

ПРИМЕНЕНИЕ ДВУХЭТАПНОГО МЕТОДА КЛАСТЕРИЗАЦИИ НА ОСНОВЕ САМООРГАНИЗУЮЩЕЙСЯ КАРТЫ КОХОНЕНА ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В СИНТЕТИЧЕСКИХ НАБОРАХ ДАННЫХ¹

В статье представлен двухэтапный метод кластеризации, основанный на применении самоорганизующейся карты Кохонена с целью выявления аномалий в синтетических наборах данных. Этот подход позволяет более точно идентифицировать аномальные объекты по сравнению с одноэтапными методами кластеризации. Представлены результаты экспериментов, проведенных на синтетических наборах данных, которые подтверждают эффективность предложенного метода.

На первом этапе исследования формируется синтетический набор данных, содержащий два или три признака. Далее данные подвергаются обработке с использованием самоорганизующейся карты Кохонена, что позволяет выделить основные кластеры и определить границы между ними.

На втором этапе обнаружения аномалий применяются алгоритмы кластеризации, основанные на анализе пространства признаков и сравнении расстояний между объектами. Аномальные объекты, находящиеся в значительном удалении от основных кластеров, классифицируются как аномалии. Проведенный анализ работы алгоритмов кластеризации на данном этапе позволил выявить, что алгоритмы DBSCAN и Isolation Forest проявили себя наиболее эффективными в обнаружении выбросов по сравнению с алгоритмом OPTICS.

Ключевые слова: обнаружение аномалий, кластеризация данных, самоорганизующаяся карта Кохонена, синтетический набор данных

¹ Исследование поддержано грантом Российского научного фонда (проект No 22-71-10095).

APPLICATION OF A TWO-STAGE CLUSTERING METHOD BASED ON SELF-ORGANISING KOHONEN MAP FOR ANOMALY DETECTION IN SYNTHETIC DATASETS

The paper presents a two-stage clustering method based on the application of a self-organizing Kohonen map to identify anomalies in synthetic datasets. This approach allows for more accurate identification of anomalous objects compared to one-stage clustering methods. Experimental results on synthetic datasets are presented, confirming the effectiveness of the proposed method.

In the first stage of the study, a synthetic dataset containing two or three features is generated. The data is then processed using a self-organizing Kohonen map to identify the main clusters and determine the boundaries between them.

In the second stage of anomaly detection, clustering algorithms based on feature space analysis and distance comparison between objects are applied. Anomalous objects that are far from the main clusters are classified as anomalies. Analysis of the performance of the clustering algorithms at this stage showed that the DBSCAN and Isolation Forest algorithms were the most effective in detecting outliers compared to the OPTICS algorithm.

Keywords: anomaly detection, data clustering, a self-organizing Kohonen map, synthetic dataset

Введение

Техническое развитие промышленных средств автоматизации и увеличение уровня интеграции промышленных и корпоративных сетей приводит к увеличению рисков проведения успешных кибератак. Реализация таких кибератак может подразумевать получение доступа к управлению важными промышленными установками, что влечёт за собой риск остановки производства или создания аварийной ситуации.

Обеспечение информационной безопасности автоматизированных систем управления технологическими процессами (АСУ ТП) требует своевременного обнаружения кибератак как известного, так и неизвестного типа. Эти кибератаки можно рассматривать как аномалии в динамических процессах, регистрируемых при работе АСУ ТП [1].

Для решения задачи поиска кластеров произвольной формы в условиях сильной зашумленности в данных обычно применяют простые и быстрые алгоритмы, основанные

на методах обучения без учителя, такие, например, как метод К-средних. Однако они неэффективны в условиях сильной зашумленности и требуют дополнительной информации, включая количество кластеров [2].

Более сложные алгоритмы иерархической кластеризации, такие, например, как OPTICS (Ordering Points To Identify the Clustering Structure), определяют кластеры произвольной формы, сохраняя информацию о всей структуре кластеров, однако требуют значительных вычислительных ресурсов при работе с высокоразмерными данными из-за сложности алгоритма [3].

