсальные подходы, однако, существующие методы все еще не идеальны.

2.1.2. Методы кластеризации

Объединение сигналов в кластеры - еще одна трудная задача при анализе данных АЭ. Сигналы обладают различными физическими параметрами, по которым можно группировать сигналы. Основными параметрами для кластеризации часто используют амплитуду, порядок датчиков, зафиксировавших сигнал, время роста сигнала и время прихода сигнала на датчик.

На сегодняшний день существует множество подходов к кластеризации данных, различающихся типом данных и алгоритмом кластеризации. Разные алгоритмы имеют разные преимущества и недостатки, а также разные требования к входным параметрам. В [16] к данным форм волн применялся алгоритм К-средних с инициализацией центров кластеров случайным равномерным распределением. В [6] использовались карта Кохонена и анализ главных компонент в совокупности с классическими статистическими методами. В [20] использовалась модификация модели гауссовой смеси применительно к временным отрезкам зафиксированных сигналов.

2.1.3. Методы классификации

Современные методы классификации все чаще применяют в своих алгоритмах машинное обучение и нейронные сети в частности. Своевременное определение серьезных дефектов является важной задачей АЭ. Разработки в области классификации сигналов АЭ преследуют цель разбить все данные на классы, соответствующие типам дефектов или их размеру/критичности, а также отделить классы шумов. Выделение шумовых классов позволяет в дальнейшем сразу определять полезность новых сигналов. В статье [7] к заранее обработанным формам волн сигналов, зафиксированным в ходе коррозии металла, помещенного в химический состав, применялась классификация методом случайного леса. В работе [1] применялись сверточные нейронные сети для классификации дефектов, возникающих в ходе нагрузки образцов из специального сплава алюминия.

2.2. Программное обеспечение

Описанные подходы лишь небольшая часть всех существующих способов анализа данных АЭ, однако, одной из проблем разработанных методов является то, что они часто не оформлены как программный продукт с возможностью использования пользователями. Методы описываются в статьях, но их код либо не выложен в открытый доступ, либо трудно используем.

Проблемой коммерческого программного обеспечения (ПО) же часто является то, что алгоритмы, используемые в нем скрыты от пользователя, а специалистам важно понимать, какие решения принимает программа при выдаче результата. Также такое ПО часто привязано к оборудованию компании, разрабатывающей его, а следовательно использование своих данных, полученных, допустим, из другой программы пользователем, может быть проблематично. Широко используемым программным обеспечением является Vallen AE Suite Software от компании Vallen Systeme, Physical Acoustics' AEWin и NOESIS, A-Line. Все они обладают схожим функционалом и реализуют большинство методов, необходимых для анализа данных АЭ: различные методы локации (двухмерной и трехмерной), кластеризации, обработку форм волн. Один из недостатков такого ПО - невозможность модифицировать используемые методы, что уменьшает применимость для определенных случаев.

Используемые технологии и разработки в области АЭ развиваются с каждым днем, а значит появляются все более и более эффективные методы, которые можно использовать. Поэтому возможность вносить изменения и добавлять свои методы также имеет значение. Библиотека методов с открытым кодом позволит пользователем достичь этого.

Разработчики Vallen Systeme также создали библиотеку pyVallenAE на языке PYTHON с открытым кодом на GITHUB. Однако, в ней реализован лишь минимальный функционал: чтение файлов данных АЭ, выделение параметров сигналов и работа со временем фиксации сигнала. Для полного анализа такого набора методов недостаточно.

Кроме коммерческого ПО была также рассмотрена работа выпуск-

ника магистратуры СПбГУ Жибаря Марка Артуровича, защищенная на кафедре программной инженерии в 2022 году. Работа Марка Артуровича также реализует функционал для обработки и кластеризации данных АЭ. Результатом работы стал программный комплекс, состоящий из нескольких приложений на разных языках программирования для обработки, локации, кластеризации и визуализации сигналов АЭ. Для локации источников сигналов применялся аналитический метод, решающий систему уравнений с использованием скорости и разницы времени прихода сигнала на датчик. Для кластеризации сигналов основным параметром использовались координаты источников и скорость сигнала, вычисленные методом локации. Значения амплитуд использовались лишь для фильтрации сигналов, не попадающих в определенный диапазон.

В данной же работе для локации используется итерационный метод, а для кластеризации сигналов амплитуда используется не просто как численный фильтр, а как дополнительный параметр, учитывающийся с помощью специально разработанной метрики. Также отличием данной работы является формирование единой библиотеки методов без необходимости использования отдельных приложений под разные цели и возможность пользователями использовать разработанные методы в своем коде.

3. Архитектура

Разработанная библиотека состоит из компонент, каждая из которых содержит методы для анализа данных АЭ (Рис. 2). Целью такого разделения является выделение основных областей методов анализа для упрощения работы с ними и потенциального расширения функционала.



Рис. 2: Компоненты библиотеки

Компонента работы с файлами позволяет считывать данные АЭ для дальнейшей их обработки. Файлы АЭ представляют из себя заголовки с физическими параметрами, зафиксированными датчиками, допол-

нительную информацию (дату записи, название объекта и прочее) и массив строк со значениями. Зафиксированные параметры необходимо обрабатывать по-разному, так как они могут отличаться не только размерностями, но и смыслом значений, а также могут быть категориальными или числовыми. Следовательно важно подготовить заранее данные для дальнейшей работы с ними.

Данные акустической эмиссии, полученные в ходе реальных экспериментов, часто содержат большое количество различных шумов (человеческий фактор, электромагнитные колебания, космические лучи [9] и др.), которые также фиксируются датчиками, поэтому крайне важно фильтровать полезную информацию от помех перед дальнейшей работой [5, 13]. Часть таких шумов возможно убрать сразу же без дополнительного анализа. Применение методов первичной обработки уменьшает объем анализируемых данных, удаляя заранее известные “шумные” данные, что ускоряет дальнейшую работу.

После подготовки данных и удаления очевидных шумов в зависимости от целей исследователя можно лоцировать источники АЭ, кластеризовать сигналы для выделения групп шумов и дефектов, классифицировать сигналы по типам.

Компонента локации содержит методы для определения расположения источников сигналов АЭ. Координаты источников можно использовать как дополнительную информацию при кластеризации и классификации. Расположенные в одной области источники вероятнее всего относятся к одному и тому же дефекту. Однако только по расположению утверждать о наличии там дефекта нельзя, ведь есть вероятность, что это могут оказаться шумы. Для более точного анализа в библиотеке реализованы компоненты кластеризации и классификации.

Кластеризация позволяет объединить схожие по физическим параметрам сигналы в группы, которые отражают текущее состояние объекта. Полученные группы можно также использовать для дальнейшего анализа. Например, определив группы шумов, их можно использовать для дальнейшей классификации новых сигналов. Используя вычисленные с помощью методов локации координаты источников вместе с физически-

ми характеристиками кластеризация позволяет сгруппировать сигналы, в кластеры, которые будут отражать область расположения источника сигналов.

После получения кластеров все новые фиксирующиеся сигналы можно классифицировать, соотнося к той или иной группе. Определение типов дефектов и соотнесение сигналов к тем или иным классам позволит специалистам лучше понимать состояние объекта и заранее принимать меры при обнаружении сигналов критических классов. Для достижения этого были реализованы методы из компоненты классификации.

Также для проверки базовой корректности работы методов библиотеки использовалось модульное тестирование. Тесты содержатся в компоненте UnitTests и заключаются в проверке совпадения получаемого результата с ожидаемым для тривиальных данных. Использование сложных реальных данных для модульного тестирования достаточно проблематично, так как время выполнения тестов будет больше и полученный результат необходимо дополнительно анализировать для проверки корректности. Апробация методов на реальных данных будет подробнее описана в разделе "Эксперименты".

