Алгоритм КЛАСТРИЗАЦИИ РАДИОЛОКАЦИОННЫХ ПАТТЕРНОВ В ПАССИВНЫХ РАДИОЛОКАЦИОННЫХ КОМПЛЕКСАХ

О. В. Глухов1, Н.В. Масалкова1, Н.И. Ефремов1, Р. С. Куликов1

1ФГБОУ ВО «НИУ «МЭИ», Москва, Россия

**Аннотация.** В данной статье рассматривается применение алгоритмов кластеризации сигнальных паттернов для увеличения эффективности пассивных радиолокационных комплексов на этапе обнаружения отметок от целей. Решение такой задачи позволит в будущем подбирать более эффективные алгоритмы вторичной (траекторной) обработки целей.

Для решения задачи кластеризации радиолокационных паттернов был разработан алгоритм кластеризации cDBSCAN – плотностной алгоритм пространственной кластеризации с присутствием шума, с применением каскада эвристических правил.

Результаты работы на полученной имитационной модели показали, что алгоритм верно определяет паттерны в выборке с высокой точностью, более 98%. При этом значение ложного обнаружения паттернов составляет около 2%.

**Ключевые слова:** cDBSCAN, DBSCAN, кластеризация, алгоритм, радиолокация, РЛС.

ALGORITHM of radar pattern clustering in passive radar systems

O.V. Glukhov1, N.V. Masalkova1, M.I. Efremov1, R. S. Kulikov1

1NRU MPEI, Moscow, Russia

**Abstract.**

**Keywords:**

# Введение

Распознавание целей в воздушном пространстве является одним из приоритетных направлений радиолокации. Обработка радиолокационной информации представляет собой важный спектр задач, решаемых при помощи радиолокационных станций (РЛС) или комплексов. Назначение обработки – подготовить к выдаче в требуемом виде полную, достоверную и современную информацию для потребителя о состоянии воздушной обстановки, появлении и местоположении воздушных объектов, параметрах их движения, возможных вариантах развития динамики изменения воздушно-помеховой обстановки. Характерным примером обработки радиолокационной информации является задача обнаружения летательных аппаратов (ЛА) и оценка их принадлежности к определенному классу или типу. Проблема аналогична известной задаче кластеризации и последующей классификации [1], необходимо из множества одиночных отметок (радиоимпульсов) выбрать несколько центров группирования, которые соответствует обнаруживаемым целям.

В большинстве систем кластеризация осуществляется с помощью полуэмпирических методов, эффективность которых невысока. Это приводит, с одной стороны, к пропуску части отметок и снижению энергетики при принятии решений о наличии целей, а также к размножению отметок от крупных целей и появлению ложных целей. С другой стороны, при завышенном пороге группирования возможно объединение отражений от разных близкорасположенных целей в один кластер, что приводит к пропуску целей и ухудшению точностных характеристик определения их траекторий. Осуществление эффективной кластеризации отметок позволит в дальнейшем адекватно формировать траектории групп целей, что приведет к энергетическому выигрышу и снижению числа ложных трасс на порядок [2].

В настоящее время растет количество работ, где решается задача кластеризации радиолокационных импульсов с помощью различных методов. Например, распознавание динамических объектов в радиолокационном пространстве с помощью алгоритмов кластеризации k-means и нечеткой кластеризации c-means по навигационным параметрам [3], модель масштабной смеси нормальных распределений для классификации и кластеризации радиолокационных излучателей [4]*,* метод автоматической классификации с использованием сети вычисления p-значений для проверки гипотез о типах излучателей, где алгоритм кластеризации основывается на методе обучаемой кластеризации векторов [5]*,* алгоритмы кластеризации с радарным сканирование на основе плотности [6],применяют и аппарат нейронных сетей, а именно глубокие рекуррентные нейронные сети (РНС) для классификации и грубой кластеризации различных групп импульсов иерархически в отношении их последовательных структур [7].

Цель работы – разработка и исследование алгоритма кластеризации последовательности радиолокационных импульсов в пассивных радиолокационных комплексах.

В данной работе рассматривается возможность повышения эффективности обработки информации в системах пассивной радиолокации за счет отождествления сигналов с целями путем кластеризации принятых радиоимпульсов алгоритмом на базе DBSCAN [8-9]. В свою очередь данный алгоритм не требует определения числа кластеров заранее и учитывает выбросы (шум) в радиолокационных данных. Важно отметить, в данной статье задача кластеризации ставится не для одиночных импульсов, а для обработки последовательности импульсов, называемых паттерном. Такая постановка усложняет классическую задачу кластеризации, так как необходимо выделить в один кластер не простые единичные импульсы, а последовательность (паттерн или сложный сигнал) от наблюдаемой цели. В связи с этим в дополнение к алгоритму DBSCAN были реализованы эвристические правила отсеивания ложных сигналов.