Среди классических алгоритмов кластеризации наиболее перспективным алгоритмом, способным выделять кластеры произвольной формы, является DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – алгоритм пространственной кластеризации на основе плотности данных. Однако следует отметить, что при увеличении количества признаков эффективность DBSCAN

может снижаться. При этом по сравнению с другими алгоритмами кластеризации этот алгоритм обладает сравнительно небольшой вычислительной сложностью [4].

При анализе данных, характеризующихся высокой размерностью, возникает необходимость в таком алгоритме кластеризации, как Isolation Forest. Это алгоритм поиска аномалий, который может использоваться и для кластеризации. В контексте высокой размерности данных Isolation Forest проявляет себя эффективно, поскольку он строит деревья решений быстро и анализирует отдельные объекты независимо от остальных, что позволяет ему хорошо справляться с данными высокой размерности, где другие методы могут столкнуться с проблемами из-за т.н. «проклятия размерности» [5].

Тем не менее, приведенные методы одноэтапного обнаружения аномалий, основанные на использовании статических пороговых значений и подверженные воздействию шума, могут оказаться ограниченными в своей гибкости, устойчивости и способности выявлять сложные аномалии в данных. С целью совершенствования этих методов, расширения их функциональных возможностей и адаптации под задачи эффективного обнаружения и предотвращения кибератак предлагается применять двухэтапный метод кластеризации, основанный на применении самоорганизующейся карты Кохонена.

Для получения точных и устойчивых решений при решении сложных задач в области обнаружения аномалий с применением двухэтапных методов кластеризации включаются такие этапы, как предварительная обработка данных, настройка параметров используемых алгоритмов, комбинирование различных моделей и учет контекста [6].

Обнаруженные аномалии в данных могут свидетельствовать о потенциальных угрозах и атаках и проявляться в различных формах, например, в виде необычных паттернов доступа к данным, отклонений от типичного поведения пользователей или изменения в структуре данных. Процесс кластеризации позволяет группи-

ровать данные в соответствии с их сходством по характеристикам, что в свою очередь способствует выявлению аномалий, отличающихся от общей тенденции данных [7].

Целью применения двухэтапного метода кластеризации, основанного на самоорганизующейся карте Кохонена, является формирование кластеров, наблюдаемых данных на первом этапе и выявление возможных угроз безопасности информации на втором этапе. Исследование проводилось с использованием онлайн-платформы Colaboratory от Google на языке программирования Python.

На первом этапе метода происходит построение самоорганизующейся карты Кохонена, которая представляет собой инструмент для визуализации многомерного пространства параметров. В процессе обучения карта формирует топологическую структуру, отражающую взаимосвязи между различными параметрами данных [8].

На втором этапе анализа результатов кластеризации, полученных от самоорганизующейся карты Кохонена, применяются алгоритмы кластерного анализа, такие как DBSCAN, OPTICS и Isolation Forest. В результате обработки данных этими алгоритмами отмечаются однозначно выделенные аномалии, что позволяет эффективно обнаруживать аномальные объекты в исследуемом наборе данных.

Описание набора данных

В работе проведен анализ синтетических двумерных и трёхмерных наборов данных с бинарной классификацией. Предполагалось, что в данных присутствует заданное количество выбросов, связанное с параметром «contamination» (который составляет 10% от общего объема данных).

Рассмотрим синтетический массив двумерных данных, соответствующий наличию двух признаков в потоке данных. В массиве содержится 110 точек, где 100 точек классифицируются как «нормальные» (inliers), а 10 точек представляют собой данные, отличающиеся от общей структуры, и считаются «выбросами» (outliers, рис. 1).