Применение методов из указанных компонент возможно как по отдельности, так и совместно, в зависимости от целей пользователя. Для исходных данных возможна последовательность использования компонент, указанная на рис. 3

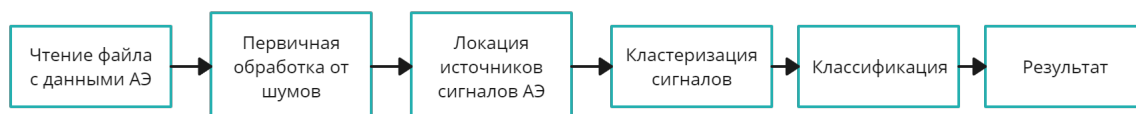


Рис. 3: Сценарий использования компонент библиотеки

Если же, например, координаты источников уже известны, то использование метода локации не нужно, а следовательно компоненту локации из данной цепочки можно удалить. Однако, будет необходимо привести данные к виду, используемому в конкретной компоненте.

4. Реализация

4.1. Первичная обработка

Существуют различные способы отделения сигналов от шума: [2, 25, 18, 21]. В данной работе записанные сигналы были отфильтрованы от явных шумов, амплитуда волны которых не попадает в диапазон волновых значений от источников АЭ от 40 [дБ] до 100 [дБ], аналогично [4]. Далее удалялись сигналы, зарегистрированные не более чем 4 датчиками. Учитывая погрешность аппаратуры, установить скольконибудь достоверное местонахождение источников АЭ с помощью менее 4-х датчиков невозможно. Затем из оставшихся данных были удалены гипотетические электромагнитные импульсы, которые также могут быть зафиксированы датчиками от внешних объектов и устройств входе эксперимента. Эти волны не несут информации о состоянии объекта и поэтому также в данном случае относятся к шуму. Такие данные отличаются тем, что их фиксация датчиками происходит практически одновременно, так как электромагнитные волны распространяются со скоростью света. В методе предполагалось, что сигналы являются электромагнитными, у которых разница в приходе на датчики составляет менее 0,2 микросекунды [мкс].

4.2. Локация

Для локации используются метод приближенного вычисления координат источника с помощью сетки, описанный в бакалаврской выпускной квалификационной работе [29]. Суть данного метода заключается в итерационном вычислении потенциальных координат источника по узлам сетки при условии, что скорость сигнала меняется незначительно. Преимуществом этого метода является то, что для каждого сигнала скорость рассчитывается индивидуально, а не задается одним глобальным значением. В данной работе метод был переписан с языка программирования R на PYTHON.

4.3. Кластеризация

Существует много алгоритмов кластеризации: k-средних, affinity propagation[12], агломеративная кластеризация, различные нейронные сети и многое другое. Перед началом разработки метода был проведен анализ некоторых из них. Так, например, метод k-средних требует заранее знать количество кластеров, зависим от выбора центроид и имеет проблемы с данными разной плотности и размера, поэтому без дополнительных модификаций его проблематично использовать. Affinity Propagation достаточно медлителен для больших объемов данных и также требуется дополнительный анализ в случае слишком сильного различия кластеров по размеру. Проблемой агломеративной иерархической кластеризации может быть несбалансированность - возникновение одного крупного кластера и множества маленьких. Для кластеризации данных был выбран алгоритм DBSCAN [24], поскольку он не требует заранее знать количество кластеров, дает одинаковый результат для одного и того же набора данных и устойчив к выбросам. Кроме того, неявным преимуществом данного алгоритма является достаточная простота его работы, так как специалистам важно понимать, почему применяемый метод выдает именно такой результат и как его интерпретировать.

Суть разработанного метода кластеризации заключается в автоматическом вычислении диапазонов характеристик, характерных для данного конкретного объекта, для одной из опасных зон, в которой заведомо имеется дефект, и на основании этого нахождения остальных зон вероятных дефектов. Для работы метода необходимы координаты прямоугольника, в котором находится искомая зона, и количество высокоамплитудных импульсов в этом прямоугольнике. Число определяется опытным путем и варьируется от 3 до 10 в зависимости от того, насколько поврежден объект, исходя из [26]). Удачный выбор такой зоны может повлиять на качество кластеризации.

Для правильного формирования кластеров важно учитывать не только расположение точек, но и другие физические параметры сигналов,

такие как скорость, амплитуда и другие. Выбор параметров основывался на следующих статьях: [26, 22, 11]. Дополнительной сложностью является то, что разные параметры обладают разной размерностью и значимостью, а следовательно применение евклидовой метрики для вычисления расстояния между элементами кластеров будет некорректным. Для решения этой проблемы была разработана и применена следующая метрика:

$$f(i, j) = w_1 * l + w_2 * v + w_3 * a, \quad (1)$$

i, j – точки из набора данных; w_1, w_2, w_3 – весовые коэффициенты расстояния, скорости и амплитуды соответственно;

$$l = \sqrt{\left(\sum_{k=0}^n \frac{x_{ik}}{n} - \sum_{k=0}^m \frac{x_{jk}}{m}\right)^2 + \left(\sum_{k=0}^n \frac{y_{ik}}{n} - \sum_{k=0}^m \frac{y_{jk}}{m}\right)^2} \quad (2)$$

$$v = \left| \sum_{k=0}^n \frac{v_{ik}}{n} - \sum_{k=0}^m \frac{v_{jk}}{m} \right| \quad (3)$$

$$a = \min\left(\sum_{k=0}^n \frac{a_{ik}}{n}; \sum_{k=0}^m \frac{a_{jk}}{m}\right) \quad (4)$$

где x, y, v, a – усредненные координаты, скорость, амплитуда точек в наборах, описанных в разделе 3.2., n, m – количество точек в этих наборах. Для каждого из параметров (расстояния, скорости и амплитуды) были использованы соответствующие пороговые значения для расчета глобального эпсилон, используемого в алгоритме DBSCAN:

$$EPS = w_1 * eps_l + w_2 * eps_v + w_3 * eps_a, \quad (5)$$

После применения алгоритма из результирующего множества кластеров удаляются кластеры с большим количеством элементов, так как в реальных случаях такие кластеры, скорее всего, являются сигналами, полученными в результате трения объекта об опору, не являющуюся дефектом. В зависимости от того, как проводился эксперимент, большими считаются кластеры из 10-20 элементов.

В результате полученные кластеры отображают области на объекте, которые с наибольшей вероятностью имеют дефекты и должны быть проверены специалистом.

4.4. Классификация

Классификация сигналов реализуется с помощью метода супер-сигналов из статьи [4]. С помощью коэффициента корреляции Пирсона для каждого кластера высчитываются супер-сигналы, как представители этого кластера, с которыми в дальнейшем будут сравниваться новые сигналы. Во время публикации статьи метод был реализован на языке программирования C++, поэтому в данной работе было необходимо его переписать на PYTHON.

4.5. Вспомогательные методы

Для реализации библиотеки также использовались вспомогательные методы из нескольких популярных библиотек, из-за которых в частности для разработки и был выбран язык программирования PYTHON. Для удобной работы с массивами данных применялись методы из библиотек Pandas и NumPy. Pandas позволяет эффективнее работать с двумерными данными различных типов, а также обрабатывать данные при отсутствии некоторой части информации. NumPy используется для применения математических методов к многомерным массивам данных и упрощает работу с ними.

Также использовались методы из библиотеки scikit-learn для кластеризации данных АЭ. Scikit-learn предоставляет различные методы машинного обучения, один из которых - используемый в данной работе алгоритм DBSCAN.

Для модульного тестирования использовался популярный фреймворк Pytest, позволяющий удобно работать с юнит-тестами.

5. Эксперимент

Для проверки новых разработанных методов использовались данные нескольких реальных экспериментов. Образование и развитие дефектов в процессе нагружения объекта контролировали путем регистрации сигналов акустической эмиссии в режиме реального времени с помощью системы Amsy-5 Vallen (Германия) по методике [14].

В первом эксперименте объектом исследования был цилиндрический сосуд высокого давления: сепаратор из стали 09Г2С-17+08Х13 высотой 3200 [мм], внутренним диаметром 2000 [мм] и толщиной стенки 12 [мм]. Было проведено гидравлическое испытание: сосуд был заполнен водой и находился под возрастающим давлением в течение 50 минут. В процессе загрузки в объекте появилось несколько трещин, которые зафиксировали специалисты. Исходные данные содержали 24579 строк данных датчиков, из которых после первичной обработки данных осталось 4856 строк, что составляет 1061 сигнал.