# Постановка задачи

Пусть пассивная система радиолокации принимает простые сигналы (импульсы или отметки) в дискретном периоде наблюдений равном N тактов. Вектор состояния описывающий простой сигнал в k-ый момент времени записывается в следующем виде:

 ,

где - время прихода импульса,  - несущая частота импульса,  - длительность импульса,  - период между k-ым и (k-1)-ым импульсами.

При этом в наблюдаемой реализации импульсов встречаются следующие сигнальные последовательности (паттерны или сложные сигналы), описывающиеся следующими матрицами состояния:

, ,

где число столбцов в обеих матрицах показывает размер этих сигнальных последовательностей.

Остальные сигналы в реализации будут считаться помехами, а их вектор состояния будет записываться как у одиночного импульса:

.

Вся принимаемая реализация импульсов на периоде наблюдения равного N записывается следующим уравнением:

,

где  - матрица состояния всей реализации импульсов размерностью ,  - матрица аддитивных белых гауссовский шумов с нулевым средним и конечной дисперсией размерностью .

Ставится задача на фоне принимаемой реализации импульсов  кластеризировать, то есть выделить в отдельные группы (кластеры), все встречающиеся паттерны типа A и B.

В качестве критерия эффективности решения будем использовать следующие метрики:

1) Точность распознавания паттернов (в %):

,

где  – число истинных кластеров, содержащих искомые паттерны, - число истинных паттернов в кластере,  - оценка числа истинных паттернов в кластере.

2) Ложное обнаружение паттернов (в %):



где  - оценка числа истинных кластеров, содержащих искомые паттерны,  - оценка числа истинных паттернов в кластере,  - оценка числа неверных паттернов в кластере.

3) Пропуск (в %):

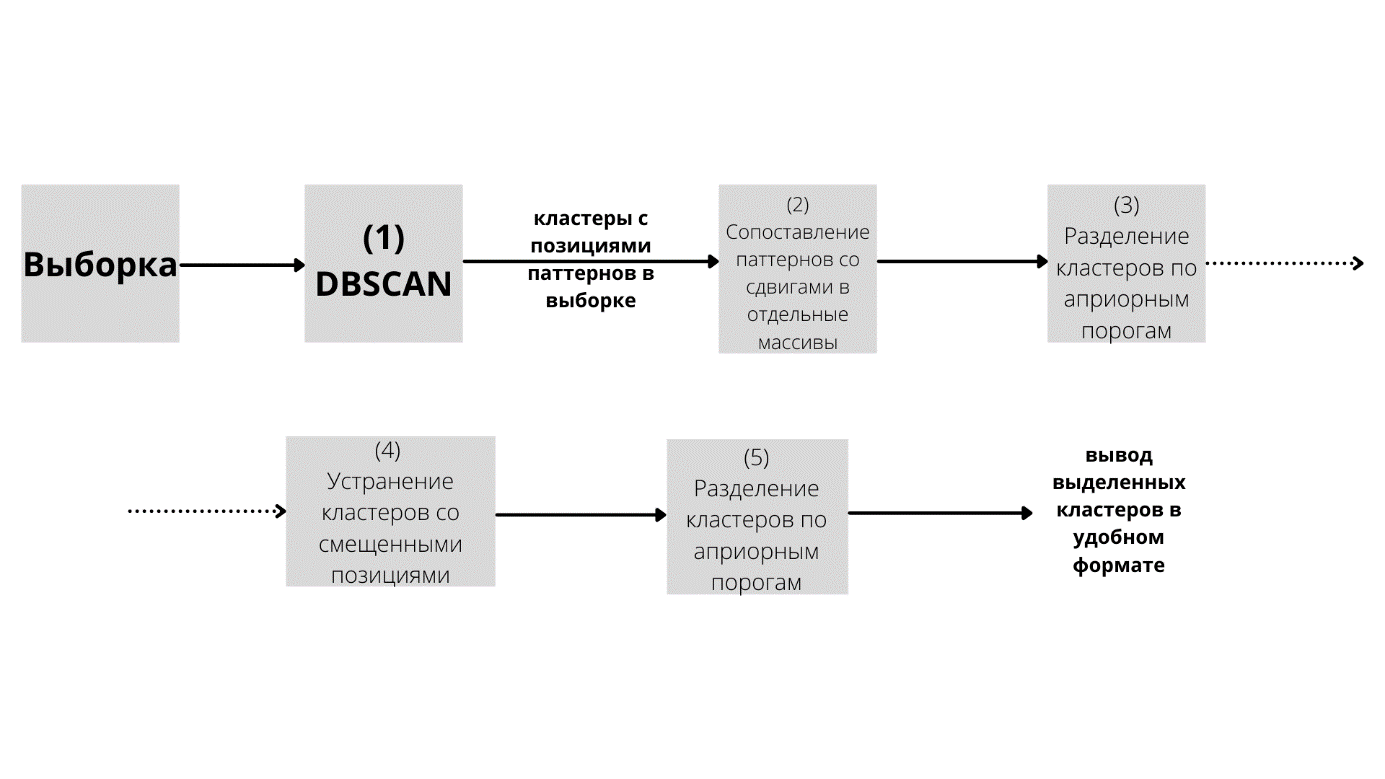


где skip принимает дискретные значения 0 и 1 (1 – пропущен кластер с искомыми паттернами, 0 – кластер с искомыми паттернами распознан),  – число истинных кластеров, содержащих искомые паттерны.

