



Санкт-Петербургский государственный университет
Кафедра системного программирования

Разработка библиотеки методов обработки данных акустической эмиссии на языке Python

Павел Андреевич Макарихин, 21.M07-мм группа

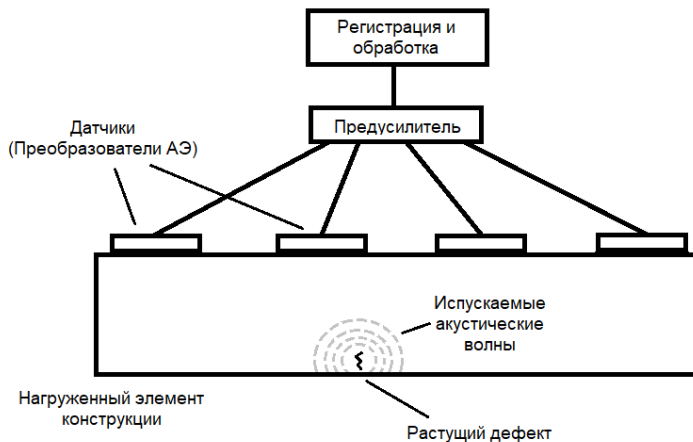
Научный руководитель: к.т.н., доц. каф. информатики А.В. Григорьева

Санкт-Петербург
2022

Введение. Неразрушающий контроль

- Акустический
- Электромагнитный
- Оптический
- Радиационный
- Ультразвуковой
- ...

Введение. Акустическая эмиссия



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

- Локация источников
- Кластеризация сигналов
- Классификация сигналов

Различные подходы для локации источников:

- Анализ мод волны
- Вейвлет-анализ волн
- Дополнительные условия
- Аналитические решения
- Анализ на основе априорно известных значений

Существует множество алгоритмов кластеризации:

- DBSCAN
- К-средних
- Affinity Propagation
- Агломеративная иерархическая кластеризация
- Нейронные сети
- прочие

Существующие подходы классификации:

- По формам волн
- По различным физическим параметрам
- С помощью эмпирических знаний
- Нейронные сети
- прочие

Широко используемое ПО:

- Vallen AE Suite Software
- Physical Acoustics' AEWIn и Noesis
- A-Line

Vallen Systeme также разработала библиотеку pyVallenAE, однако, там очень небольшой функционал для анализа.

ВКР магистра Жибаря Марка Артуровича, защищенная на кафедре программной инженерии в 2022 году: Программный комплекс для анализа данных АЭ, состоящий из нескольких приложений, написанных на разных языках программирования (Java и Python).