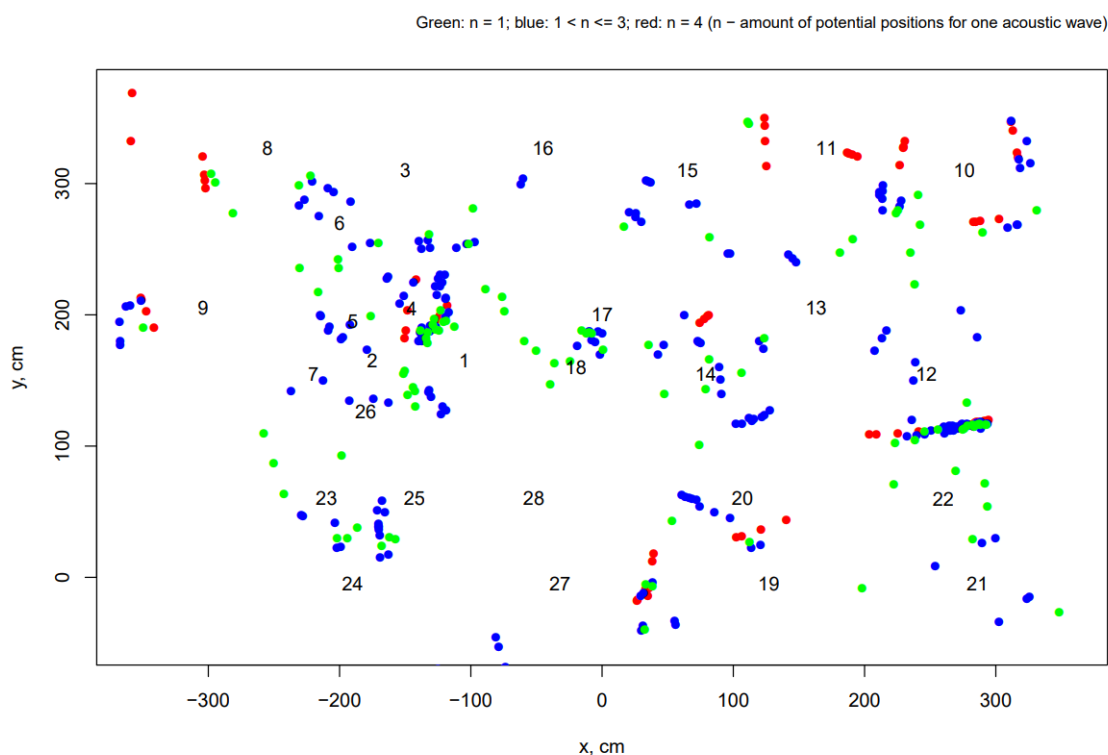


Рис. 4: Множество потенциальных позиций источников сигналов

Во втором эксперименте объектом исследования служил сепаратор

низкого давления из стали 09Г2С-12+08Х13 высотой 6674 [мм], внутренним диаметром 2400 [мм]. и толщиной стенки 14 [мм]. Вместимость аппарата - 31,0 м³. Проведено то же гидравлическое испытание, что и в первом эксперименте.

После применения метода локации к данным первого эксперимента были рассчитаны наборы возможных координат источников АЭ для 206 сигналов из набора данных. Для каждого из результирующих сигналов были получены наборы и отображены на одной диаграмме местоположения на рис. 4.

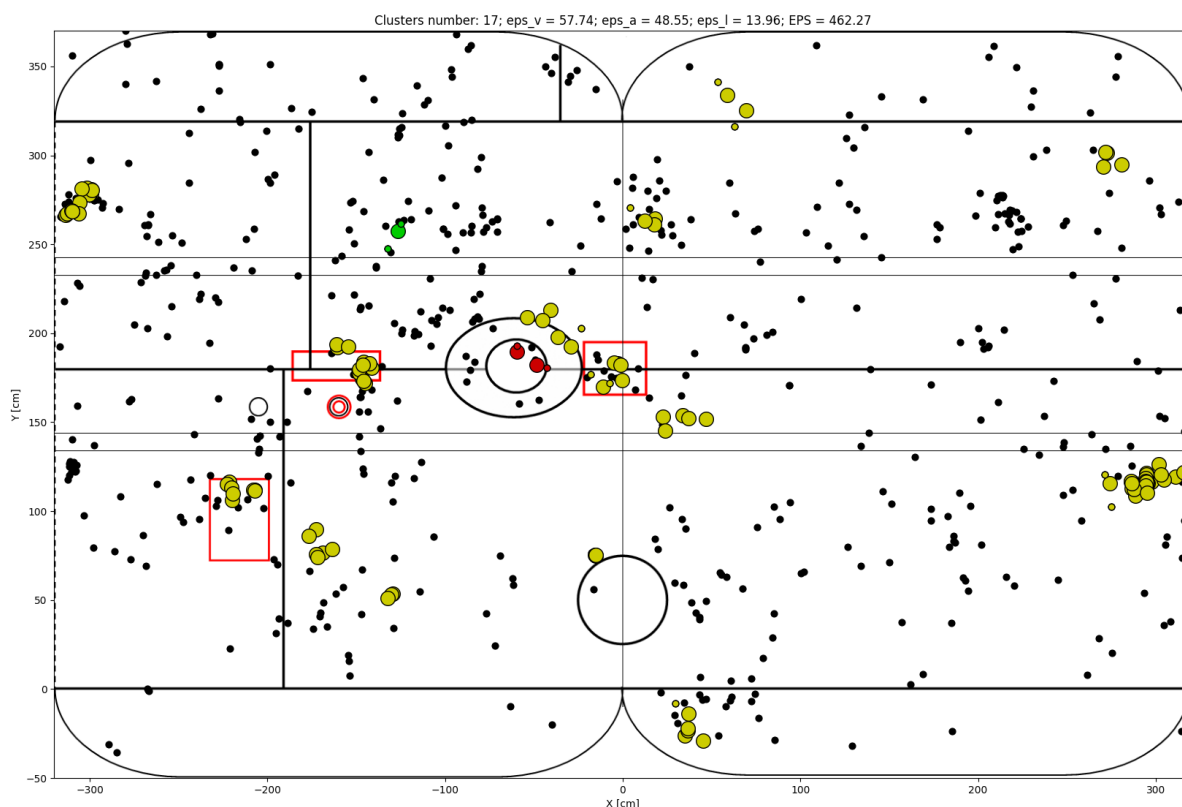


Рис. 5: Результат работы метода для первого эксперимента

В результате кластеризации было получено 17 кластеров. Сигналы с 1 решением показаны зеленым, сигналы с 2 или 3 решениями – синим, сигналы с 4 решениями – красным. Важно отметить, что скорость рассчитывается для каждого сигнала и для каждого решения в отдельности.

Таким образом, на рис. 4 показаны все точки расположения источников АЭ, полученные данным методом для исследуемого объекта. При

этом скорости различных сигналов АЭ различны. Они могут быть разными даже для каждого из четырех решений одного сигнала. Области скопления точек совпадают с участками, указанными специалистами для дополнительного контроля, в связи с возможным наличием в этих участках дефектов. Показана правильность применения этого метода при обследовании объектов данного типа. Во втором эксперименте после применения метода локации был получен набор данных из 534 точек. В результате кластеризации было получено 5 кластеров.