4) Оценка числа истинных кластеров .

# Разработка алгоритма

Для решения поставленной задачи был применен разработанный каскад обработки радиолокационных данных на базе алгоритма DBSCAN, представленный на рис.



Для удобства назовем его cascade DBSCAN или сокращенно cDBSCAN.

Блок (1) характеризуется непосредственно алгоритмом кластеризации DBSCAN.

Концепция алгоритма DBSCAN состоит в выделении областей высокой плотности, которые отделены друг от друга областями низкой плотности. И чтобы правильно применить данный алгоритм кластеризации для решения поставленной задачи, требуется подробнее определить входные параметры и характеристики алгоритма. На вход алгоритма поступает матрица близости и два параметра – радиус 𝑒𝑝𝑠𝑖𝑙𝑜𝑛-окрестности и 𝑚𝑖𝑛𝑝𝑡𝑠-минимальное количество соседей. Для определения 𝑒𝑝𝑠𝑖𝑙𝑜𝑛 и 𝑚𝑖𝑛𝑝𝑡𝑠 требуется ввести несколько определений. Пусть задана некоторая симметричная функция расстояния 𝜌(𝑥, 𝑥 ′ ) и константы 𝜀 и 𝑚.

Тогда:

1. Назовем область 𝐸(𝑥), для которой ∀𝑥: 𝜌(𝑥, 𝑥 ′ ) ≤ 𝜀, где 𝜀 – окрестность объекта 𝑥

2. Центральным объектом или ядерным объектом степени 𝑚 называется объект, 𝜀-окрестность которого содержит не менее 𝑚 объектов: |𝐸(𝑥)| ≥ 𝑚

3. Объект 𝑝 непосредственно плотно-достижим из объекта 𝑞, если 𝑝 ∈ 𝐸(𝑞) и 𝑞 – корневой объект

4. Объект p плотно-достижим из объекта 𝑞, если ∃𝑝1, 𝑝2 … 𝑝𝑛, 𝑝1 = 𝑞, 𝑝𝑛 = 𝑝, такие что ∀𝑖 ∈ 1 … 𝑛 − 1: 𝑝𝑖 + 1 непосредственно плотно-достижим из 𝑝𝑖

Следуя определению плотной области, точка может быть классифицирована как основная точка если |𝐸(𝑥)| ≥ 𝑚. Центральные точки, как следует из названия, находятся внутри кластера. Пограничная точка (достижимая по плотности точка) имеет меньше, чем m в своей 𝐸(𝑥) области, но лежит в окрестности другой центральной точки. Шум (выпадающая точка) – это любая точка данных, которая не является ни основной, ни пограничной. Достижимость не является симметричным отношением, поскольку, по определению, никакая точка не может быть достигнута из неосновной точки, независимо от расстояния (так что неосновная точка может быть достижимой, но ничто не может быть достигнуто из неё). Поэтому дальнейшее понятие связности необходимо для формального определения области кластеров, найденных алгоритмом DBSCAN. Две точки 𝑝 и 𝑞 связаны по плотности, если имеется точка 𝑜, такая что и 𝑝, и 𝑞 достижимы из 𝑜. Связность по плотности является симметричным отношением. Тогда кластер удовлетворяет двум свойствам:

1. Все точки в кластере попарно связны по плотности.

2. Если точка достижима по плотности из какой-то точки кластера, она также принадлежит кластеру

DBSCAN отлично работает на плотных, хорошо отделенных друг от друга кластерах. При этом форма кластера совершенно не важна. Алгоритм отлично обнаруживает кластеры малой размерности. Успешно применяется для большого датасета , причем сложность элементов датасета значения не имеет. Количество элементов в кластере может варьироваться, количество выбросов значения не имеет, если они рассеяны по большому объему.

Перед предъявлением данных алгоритму кластеризации требуется их нормализовать, либо стандартизировать. В данной работе была выбрана Z-стандартизация параметров матрицы наблюдения. Где Z-стандартизация – это такое преобразование данных, которое позволяет перевести шкалу на Z шкалу, где среднее значение будет равняться нулю, а стандартное отклонение равняется единице.

Параметр времени прихода импульса не используется в алгоритме кластеризации, т.к. он является не информативным. Поэтому входными параметрами для алгоритма кластеризации являются следующие три параметра: несущая частота импульса, период между импульсами и длительность импульса.