Отличия данной работы

- Единая библиотека
- Разные применяемые методы



Первичная обработка

```

Id  HMMSS      MSEC CHAN  A      TRAI  E(TE)      R      D THR      RMS  CNTS  ALIN
[hmmss] [ms.µs]  [dB]
La Label 1: '14:11 Resume'
DT 7 Ноябрь 2017 г., Host Time: 14:11
Ev00:00:03 66,2965 3 41,1 215E-1 0,4 5,0 40,0 3,7 1 114
Ev00:00:18 783,5382 10 44,1 986E-1 40,2 195,2 40,0 4,5 7 161
Ev00:00:19 223,2265 4 45,2 1 896E-1 7,4 76,0 40,0 3,8 4 183
Ev00:00:24 41,7766 4 48,6 2 143E00 56,6 173,0 40,0 3,7 16 270
Ev00:00:36 840,8211 2 48,3 3 719E-1 10,4 55,8 40,0 3,7 6 259
Ev00:00:49 313,9390 3 43,7 4 116E00 5,0 219,0 40,0 3,7 9 154
Ev00:01:01 313,9168 3 40,7 297E-1 0,2 0,8 40,0 3,7 1 109
Ev00:01:17 277,1777 3 52,0 5 211E00 37,4 120,0 40,0 3,7 11 399
Ev00:01:20 563,9134 3 42,6 334E-1 4,4 9,2 40,0 3,7 1 135
Ht00:01:20 564,0599 9 45,2 125E00 121,2 194,0 40,0 4,5 6 183
Ev00:01:21 537,8829 4 48,6 6 430E00 223,4 626,0 40,0 3,8 28 270
Ev00:01:34 14,4855 2 44,5 7 148E00 66,0 300,8 40,0 4,2 11 168
Ht00:01:34 14,6637 6 42,6 8 797E-1 41,8 161,0 40,0 3,9 4 135
Ev00:01:36 160,8530 3 41,9 489E-1 4,6 27,4 40,0 3,7 2 124
Ev00:01:48 669,3676 10 53,2 100E01 197,0 435,0 40,0 4,5 44 455
Ev00:02:13 728,6979 10 52,0 184E00 8,0 81,6 40,0 4,5 9 399
Ev00:02:51 261,4181 1 45,2 9 139E00 70,8 156,4 40,0 4,0 8 183
Ev00:03:01 459,7859 1 43,0 209E-1 29,2 33,8 40,0 4,0 1 141
LE00:03:04 543,9642 1 55,4 10 279E01 47,0 2468,0 40,0 4,0 131 589
Ht00:03:04 544,0345 7 53,5 11 209E01 464,2 2153,6 40,0 3,5 110 475
Ht00:03:04 544,0361 3 74,2 12 262E03 139,0 5803,2 40,0 3,7 540 5142
Ht00:03:04 544,0429 9 93,4 318E04 384,0 3601,6 40,0 4,5 314 46848
Ht00:03:04 544,0595 5 63,7 13 245E02 126,8 3307,2 40,0 3,6 308 1529
Ht00:03:04 544,0896 10 56,5 503E01 523,6 2617,6 40,0 4,4 169 671
Ht00:03:04 544,1383 8 55,4 14 476E01 814,8 2293,6 40,0 4,3 156 589
LE00:03:10 761,0621 1 55,8 18 259E01 51,6 1910,4 40,0 4,0 116 616
Ht00:03:10 761,1338 3 74,6 19 160E03 143,4 4926,4 40,0 3,7 429 5370

```

#	Id	Channel	Time	MGec	Amplitude	Energy	Duration	Counts
0	LE	9	00:00:43	937530	88,5	9370000	11894,4	1174
0	HI	1	00:00:43	9375959	96,5	3800	3500,8	161
0	HI	7	00:00:43	9375956	60,3	11300	3237,6	255
0	HI	3	00:00:43	9376193	63,3	25000	4144	356
0	HI	5	00:00:43	9376333	99,9	9480	2916,8	214
0	HI	8	00:00:43	9376708	54,7	4350	2533,6	191
0	HI	6	00:00:43	9376803	53,5	2140	2072	118
1	LE	1	00:00:47	7141407	57,7	3830	3544	179
1	HI	9	00:00:47	714143	82,3	13300000	8819,2	995
1	HI	7	00:00:47	7142131	62,6	18100	3229,6	264
1	HI	3	00:00:47	7142197	69,3	51700	4107,2	348
1	HI	5	00:00:47	7142382	64,4	18400	3000,8	243
1	HI	8	00:00:47	7142434	58	6010	2740,8	211
1	HI	10	00:00:47	7142654	52,8	3150	2626,4	166
2	LE	9	00:00:51	8014368	85,5	3060000	7366,4	718
2	HI	1	00:00:51	8014843	95	1640	1577,2	74
2	HI	7	00:00:51	8015094	57,7	4320	2508,8	161
2	HI	3	00:00:51	8015151	61,8	13300	3073,6	248
2	HI	5	00:00:51	8015298	96,5	4220	2428,8	142
2	HI	6	00:00:51	8016042	49,8	807	1631,6	58
2	HI	8	00:00:51	8016172	52	2080	2307,2	128
3	LE	9	00:00:58	6017406	91,9	15100000	9504	890
3	HI	1	00:00:58	6017457	90,3	7240	3435,2	207
3	HI	3	00:00:58	6018186	96,7	54300	4300,8	375
3	HI	5	00:00:58	6018404	62,9	26100	3606,4	279

Локация источника

Максимальное расстояние источника сигнала, относительно первых зафиксировавших датчиков:

$$d_{max} = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{2}, \quad (1)$$

где x_i, y_i - координаты i -го датчика, по порядку фиксирования.