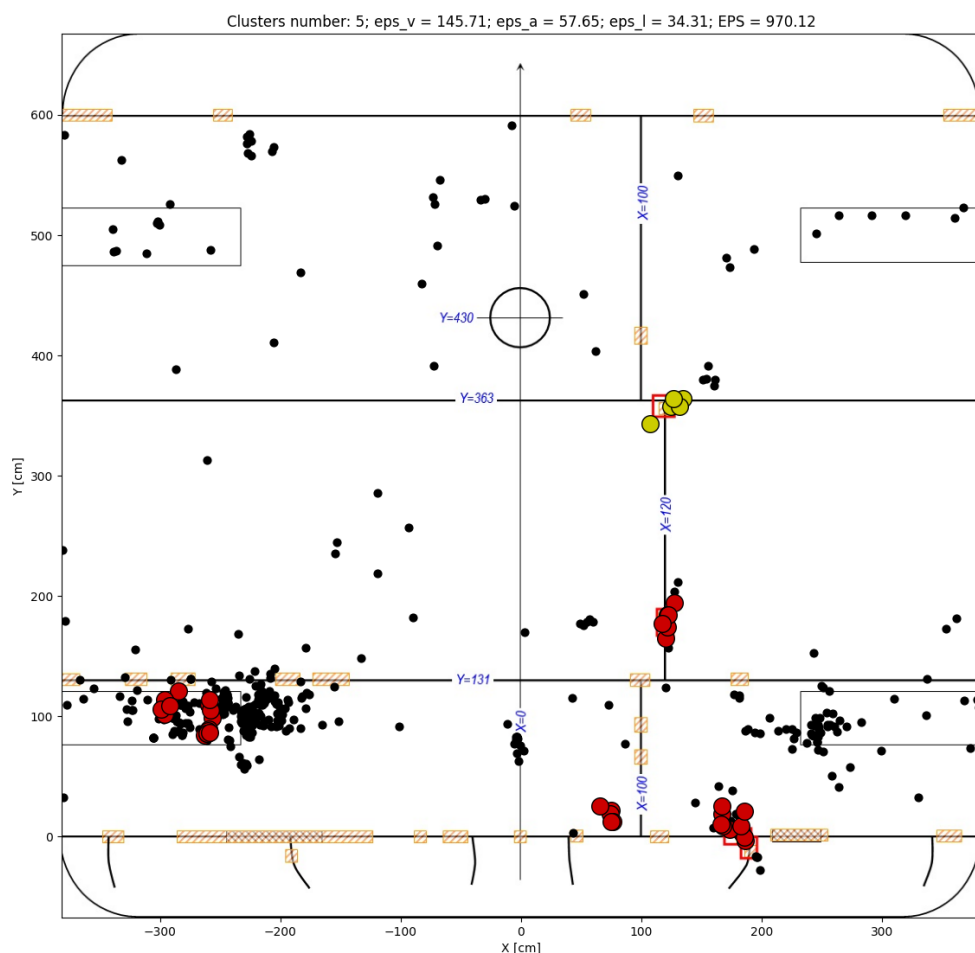


Рис. 6: Результат работы метода для второго эксперимента

На рис. 5 и рис. 6 красными прямоугольниками отмечены области, в которых действительно были обнаружены дефекты. Точки с амплитудой от 40 [дБ] до 55 [дБ] окрашены в зеленый цвет, с амплитудой от 55 [дБ]

до 60 [дБ] — в желтый, а с амплитудой более 60 [дБ] — в красный. Точки, не попавшие ни в один кластер, окрашены в черный цвет. На рис. 6 видно, что кластеры образовались рядом с каждой проблемной зоной, а несколько кластеров образовались вблизи области опор конструкции, где часто возникает много сигналов АЭ.

В третьем эксперименте проводилась калибровка датчиков, в ходе которой имитировались события АЭ с помощью специального прибора в разных местах объекта. После применения метода локации были корректно выделены 5 групп источников АЭ, соответствующие имитированным сигналам рис. 7.

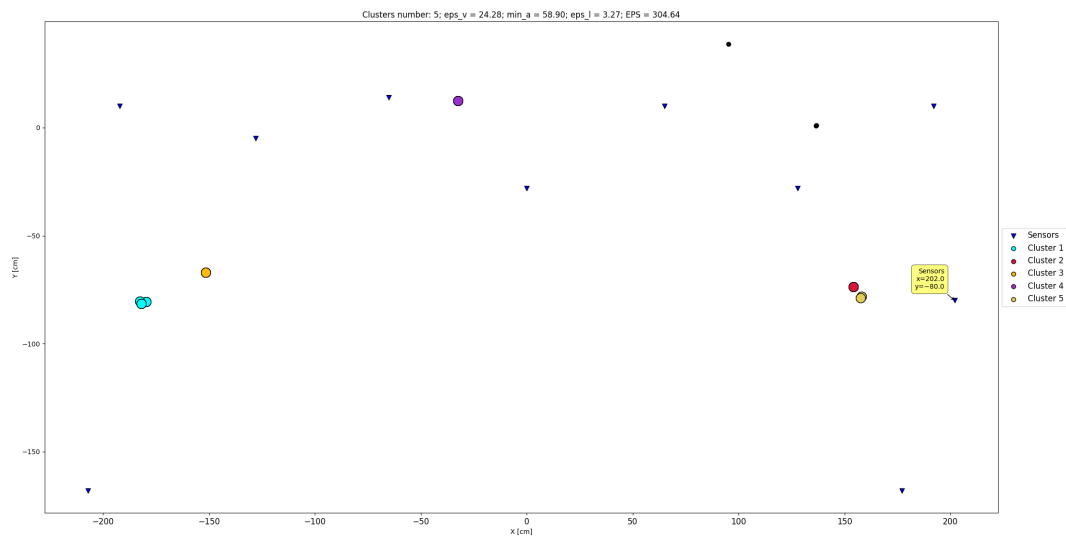


Рис. 7: Результат работы метода для третьего эксперимента

Кроме проверки корректности разработанных методов, также проводился анализ результатов различных алгоритмов кластеризации, применяемых к данным второго эксперимента. Для сравнение с разработанным методом были выбраны широко используемые алгоритмы: К-средних, Affinity Propagation и агломеративная иерархическая кластеризация. Для этого использовались их реализации из библиотеки scikit-learn. Также в них не использовалась разработанная метрика для расчета расстояния в виду особенностей этих алгоритмов.

Разработанный метод с использованием метрики получил результат,

отображенный на рис.8.