Перед нами ставилась задача находить паттерны из выборки импульсов. Поэтому требуется так преобразовать входные данные, чтобы алгоритм кластеризации искал именно набор импульсов, сгруппированный в паттерн. Для этого матрицу наблюдений требуется видоизменить, расширив её до размерности . В нашем случае минимальный размер паттерна равен 3 импульсам. Назовем эту операцию «окном смещения».

Изображение выглядит как текст, часы

Автоматически созданное описание

Рис. Операция окно смещения

Для работы алгоритма кластеризации DBSCAN требуется задать следующие параметры: число минимальных соседей и радиус поиска соседей .

Число минимальных соседей выбирается эвристически и должно быть на одно значение больше числа признаков в обрабатываемых данных. В нашем случае мы обрабатываем три параметра, поэтому число соседей берем равным 4. выбирается исходя из отсортированной по возрастанию матрицы попарных расстояний с евклидовой метрикой.

Подробнее представим выбор -окрестности на рис.

Изображение выглядит как стол

Автоматически созданное описание

Рис. 5‑6 Выбор -окрестности.

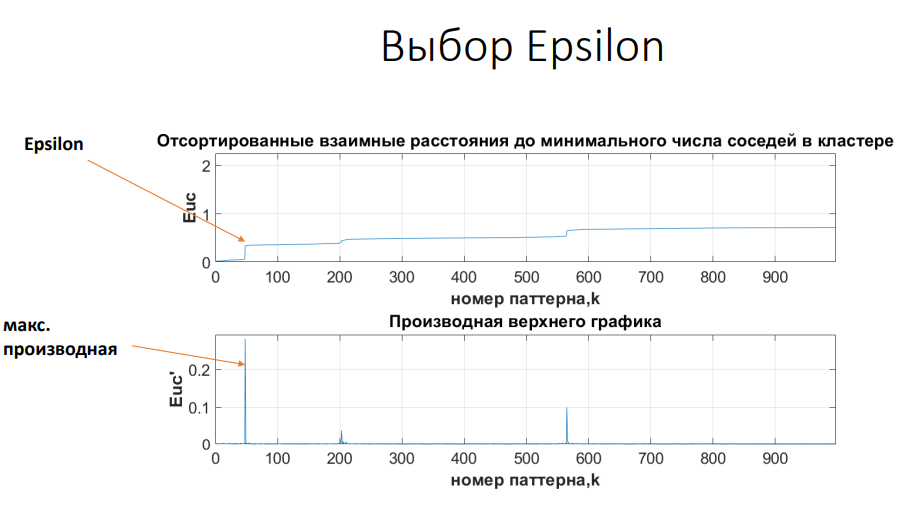


Рис. 5‑7 Выбор -окрестности в увеличенном масштабе.

В итоге после подбора всех необходимых параметров алгоритма DBSCAN на его выходе получаем оцененные кластеры с позициями выявленных в них паттернов из подаваемой выборки. Из-за поиска паттернов различной длины, а именно паттерна длиной в 3 импульса и паттерна в 7 импульсов, и применяемого метода «окна» при формировании нужного формата данных для алгоритма, на его выходе можно получить побочные смещения позиций паттерна большей размерности. То есть алгоритм DBSCAN не может однозначно оценить кластер с паттернами размерностью 7 импульсов. Для чего необходимо дополнить полученный алгоритм рядом эвристических правил, указанных на рисунке со структурной схемой обработки радиолокационной информации. Это реализуется в блоках (2) и (5). Таким образом реализуется новый алгоритм кластеризации cDBSCAN.

В качестве априорных порогов в блоках (3) и (5), применяемых для дополнительного разделения кластеров и фильтрации ложных паттернов применяется пороги по частоте равный 240 МГц и 200 нс по длительности импульсов. Данные значения взяты после анализа реальных измерений с радиолокационного пассивного комплекса.

# Моделирование

Для решения задачи и с целью качественной оценки эффективности алгоритма кластеризации сDBSCAN на данных, которые представляют собой элементарные радиоимпульсы, требуется описать параметры имитационной модели сигнала.

* 1. ***Исходные данные моделирования***

Для того, чтобы описать, что из себя представляют элементарные импульсы, был проведен анализ реальных записей обнаруживаемых элементарных импульсов с пассивного радиолокационного комплекса. Одиночные импульсы имеют следующие параметры: время прихода импульса , длительность импульса , период импульса , это разность времен между текущим импульсом и предыдущим импульсом, несущая частота . Анализируя реальные записи, было выяснено, что импульсы с некоторыми параметрами встречаются чаще всего. А именно: минимальное значение периода , наиболее часто встречающиеся несущие частоты сигналов представляют собой следующий дискретный набор , наиболее часто встречающиеся длительности импульса представляют собой дискретный набор .Параметры этих импульсов взяты за основу сигналов в имитационной модели.

Реальные записи всегда содержат шумы наблюдения, поэтому это следует учесть и для имитационной модели. В итоге подобрались следующие шумы наблюдения: для времени прихода импульса , по несущей частоте , для длительности импульса .