Первые несколько inliers:

```
[[ 0.1, 0.9],  
 [-0.7, 0.2],  
 [ 0.4, -0.8],  
 ... ]
```

Первые несколько outliers:

```
[[ -5.2, 3.1],  
 [ 0.5, -7.3],  
 [-8.0, 0.0],  
 ... ]
```

Рис. 1. Представление массива синтетического набора данных

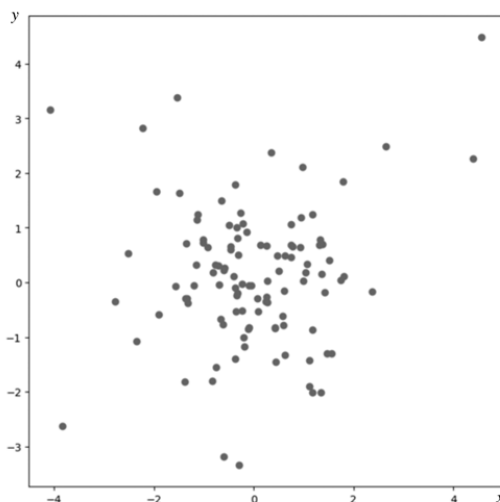


Рис. 2. Диаграмма рассеяния синтетического набора данных

В ходе выполнения алгоритма происходит генерация случайных выбросов, которые могут изменяться при каждом запуске программы. Затем формируется диаграмма рассеяния, которая наглядно отображает распределение точек данных по координатам x и y на декартовой плоскости (рис. 2, 8).

Необходимо отметить, что в процессе синтеза данных не задается фиксированное значение радиуса или расстояния, которое определяло бы данные как аномалии. Процесс обнаружения аномалий изменяется динамически в зависимости от структуры данных. При моделировании данных используется распределение внутри двух концентриче-

ских окружностей. Такой способ моделирования данных основан на принципе сферической симметрии и позволяет эффективно выявлять аномалии в данных, структура которых соответствует структуре окружающего пространства данных. Равномерное распределение данных по всей поверхности сферы обеспечивает их симметрию относительно центра, что на фоне симметрии упрощает выделение и анализ аномалий [9].

Для визуального представления данных в виде «выбросов» определено пороговое значение ошибки квантования (рис. 3, 9). На графике представлены значения ошибок квантования по оси X , и частота их появления по оси

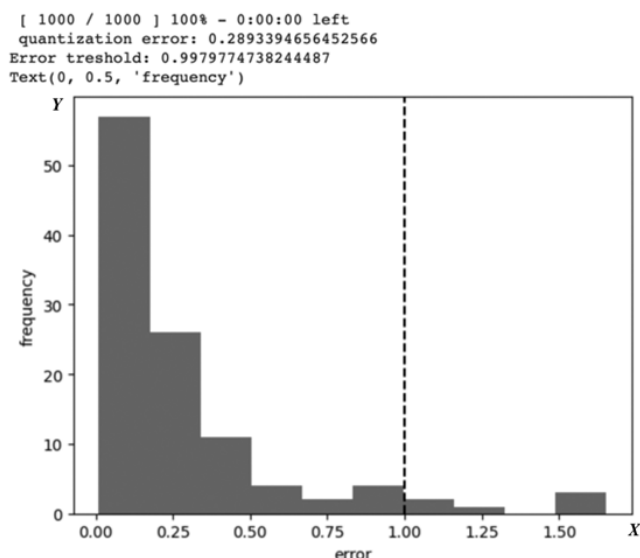


Рис. 3. Гистограмма ошибки квантования

У. Ось X отражает величину ошибок квантования, т.е. меру расстояния между исходными данными и соответствующими векторами после обучения картой Кохонена. Эти ошибки нормализованы, показывая, насколько далеко точка данных находится от ближайшего нейрона в весовом пространстве. Ось Y отображает частоту точек данных с каждым значением ошибки на оси X. Следовательно, полученная гистограмма визуализирует распределение ошибок в наборе данных. Вертикальная черная пунктирная линия на графике обозначает порог ошибки, превышение которого классифицирует данные как выбросы.