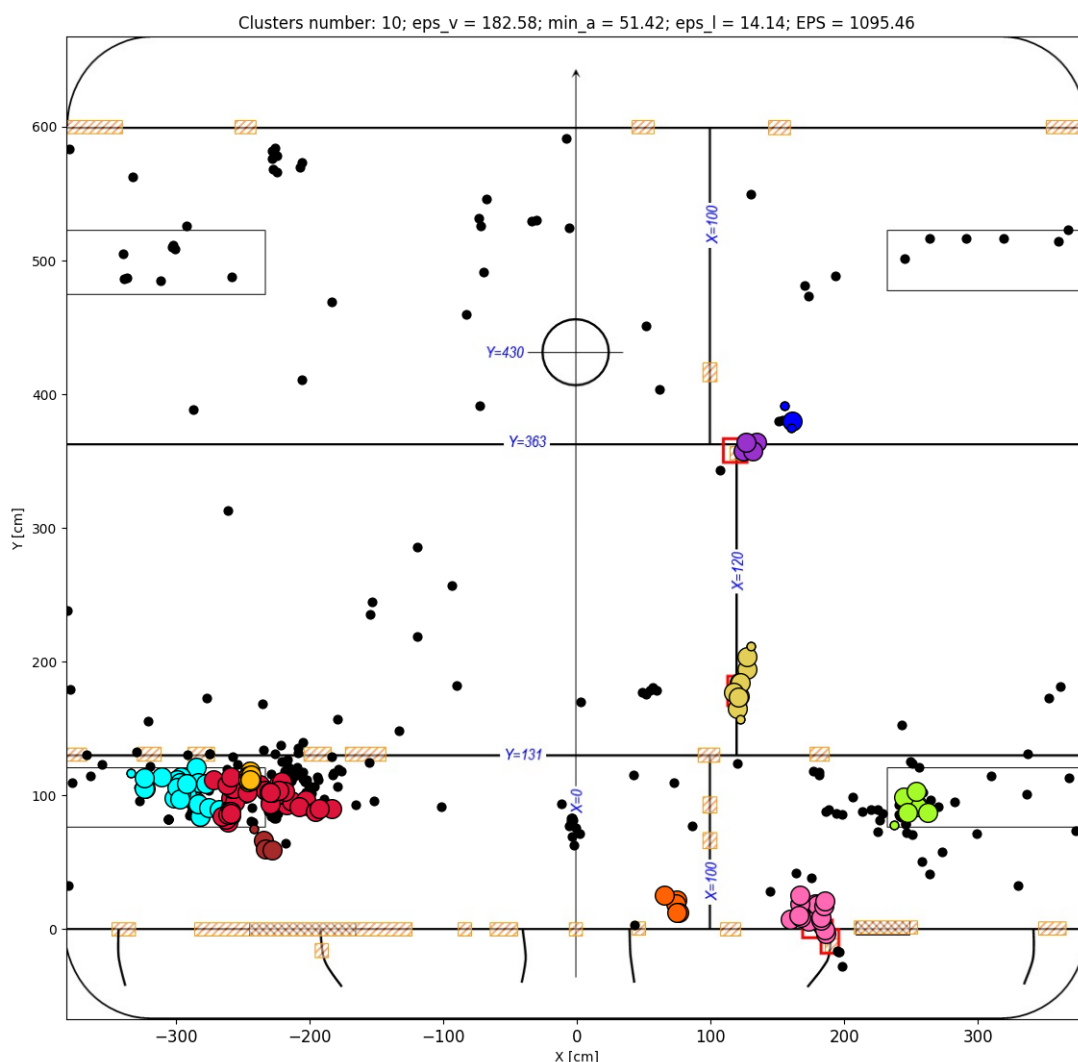


Рис. 8: Результат работы разработанного метода с использованием разработанной метрики

В данном случае было получено 10 кластеров выделенных разными цветами, которые сгруппированы в ожидаемых областях и отделены от шумов. Такое разбиение обусловлено тем, что учитываются не только амплитуда и скорость сигнала, но также и координаты, которые ограничивают кластеры по области, что достигнуто благодаря применению разработанной метрики. Также для сравнения был получен результат с использованием базовой метрики - евклидова расстояния (рис.9). Однако использование евклидова расстояния для данных параметров некорректно, так как различные параметры имеют различные диапазоны значений, иногда отличающиеся на порядки (например, скорости

могут варьироваться численно от 100 до 500, а амплитуды от 40 до 80), а следовательно должны учитываться по-разному. В данном случае было получено 19 кластеров, эpsilon расстояние было взято значение 35. Хотя и разбиение на кластеры достаточно похоже на рис. 8, с точки зрения АЭ оно нарушает некоторые моменты. Так, например, бирюзовый кластер в левой нижней части рисунка в длину достигает почти 1 метра, что весьма значительно и не может быть связано с одним дефектом.

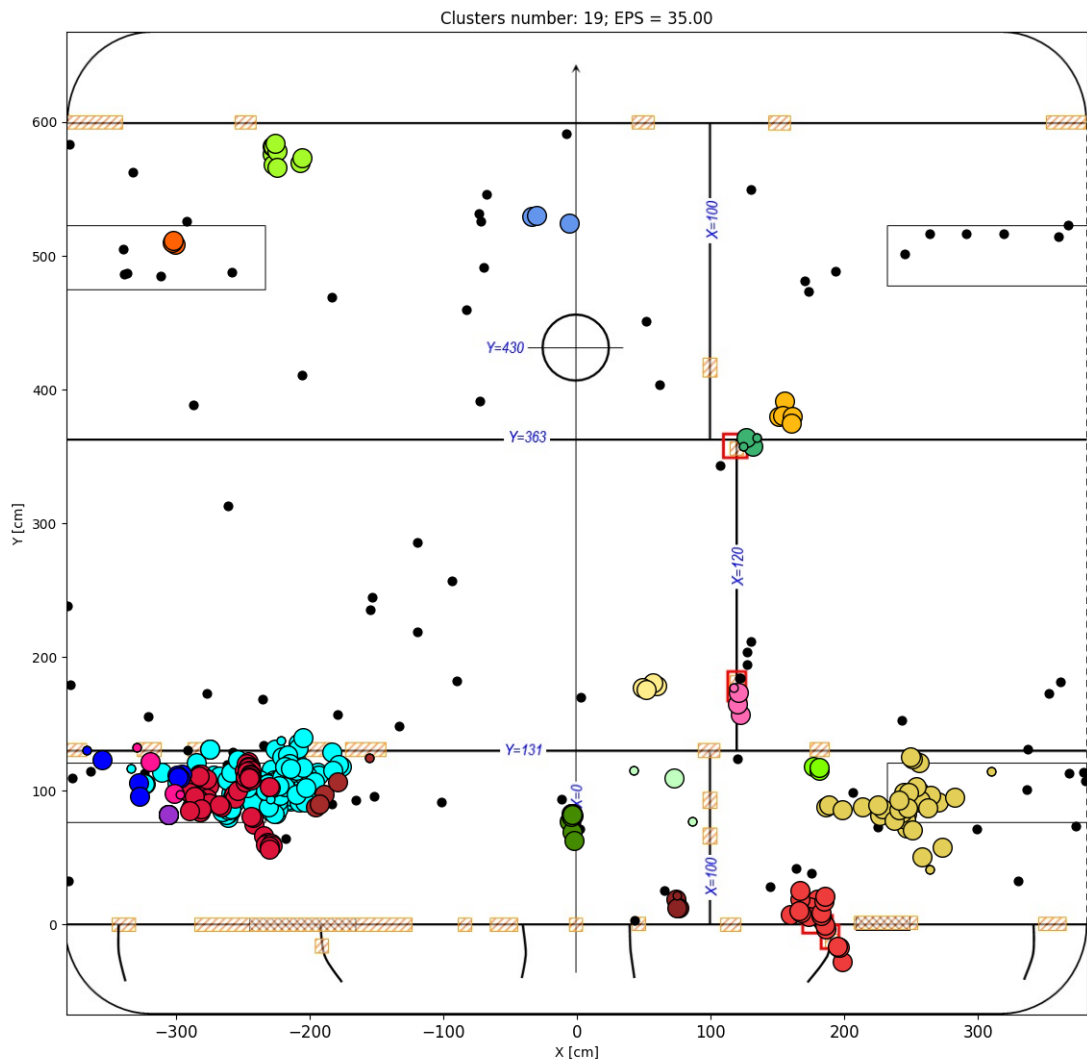


Рис. 9: Результат работы разработанного метода с использованием евклидова расстояния

Следующим исследуемым алгоритмом стал метод К-средних. Так как он требует заранее указывать количество кластеров, а также использует псевдослучайное число `random_state` для начального определения центров кластеров, было проведено несколько запусков для разных

значений. Также К-средних не позволяет использовать свою метрику в реализации библиотекой `scikit-learn`. На рисунках 10 и 11 показаны несколько результатов. У метода К-средних, применяемого к таким данным, возникает несколько проблем. Во-первых, он разбивает по кластерам все элементы, не отбрасывая шумы (например, обособленные точки, вокруг которых нет соседей), что обусловлено самим алгоритмом. Во-вторых, некорректным использованием параметров обычной метрикой, считающей их равнозначными. Таким образом, с точки зрения АЭ данные результаты являются некорректными и бессмысленными. Например, рассчитанные центры кластеров (точки большего размера на рис. 10 и рис.11) удалены от своих кластеров на значительное расстояние.