Кроме шума наблюдения существуют также отклонения от мгновенных значений. Положим следующее: девиация частоты в долях от несущей принимает равновероятные значения в следующем диапазоне целых чисел , девиация длительности импульса в долях **.**

* 1. ***Формирование выборки***

Задача следующая. Требуется сформировать выборку импульсов, в которой будут присутствовать повторяющиеся несколько раз сигналы с одинаковыми параметрами – паттерны. Именно эти паттерны должен выделить алгоритм сDBSCAN при кластеризации.

Формирование выборки импульсов происходит следующим образом. Задаются параметры импульса:

1. Формируется случайным образом период между импульсами:

|  |  |
| --- | --- |
|  |  |

Где  *–* непрерывное равномерное распределение целых чисел,  – нормальный закон распределения с нулевым средним значением и дисперсией .

1. Формируется несущая частота импульса:

|  |  |
| --- | --- |
|  |  |

1. Формируется случайным образом длительность импульса:

|  |  |
| --- | --- |
|  |  |

1. Далее задается длина паттерна
2. Формируется выборка импульсов из элементов, где случайным образом распределены паттерны. К параметрам импульсов добавляется шум наблюдения, распределенный по нормальному закону.

В итоге модель наблюдений описывается следующим образом:

|  |  |
| --- | --- |
|  |  |

где - матрица шумов наблюдений размерностью - матрица состояния размерностью - матрица наблюдений размерностью

* 1. ***Результаты эксперимента***

Требуется сформировать выборку импульсов и добавить в эту выборку два паттерна случайным образом. Размер 1-го паттерна возьмём равным трем импульсам, причём было принято, что три импульса — это минимальный размер паттерна. Размер 2-го паттерна равен семи импульсам.

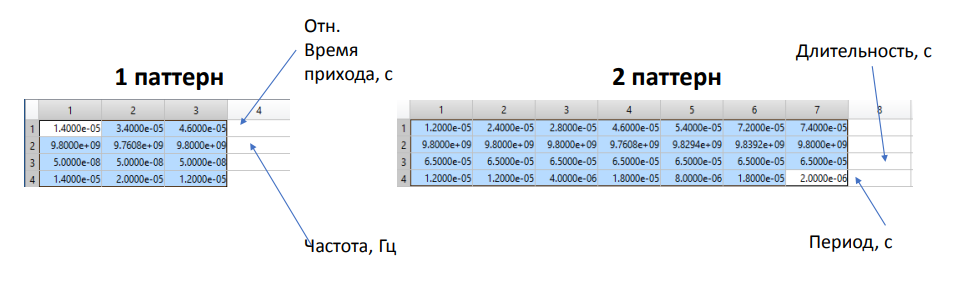


Рис. 5‑1 Пример паттернов.

Предполагается, что алгоритмом кластеризации сDBSCAN можно будет выделить добавленные паттерны в отдельные кластеры, тем самым сформировав два кластера со схожими паттернами. Остальные импульсы он будет считать помехами и шумами (выбросами).

Подробнее проанализируем смоделированную выборку. Для этого построим вероятностные распределение параметров матрицы наблюдений (рис. 5-2) и представление импульсов в выборке (рис. 5-3, 5-4).

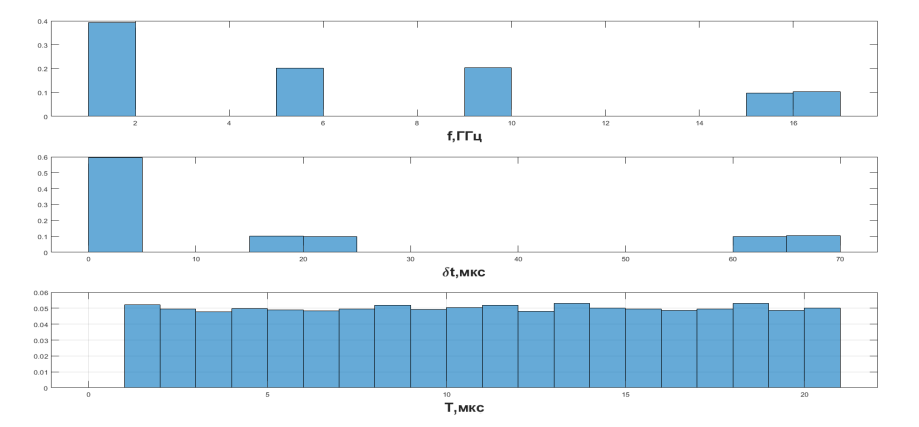


Рис. 5‑2 Вероятностные распределения параметров матрицы наблюдений (без времен прихода)