$$v_1 - v_2 < accuracy, \quad (2)$$

где accuracy - заранее заданная точность

$$\begin{cases} v_1 = \left| \frac{\sqrt{(x_2 - \bar{x})^2 + (y_2 - \bar{y})^2} - \sqrt{(x_1 - \bar{x})^2 + (y_1 - \bar{y})^2}}{t_2 - t_1} \right| \\ v_2 = \left| \frac{\sqrt{(x_3 - \bar{x})^2 + (y_3 - \bar{y})^2} - \sqrt{(x_2 - \bar{x})^2 + (y_2 - \bar{y})^2}}{t_3 - t_2} \right| \end{cases} \quad (3)$$

где \bar{x}, \bar{y} - координаты вершины сетки, x_i, y_i - координаты датчиков, t_i -

Для кластеризации данных был выбран алгоритм DBSCAN, поскольку он:

- Не требует заранее знать количество кластеров
- Дает одинаковый результат для одного и того же набора данных
- Устойчив к выбросам
- Достаточно прост для понимания специалистом

Алгоритм DBSCAN

Пусть i - основной элемент, если в его окрестности радиуса \mathcal{E} находится хотя бы m элементов (\mathcal{E} -соседей). Тогда абстрактно алгоритм состоит из следующих шагов:

- Для каждого элемента проверяем, основной он или нет
- Если основной, то формируем кластер и переходим на \mathcal{E} -соседей
- Если не основной, то добавляем его к кластеру ближайшего основного \mathcal{E} -соседа, иначе считаем шумом
- Для \mathcal{E} -соседей аналогично
- Если \mathcal{E} -соседей нет, то берем случайный следующий

Метрика расстояния для двух точек

$$f(i, j) = w_1 * l + w_2 * v + w_3 * a, \quad (4)$$

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

$$l = \sqrt{\left(\sum_{k=0}^n \frac{x_{ik}}{n} - \sum_{k=0}^m \frac{x_{jk}}{m}\right)^2 + \left(\sum_{k=0}^n \frac{y_{ik}}{n} - \sum_{k=0}^m \frac{y_{jk}}{m}\right)^2} \quad (5)$$

$$v = \left| \sum_{k=0}^n \frac{v_{ik}}{n} - \sum_{k=0}^m \frac{v_{jk}}{m} \right| \quad (6)$$

$$a = \min\left(\sum_{k=0}^n \frac{a_{ik}}{n}; \sum_{k=0}^m \frac{a_{jk}}{n}\right) \quad (7)$$

где x, y, v, a – усредненные координаты, скорость, амплитуда из наборов, полученных в результате локализации, n, m – количество точек в этих наборах.

Гиперпараметр \mathcal{E} , использующийся в DBSCAN, рассчитывается по формуле:

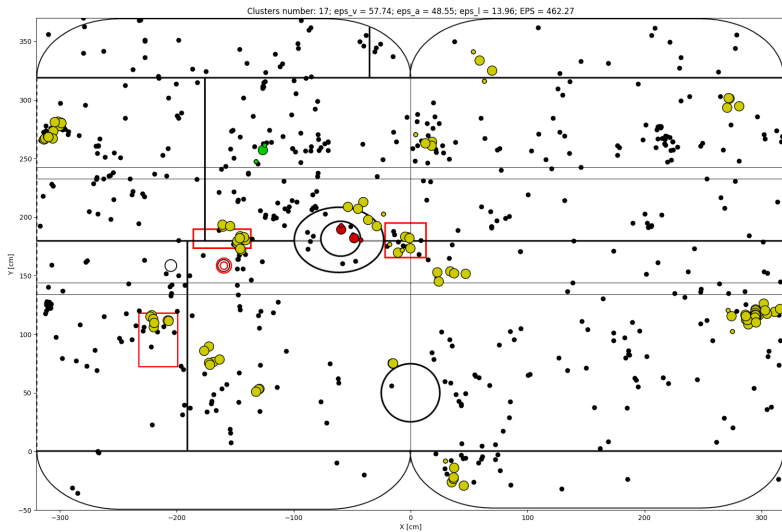
$$\mathcal{E} = w_1 * \mathcal{E}_l + w_2 * \mathcal{E}_v + w_3 * \mathcal{E}_a \quad (8)$$

Для расчета \mathcal{E}_l , \mathcal{E}_v , \mathcal{E}_a выбирается одна из зон, в которой известно, что был дефект. Далее на основе значений параметров точек, попавших в выбранную зону происходит расчет.