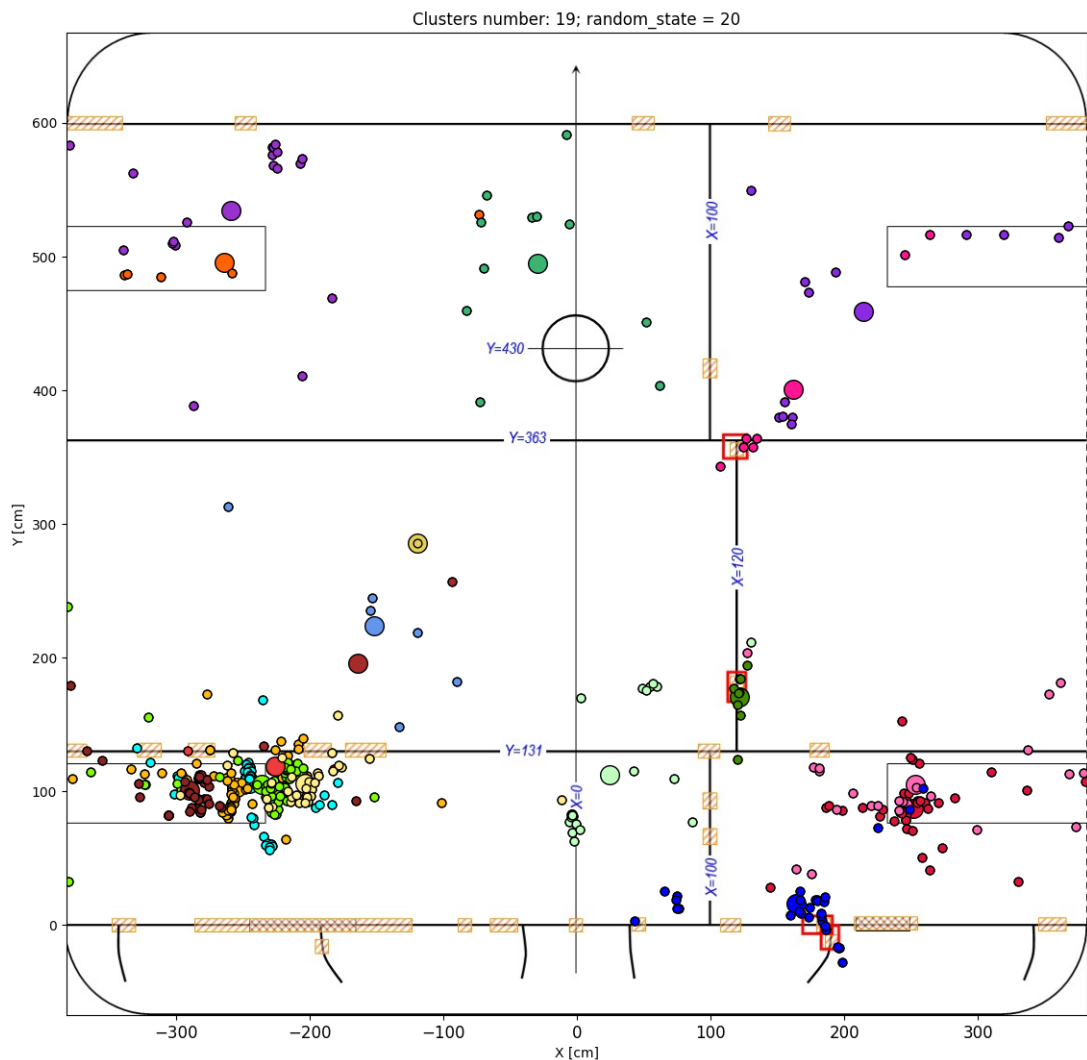


Рис. 10: Результат работы метода К-средних для 19 кластеров и `random_state = 20`

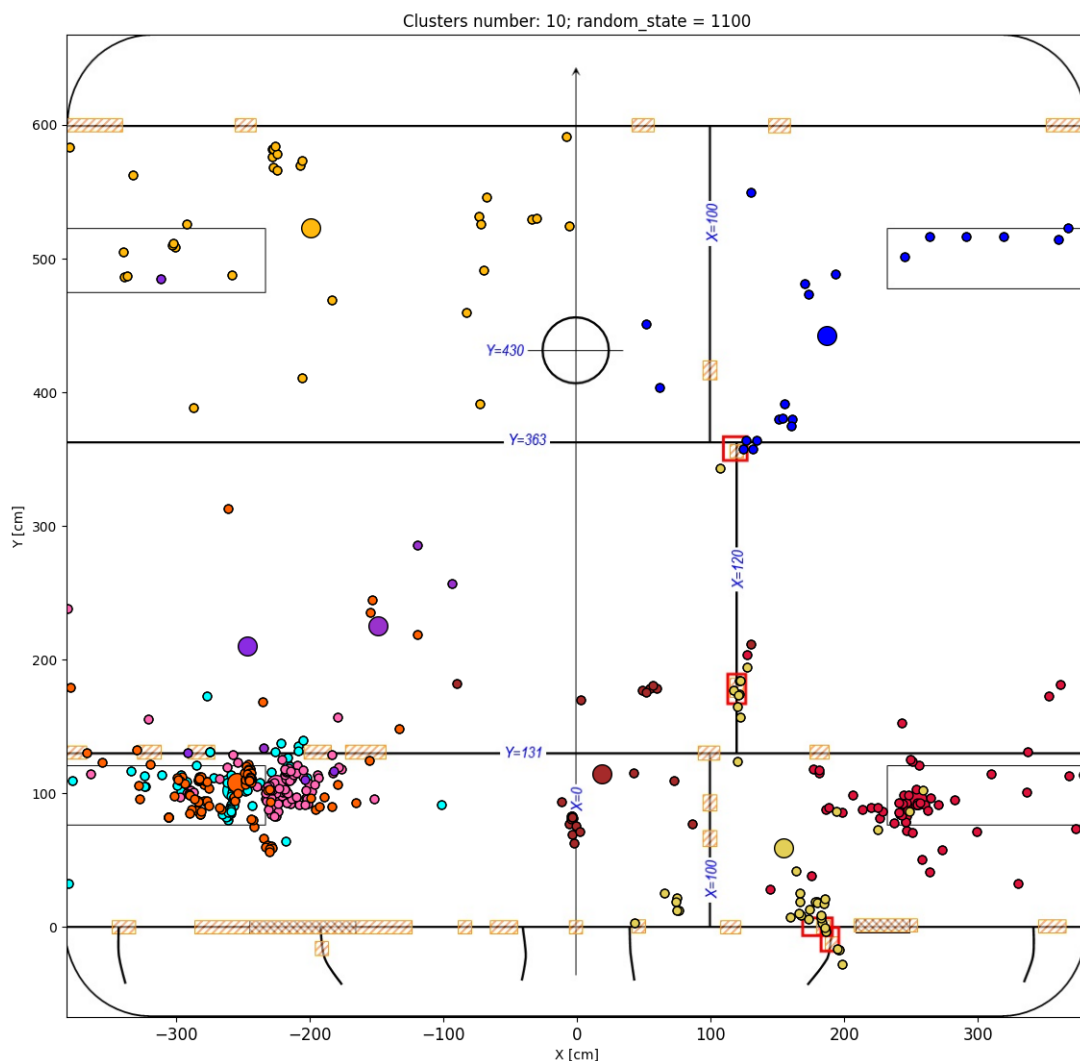


Рис. 11: Результат работы метода К-средних для 10 кластеров и $\text{random_state} = 1100$

Далее был использован алгоритм Affinity Propagation. Он не требует заранее знать число кластеров, но также как и К-средних выбирает случайно начальные значения, используя псевдослучайное число random_state . Вместо функции метрики реализация алгоритма в `scikit-learn` требует заранее посчитанной матрицы близости всех элементов, что может быть проблематично при большом их количестве. Были рассмотрены результаты при разных начальных значениях, однако при увеличении random_state , метод переставал сходиться и начинал некорректно формировать кластеры так, что их количество превышало 100. Для небольших же значений метод получил разбиение на 19 кластеров (рис 12), но их разбиение также было бессмысленным - элементы

кластеров были разбросаны очень далеко друг от друга.