Рис. 5‑3 Двумерное представление импульсов

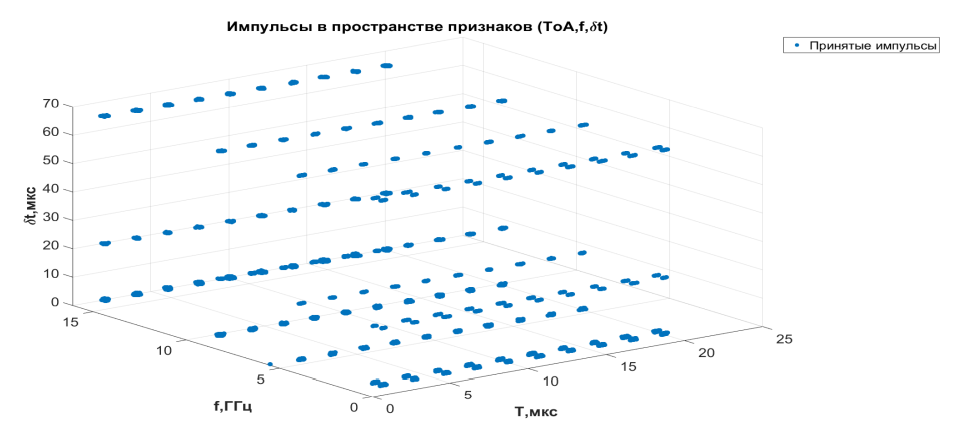


Рис. 5‑4 Трехмерное представление импульсов с осью значений периодов

Далее представлены результаты алгоритма кластеризации cDBSCAN для данной выборки, которая содержит два паттерна разной длины.

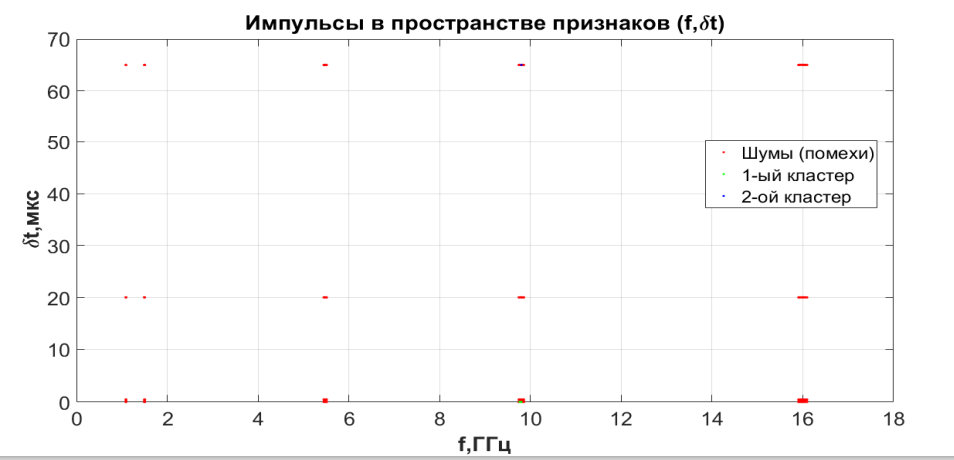


Рис. 5‑8 Импульсы в пространстве признаков частоты и длительности импульса.

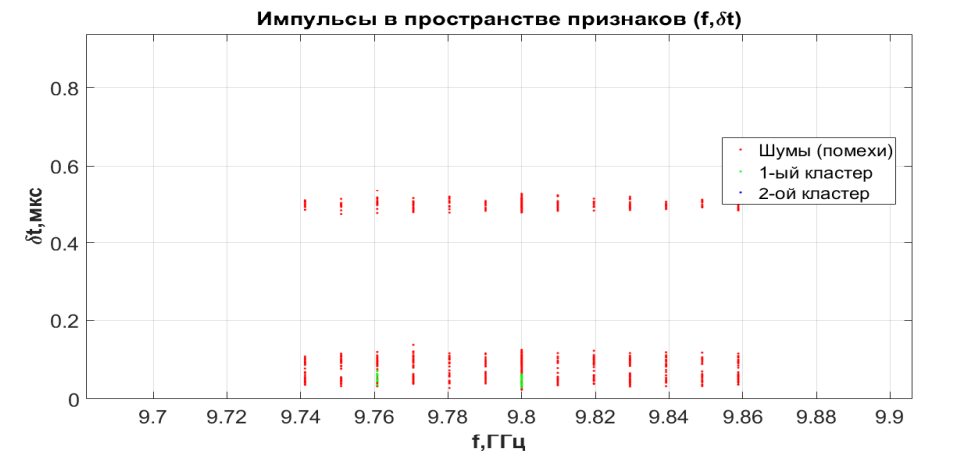


Рис. 5‑9 Импульсы в пространстве признаков частоты и длительности импульса в увеличенном масштабе первый кластер.