Обычно, для определения порогового значения используются различные статистические методы, такие как, например, методы на основе стандартного отклонения или квантилей [10]. В рассматриваемом случае для вычисления порога выявления аномалий использовано среднее значение ошибок квантования `np.mean(quantization_errors)` и стандартное отклонение `np.std(quantization_errors)` этих ошибок. Среднее значение ошибок квантования (`np.mean`) вычисляется с помощью формулы:

$$\text{mean} = \frac{1}{N} \sum_{i=1}^N x_i$$

где N – общее количество значений, а x_i – значения ошибок квантования.

Стандартное отклонение (`np.std`) вычисляется с помощью формулы:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2},$$

где μ – среднее значение вероятностного распределения.

Вычисление порога обнаружения аномалий на основе средних значений и стандартных отклонений ошибок квантования осуществлялось по формуле [11]:

$$T = \mu + k \cdot \sigma,$$

где $\mu = \text{np.mean(quantization_errors)}$ – среднее значение ошибок квантования;

$\sigma = \text{np.std(quantization_errors)}$ – стандартное отклонение ошибок квантования;

k – коэффициент, определяющий порог определения аномалий.

В случае, когда значение ошибки кванто-

вания для определенной точки превышает установленный порог T , данная точка классифицируется как аномальная.

Алгоритм двухэтапной кластеризации, основанный на самоорганизующейся карте Кохонена

На первом этапе обработки синтетических данных применена самоорганизующаяся карта Кохонена, которая представляет собой инструмент первоначальной кластеризации данных с целью выявления групп, обладающих схожей структурой данных, отражающих как нормальное, так и аномальное функционирование системы.

Для реализации процесса кластеризации данных построена карта Кохонена размером 8x8 узлов с установленной топологией размещения нейронов. Перед обучением карты произведено нормирование обучающего датасета, а во время обучения использована линейная инициализация весов нейронов и применен пакетный обучающий алгоритм. В контексте карт Кохонена пакетный алгоритм обучения подразумевает использование всех обучающих данных для обновления весов нейронной сети за одну итерацию. Это позволяет более эффективно использовать данные и сделать процесс обучения более стабильным. По завершению обучения карты формируется демонстрационная унифицированная матрица расстояний - представление самоорганизующейся карты Кохонена, которое визуализирует расстояние между нейронами [12].

Исследование начинается с применения модели ко всему набору синтетических данных. Затем каждый входной образец из данных отображается на карту. Для каждого образца вычисляется евклидово расстояние между исходным значением и его аппроксимацией на карте. Темная окраска на карте соответствует увеличению расстояния между нейронами, что свидетельствует о возможном наличии выбросов в данных (рис. 4, 10).

В ходе анализа данных с использованием карты Кохонена были выявлены точки, которые могут быть интерпретированы как выбросы. Проведенное исследование выявило наличие двух кластеров в обученной карте Кохонена. Среди общего числа выбросов, представленных в датасете, было обнаружено 6 точек из 10 заданных (рис. 5, 11).

Эти значения представлены в классификационном отчете и на матрице ошибок, используемой в качестве инструмента визуализации