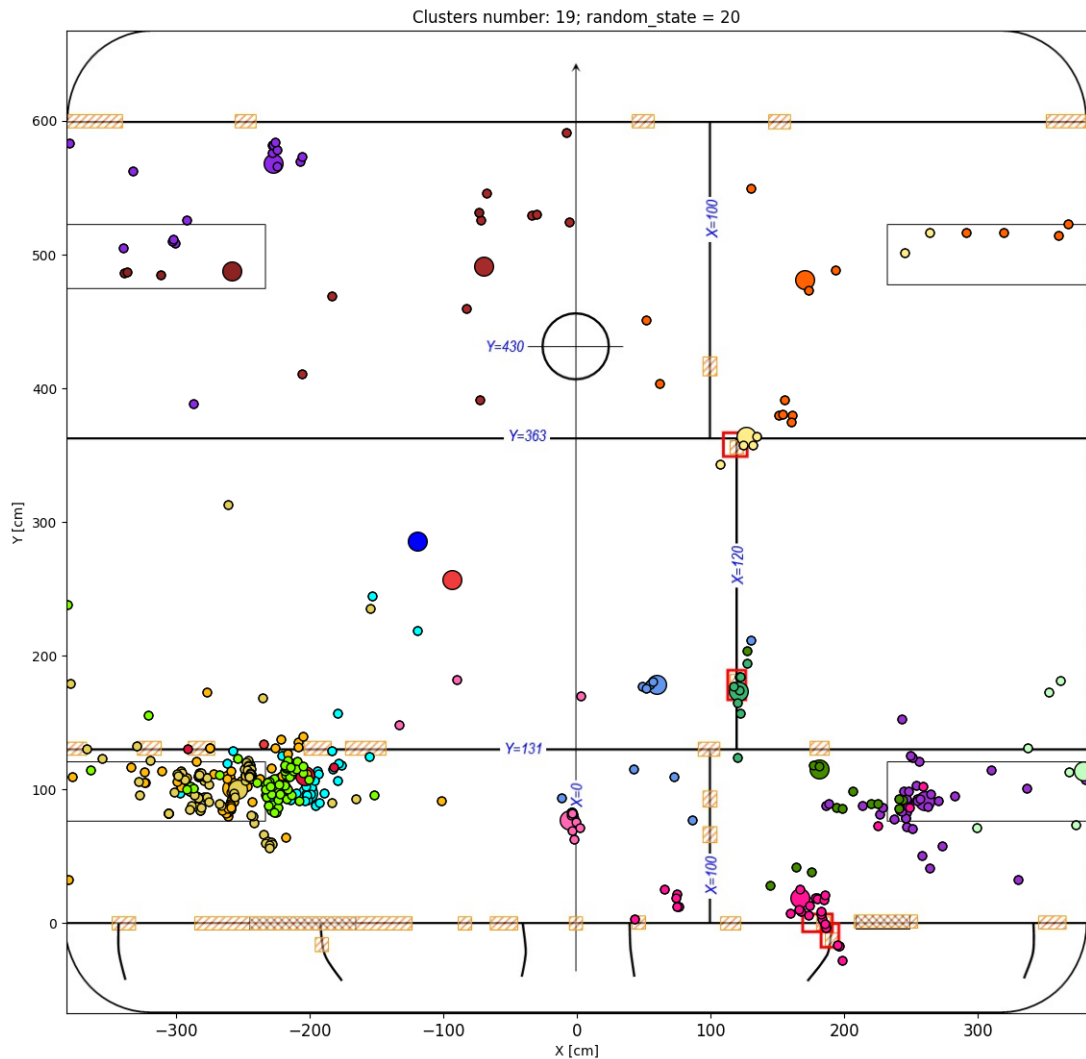


Рис. 12: Результат работы метода Affinity Propagation для 19 кластеров и `random_state = 20`

И последним рассмотренным алгоритмом была иерархическая агломеративная кластеризация. Ее реализация также есть в библиотеке `scikit-learn`. Данный алгоритм принимает на вход либо число кластеров, и тогда самостоятельно рассчитывает пороговое расстояние между элементами для объединения, либо пороговое расстояние, и затем сам определяет число кластеров. Примечательно то, что при разбиении на то же количество кластеров, что и описанным в работе методом, пороговое расстояние становится приблизительно равно эписилону, используемому в DBSCAN, однако по составу кластеры сильно отличаются (рис. 13). Элементы кластеров так же, как и в предыдущих случаях, располагаются

друг от друга на значительном расстоянии, что некорректно.

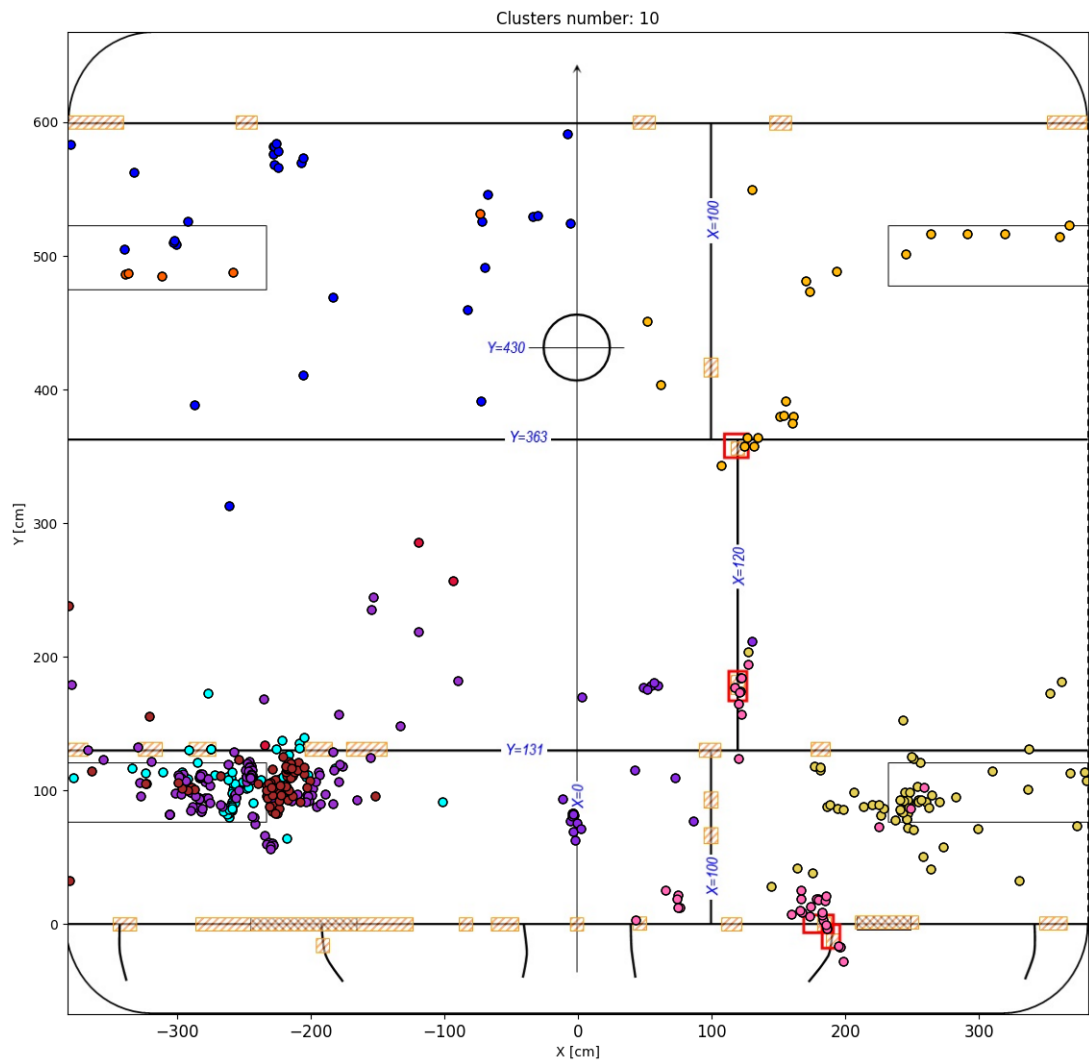


Рис. 13: Результат работы метода иерархической агломеративной кластеризации для 10 кластеров

В итоге, хотя рассмотренные результаты и получают такое же число кластеров, что и разработанный метод, однако объединение в эти кластеры происходит некорректно, что лишает смысла кластеризацию. Проведенный сравнительный анализ подтвердил преимущества DBSCAN в данной области и доказал необходимость использования специальной метрики для данной задачи.

Заключение

В ходе работы были выполнены следующие задачи:

- Разработанные за последние 3 года обучения в бакалавриате и магистратуре СПбГУ методы локации и классификации были переписаны на язык программирования Python;
- Была разработана метрика для вычисления расстояния между элементами кластеров с учетом значимости параметров;
- Был разработан метод кластеризации данных АЭ с использованием алгоритма DBSCAN;
- Была сформирована библиотека методов;
- Методы библиотеки были апробированы на нескольких экспериментах.