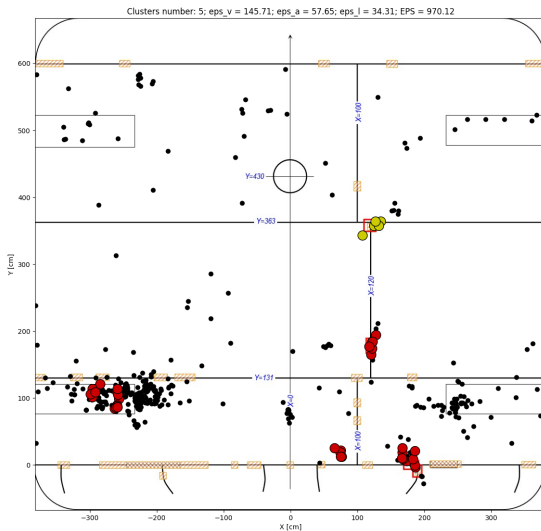
Весовые коэффициенты w_1 , w_2 и w_3 можно задать вручную, либо также рассчитать автоматически.

Классификация сигналов реализуется с помощью метода супер-сигналов из статьи *Assessment of the Integrity of the Object Based on the Correlation of Super-Signals*. С помощью коэффициента корреляции Пирсона для каждого кластера высчитываются супер-сигналы, как представители этого кластера, с которыми в дальнейшем будут сравниваться новые сигналы.

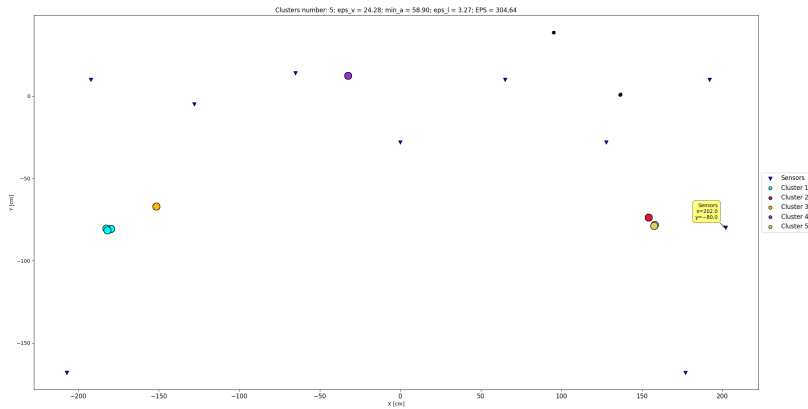
Эксперимент 1



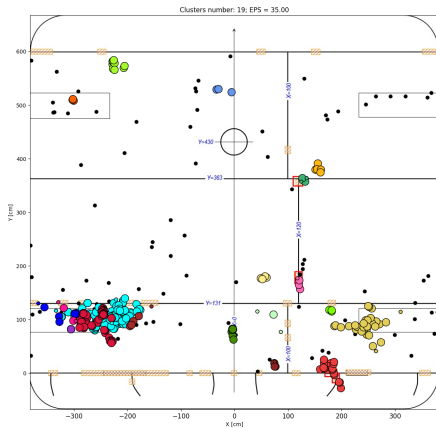
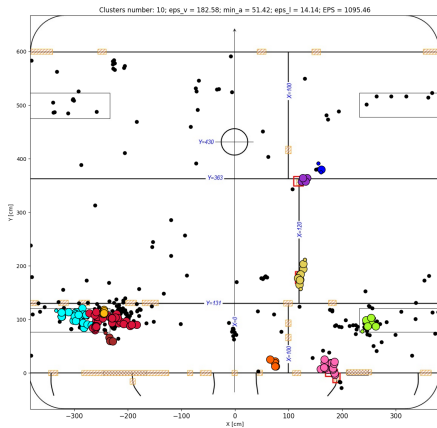
Эксперимент 2