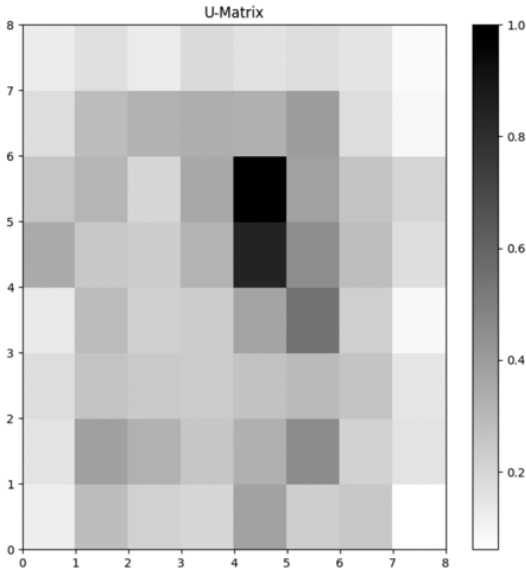


Рис. 4. Унифицированная матрица расстояний (U-matrix) между нейронами

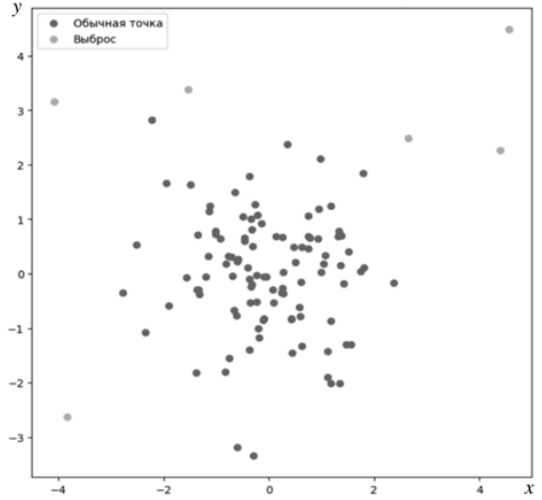
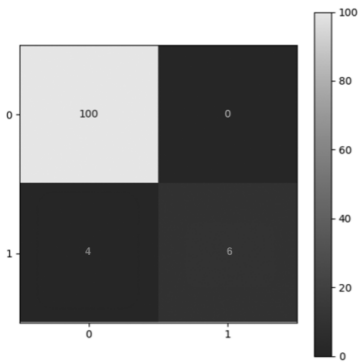


Рис. 5. Диаграмма рассеяния после обработки данных картой Кохонена

зации качества работы классификатора, демонстрирующего, сколько объектов каждого класса было классифицировано правильно или неправильно (рис. 6). Диагональные элементы матрицы отражают количество правильно классифицированных объектов для

каждого класса. Ошибки первого и второго рода вычисляются на основе данных из матрицы ошибок, которая содержит информацию о фактических и предсказанных классах.

На втором этапе обработки данных происходит процесс идентификации кластеров, к которым были присвоены данные, описывающие аномальное поведение наблюдаемых процессов на обученной карте. Для выполнения этого этапа исследования применены алгоритмы кластеризации DBSCAN, OPTICS и Isolation Forest (рис. 7, 12).



Матрица ошибок:
 $\begin{bmatrix} 100 & 0 \\ 4 & 6 \end{bmatrix}$
 Количество False Positive: 0
 Количество False Negative: 4

a)

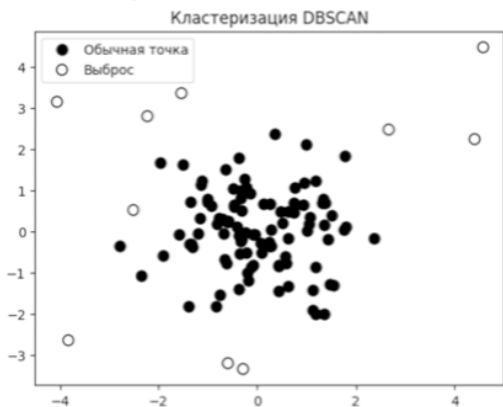
	precision	recall	f1-score	support
0.0	0.96	1.00	0.98	100
1.0	1.00	0.60	0.75	10
accuracy			0.96	110
macro avg	0.98	0.80	0.87	110
weighted avg	0.97	0.96	0.96	110

Precision: 1.0
 Recall: 0.6
 F1 Score: 0.7499999999999999

б)