Полученные результаты демонстрируют корректность применяемых методов для работы с реальными данными. Применение разработанного метода кластеризации значительно сокращает площади объекта, подлежащие проверке специалистом. На основе проведенного анализа различных методов кластеризации данных была написана статья "Сравнение и настройка разных алгоритмов машинного обучения для кластеризации акустических импульсов в промышленных объектах", которая была представлена на конференции СИТИМ-2023, входящую в Ринц. Разработанный метод кластеризации вошел в статью, которая будет опубликована в журнале PROCEEDINGS ON ENGINEERING SCIENCES, входящий в Scopus.

Объединение методов в библиотеку упрощает процесс анализа данных АЭ, а также позволяет пользователям модифицировать их и применять вместе с другими методами. Также в дальнейшем возможно расширение функционала библиотеки.

Список литературы

- [1] Acoustic Emission and Deep Learning for the Classification of the Mechanical Behavior of AlSi10Mg AM-SLM Specimens / Claudia Barile, Caterina Casavola, Giovanni Pappaletta et al. // Applied Sciences. — 2022. — Vol. 13, no. 1. — P. 189.
- [2] Agletdinov Einar, Merson Dmitry, Vinogradov Alexei. A new method of low amplitude signal detection and its application in acoustic emission // Applied Sciences. — 2019. — Vol. 10, no. 1. — P. 73.
- [3] Assessment of cooperativity in complex systems with non-periodical dynamics: Comparison of five mutual information metrics / Nikita S Pyko, Svetlana A Pyko, Oleg A Markelov et al. // Physica A: Statistical Mechanics and its Applications. — 2018. — Vol. 503. — P. 1054–1072.
- [4] Balagurin Pavel, Makarikhin Pavel, Grigorieva Anastasia. Assessment of the Integrity of the Object Based on the Correlation of Super-Signals. — 2020. — P. 73–80.
- [5] Barat Vera, Borodin YRIJ, Kuzmin ALEXEY. INTELLIGENT AE SIGNAL FILTERING METHODS. // Journal of Acoustic Emission. — 2010. — Vol. 28.
- [6] Calabrese L, Campanella G, Proverbio E. USE OF CLUSTER ANALYSIS OF ACOUSTIC EMISSION SIGNALS IN EVALUATING DAMAGE SEVERITY IN CONCRETE STRUCTURES. // Journal of Acoustic Emission. — 2010. — Vol. 28.
- [7] Classification of acoustic emission signals using wavelets and Random Forests: Application to localized corrosion / Nicolas Morizet, Nathalie Godin, J Tang et al. // Mechanical Systems and Signal Processing. — 2016. — Vol. 70. — P. 1026–1037.
- [8] DONG Long-jun, LI Xi-bing. Three-dimensional analytical solution of acoustic emission or microseismic source location under cube monitoring

- network // Transactions of Nonferrous Metals Society of China. — 2012. — Vol. 22, no. 12. — P. 3087–3094.
- [9] EAS observation conditions in the SPHERE-2 balloon experiment / Elena Bonvech, Dmitry Chernov, Miroslav Finger et al. // Universe. — 2022. — Vol. 8, no. 1. — P. 46.
- [10] Estimating the Probability of Detecting a Delamination in the Wall of Equipment Depending on the Set of Used Methods of Nondestructive Testing and Ways of Its Improvement / IA Rastegaev, VP Gomera, SA Tyupin et al. // Russian Journal of Nondestructive Testing. — 2018. — Vol. 54. — P. 619–629.
- [11] Evaluating the Probability of Detecting Acoustic Emission Signals Using the Amplitude Discrimination Method with a Low Signal-to-Noise Ratio / IA Rastegaev, IS Yasnikov, II Rastegaeva et al. // Russian Journal of Nondestructive Testing. — 2022. — Vol. 58, no. 3. — P. 157–166.
- [12] Frey Brendan J, Dueck Delbert. Clustering by passing messages between data points // science. — 2007. — Vol. 315, no. 5814. — P. 972–976.
- [13] Identification of acoustic-emission sources during testing of technological equipment with a high noise level / DG Davydova, AN Kuz'min, RG Rizvanov, EG Aksel'rod // Russian Journal of Nondestructive Testing. — 2015. — Vol. 51. — P. 292–302.
- [14] In-situ estimation of defect volume from parameters of acoustic emission signals / EE Damaskinskaya, VL Hilarov, IA Panteleev et al. — 2021. — Vol. 2103, no. 1. — P. 012066.
- [15] Jingpin Jiao, Bin Wu, Cunfu He. Acoustic emission source location methods using mode and frequency analysis // Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures. — 2008. — Vol. 15, no. 4. — P. 642–651.

- [16] Long Feifei, Xu Haifeng. Validity identification and classification technique of tank acoustic emission testing signals based on clustering analysis. — 2011. — Vol. 3. — P. 2000–2003.
- [17] Mostafapour A, Davoodi S, Ghareaghaji M. Acoustic emission source location in plates using wavelet analysis and cross time frequency spectrum // Ultrasonics. — 2014. — Vol. 54, no. 8. — P. 2055–2062.
- [18] New automatic localization technique of acoustic emission signals in thin metal plates / Petr Sedlak, Yuichiro Hirose, Sabrina A Khan et al. // Ultrasonics. — 2009. — Vol. 49, no. 2. — P. 254–262.
- [19] Nosov VV, Zelenskii NA. Estimating the strength of welded hull elements of a submersible based on the micromechanical model of temporal dependences of acoustic-emission parameters // Russian journal of non-destructive testing. — 2017. — Vol. 53, no. 2. — P. 89–95.
- [20] Ramasso Emmanuel, Denoeux Thierry, Chevallier Gael. Clustering acoustic emission data streams with sequentially appearing clusters using mixture models // Mechanical Systems and Signal Processing. — 2022. — Vol. 181. — P. 109504.
- [21] Vaezi Yoonas, Van der Baan Mirko. Comparison of the STA/LTA and power spectral density methods for microseismic event detection // Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society. — 2015. — Vol. 203, no. 3. — P. 1896–1908.
- [22] Yaroslavkina E. Investigation of methods for determining the integral characteristics of periodic signals separated in time. — 2018. — P. 1–4.
- [23] Zhang Zhong Ning, Tian Jian. A New Acoustic Emission Source Location Method Based on the Linear Layout of Sensors. — 2011. — Vol. 267. — P. 561–564.
- [24] A density-based algorithm for discovering clusters in large spatial databases with noise. / Martin Ester, Hans-Peter Kriegel, Jörg Sander et al. — 1996. — Vol. 96, no. 34. — P. 226–231.

- [25] A time-frequency based approach for acoustic emission assessment of sliding wear / Igor Rastegaev, Dmitry Merson, Inna Rastegaeva, Alexei Vinogradov // Lubricants. — 2020. — Vol. 8, no. 5. — P. 52.
- [26] Гомера Виктор Петрович, Смирнов Анатолий Джумамуратович, Нефедьев Евгений Юрьевич. Повышение достоверности АЭ-контроля оборудования, содержащего внутренние устройства // В мире неразрушающего контроля. — 2014. — no. 3. — P. 20–24.
- [27] Григорьева А. В., Максименко М. В. Метод обработки данных акустико-эмиссионного контроля для определения скорости и локации каждого сигнала // Компьютерные исследования и моделирование. — 2022. — Vol. 14, no. 5. — P. 1029–1040.
- [28] Иванов В.И., Бигус Г.А., Власов И.Э. Акустическая эмиссия: учеб. Пособие/под общ. Ред. ВВ Ключева.—2-е изд. М.: Издательский дом «Спектр», 2015.—192с.: ил // Диагностика безопасности.
- [29] Макарихин П. А. Обработка данных акустической сигналов для применения к ним алгоритмов классификации и кластеризации : ВКР бакалавра / А. Макарихин, П. ; Санкт-Петербургский Государственный Университет. — 2021.