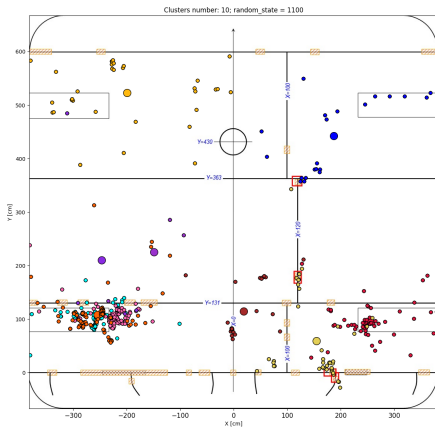
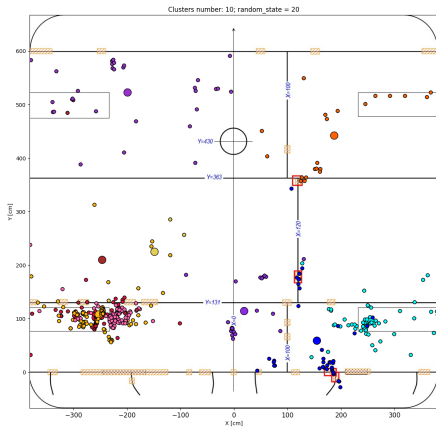
Эксперимент 3



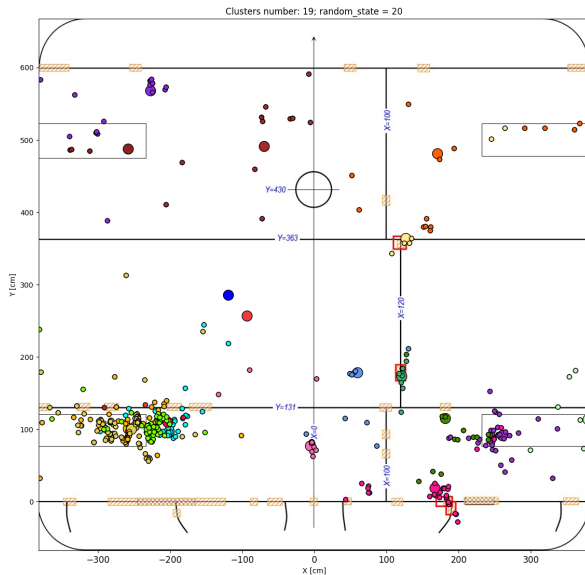
Использование разработанной метрики



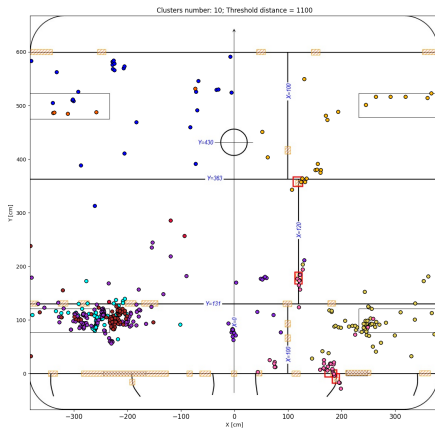
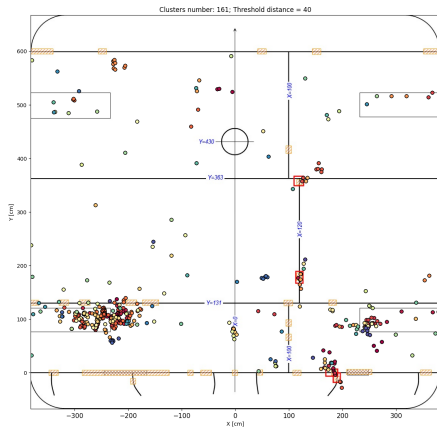
К-средних



Affinity Propagation



Агломеративная иерархическая кластеризация



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.

На основе новых методов было написано 2 статьи, одна из которых вошла в перечень РИНЦ, а другая - в Scopus.