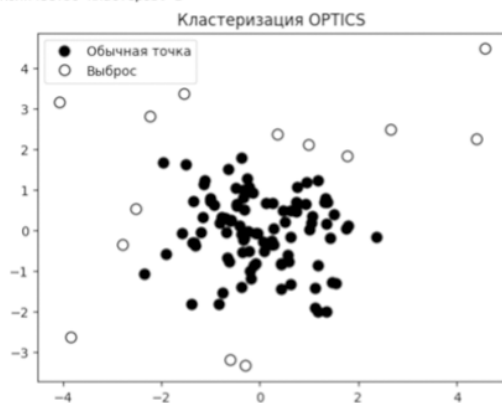
Рис. 6. Метрики точности работы алгоритма: матрица ошибок (а); классификационный отчет (б)

Количество шумовых точек: 10
Количество кластеров: 2



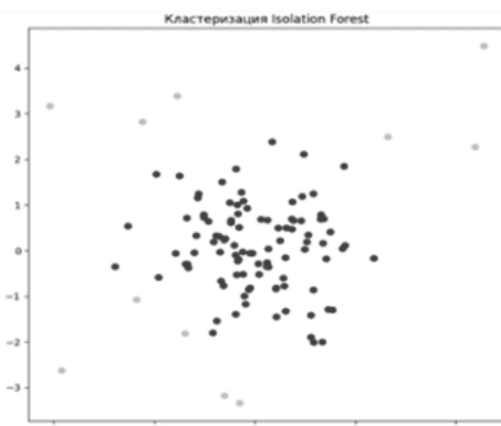
а)

Количество шумовых точек: 14
Количество кластеров: 2



б)

Количество кластеров: 2
Количество шумовых точек: 11



в)

Рис. 7. Диаграммы рассеяния второго этапа кластеризации методами: DBSCAN (а); OPTICS (б) и Isolation Forest (в)

На входе алгоритмов используются векторы весовых коэффициентов, соответствующие каждому из 64 нейронов обученной карты Кохонена.

Рассматривая работу кластеризаторов на втором этапе, можно отметить, что они успешно выявили оставшиеся выбросы. Наиболее результативными при этом оказались методы DBSCAN и Isolation Forest.

Для каждого алгоритма проводилась настройка параметров. Например, настройка работы алгоритма DBSCAN включает в себя два этапа: указание минимального числа соседей, которые ищутся для каждой точки из набора входных данных, и указание радиуса окрестности, в котором ведется соответствующий поиск соседей для каждой точки [2].

Подбор оптимальных параметров был осуществлен с использованием метода пе-

ребора. Установленные значения настроечных параметров, такие как радиус окрестности $\epsilon = 1,0$ и минимальное число соседей $m = 8$, использовались для выполнения алгоритма DBSCAN после предварительного обучения картой Кохонена. Результатом работы алгоритма было разделение исходных данных на кластеры, где оранжевые точки были определены как выбросы, а голубая область представляла собой единственный обнаруженный кластер, соответствующий нормальному состоянию. Следует отметить, что в методах неконтролируемой кластеризации реальные метки классов часто неизвестны, что ограничивает возможности оценки точности кластеризации. В таких случаях можно прибегнуть к косвенным методам оценки, например, анализу свойств сформированных кластеров.

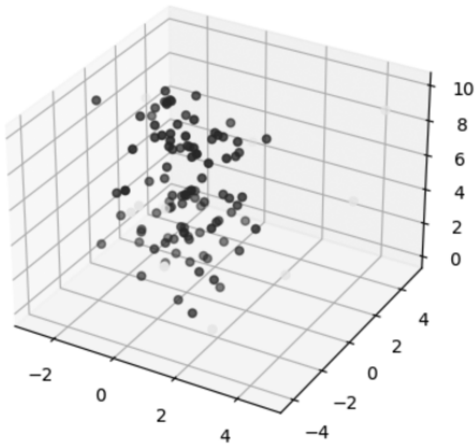


Рис. 8. Диаграмма рассеяния синтетического набора данных с тремя признаками

```
[ 1000 / 1000 ] 100% - 0:00:00 left
quantization error: 1.3188452413536078
Error threshold: 3.120419799230862
Text(0, 0.5, 'frequency')
```