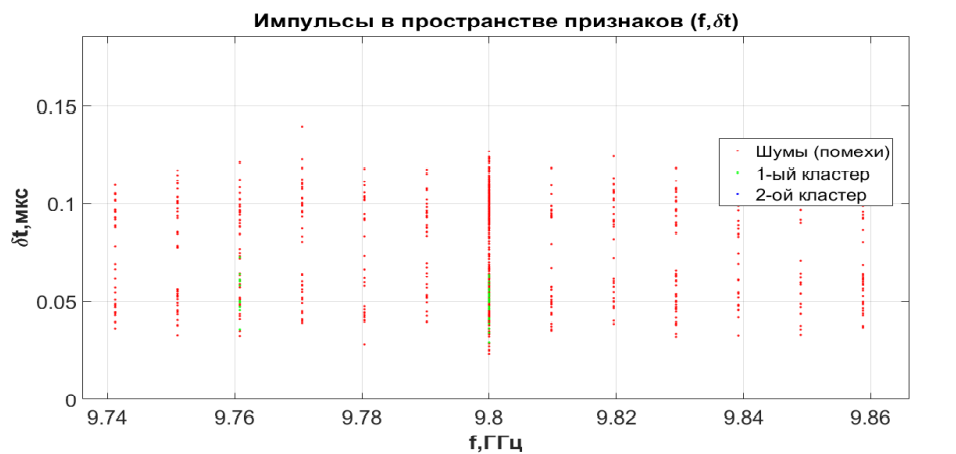


Рис. 5‑10 Импульсы в пространстве признаков частоты и длительности импульса в увеличенном масштабе первый кластер.

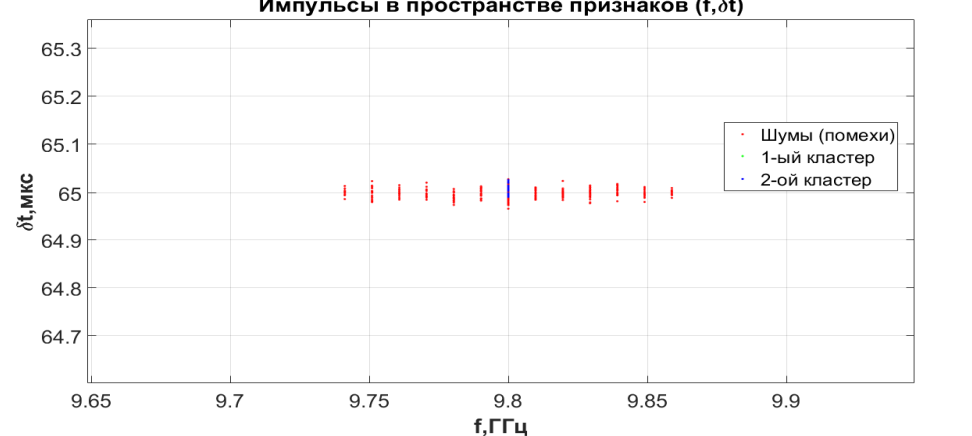


Рис. 5‑11 Импульсы в пространстве признаков частоты и длительности импульса в увеличенном масштабе второй кластер.

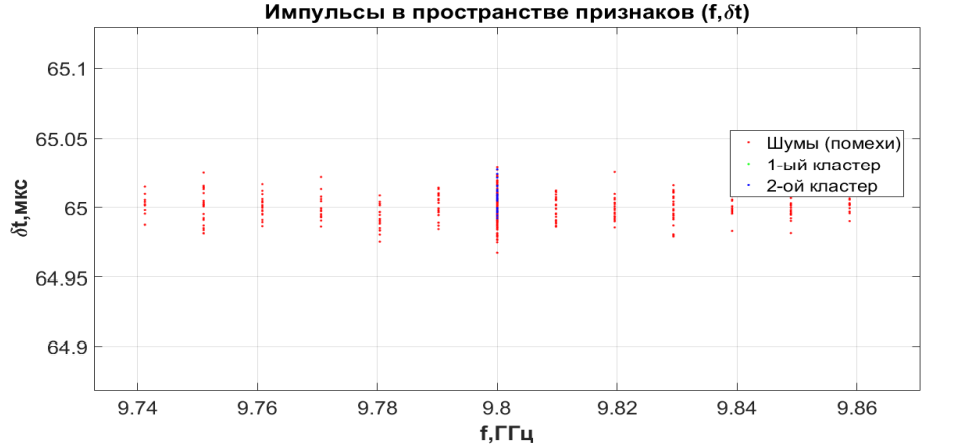


Рис. 5‑12 Импульсы в пространстве признаков частоты и длительности импульса в увеличенном масштабе второй кластер.

Алгоритм сDBSCAN явно выделил два кластера. Теперь требуется проверить точность оценки качества кластеризации на множестве реализаций.

В итоге получили следующие значения заявленных метрик качества.

Результат кластеризации на 100 случаях, средняя точность определения паттернов составила - 98.2632%

Средний процент ложных обнаружений паттернов составил - 2.0087%

Средний процент пропущенных кластеров - 1%

Среднее число выявляемых кластеров составило - 1.99

# Заключение

В работе были рассмотрены современные методы и алгоритмы решения задачи кластеризации радиолокационной информации. Было показано, что применение алгоритмов статистики и машинного обучения к системам обработки радиолокационной информации считается актуальным и многообещающим направлением. В данной работе было предложено применить алгоритмы кластеризации сигнальных паттернов на этапе обнаружения отметок от целей. Решение такой задачи позволит в будущем подбирать более эффективные алгоритмы вторичной (траекторной) обработки целей.

Для решения задачи кластеризации радиолокационных паттернов был разработан алгоритм кластеризации cDBSCAN – плотностной алгоритм пространственной кластеризации с присутствием шума, с применением каскада эвристических правил.

Также в работе была смоделирована имитационная модель элементарных радиоимпульсов, основанная на анализе реальных записей обнаруживаемых элементарных импульсов с пассивного радиолокационного комплекса. Применение алгоритма кластеризации cDBSCAN к данным, сформированным на модели, показало, что алгоритм верно определяет паттерны в выборке с высокой точностью, более 98%.

Дальнейшая работа в данном направлении – это кластеризовать реальные записи с пассивного радиолокационного комплекса с целью выделения повторяющихся паттернов и дальнейшего анализа записей с применением cDBSCAN.

## Литература

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мегиалкин Л Д. Прикладная статистика: Классификация и снижение размерностей / Справочное издание под ред. С.А. Айвазяна. М.: Финансы и статистика, 1989. 607
2. Татузов А.Л.. Нейронные сети в задачах радиолокации. Кн. 28. – М.: Радиотехника, 2009. – 432 с.: ил. (Научная серия «Нейрокомпьютеры и их применение).
3. Князев Н.Л., Денисова Л.А., Методы распознавания динамических объектов в радиолокационном пространстве. УДК 621.396.96:004.021. Омский государственный технический университет, г. Омск, Россия.
4. G. Revillon, A. Mohammad-Djafari and C. Enderli, "Radar emitters classification and clustering with a scale mixture of normal distributions," 2018 IEEE Radar Conference (RadarConf18), 2018, pp. 1371-1376, doi: 10.1109/RADAR.2018.8378764.
5. Kim, L.S., Kil, R.M., Jo, C.H. (2015). Radar Pattern Classification Based on Class Probability Output Networks. In: Arik, S., Huang, T., Lai, W., Liu, Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science(), vol 9489. Springer, Cham.
6. Ramirez-Figueroa, John A. et al. “A new principal component analysis by particle swarm optimization with an environmental application for data science.” Stochastic Environmental Research and Risk Assessment (2021): 1-16.
7. Z. -M. Liu, "Recognition of Multifunction Radars Via Hierarchically Mining and Exploiting Pulse Group Patterns," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 6, pp. 4659-4672, Dec. 2020, doi: 10.1109/TAES.2020.2999163.
8. Jang, Jennifer and Heinrich Jiang. “DBSCAN++: Towards fast and scalable density clustering.” ArXiv abs/1810.13105 (2019): n. pag.
9. Ester, Martin et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” KDD (1996).

## References

1. Aivazyan S.A., Buchbacker V.M., Yenyukov I.S., Megyalkin L D. Applied Statistics: Classification and Dimensional Reduction / Reference Edition under Rev. S.A. Aivazyan. M.: Finance and statistics,1989. 607
2. Tatozov A.L.. Neural networks in radar tasks. Kn. 28. - M.: Radio Engineering, 2009. - 432 pp.: il. (Scientific series Neurocomputers and their applications).
3. Knyazev N.L., Denisova L.A., Methods of recognition of dynamic objects in radar space. UDC 621.396.96:004.021. Omsk State Technical University, t. Omsk, Russia.
4. G. Revillon, A. Mohammad-Djafari and C. Enderli, "Radar emitters classification and clustering with a scale mixture of normal distributions," 2018 IEEE Radar Conference (RadarConf18), 2018, pp. 1371-1376, doi: 10.1109/RADAR.2018.8378764.
5. Kim, L.S., Kil, R.M., Jo, C.H. (2015). Radar Pattern Classification Based on Class Probability Output Networks. In: Arik, S., Huang, T., Lai, W., Liu, Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science(), vol 9489. Springer, Cham.
6. Ramirez-Figueroa, John A. et al. “A new principal component analysis by particle swarm optimization with an environmental application for data science.” Stochastic Environmental Research and Risk Assessment (2021): 1-16.
7. Z. -M. Liu, "Recognition of Multifunction Radars Via Hierarchically Mining and Exploiting Pulse Group Patterns," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 6, pp. 4659-4672, Dec. 2020, doi: 10.1109/TAES.2020.2999163.
8. Jang, Jennifer and Heinrich Jiang. “DBSCAN++: Towards fast and scalable density clustering.” ArXiv abs/1810.13105 (2019): n. pag.
9. Ester, Martin et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” KDD (1996).