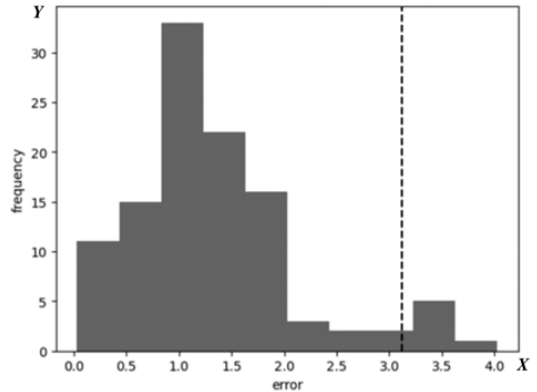


Рис. 9. Гистограмма ошибки квантования для данных с тремя признаками

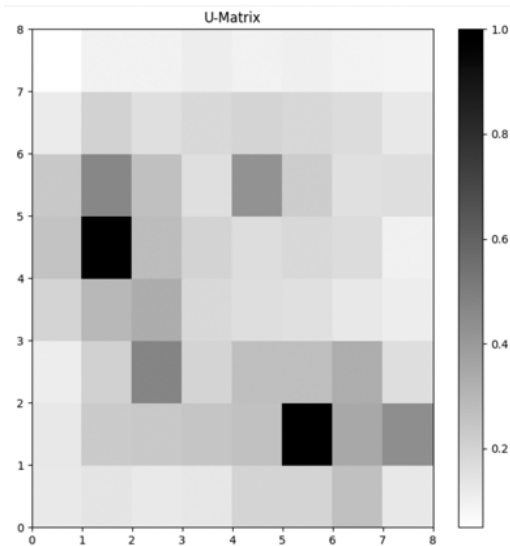


Рис. 10. Унифицированная матрица расстояний (U-matrix) для данных с тремя признаками

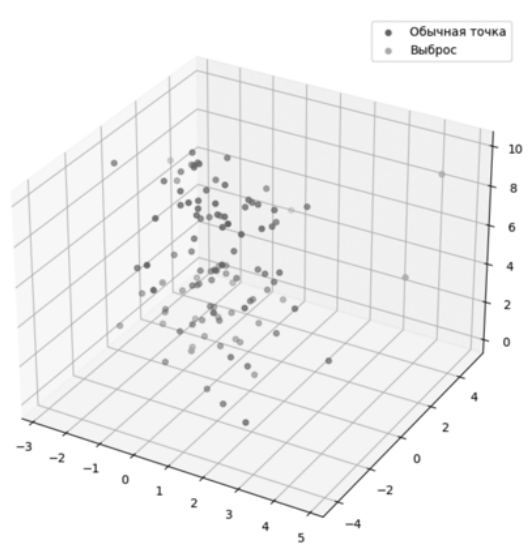


Рис. 11. Диаграмма рассеяния после обработки данных с тремя признаками картой Кохонена

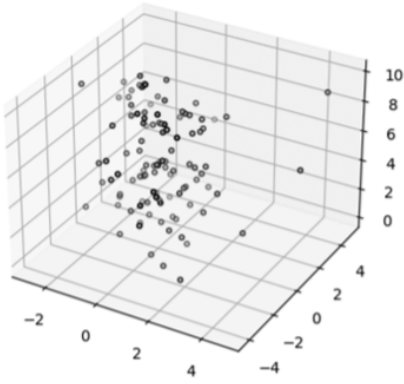
Аналогичный эксперимент проводился и на более сложных данных, с увеличением признаков до трёх (рис. 8). В результате применения метода карты Кохонена на первом этапе были выделены также шесть из десяти исходных точек. При обработке синтетического трехмерного набора данных алгоритмами кластеризации на втором этапе наблюдалось, что алгоритм Isolation Forest продемонстрировал наилучшую производительность, в отличие от алгоритмов DBSCAN и OPTICS, точно выделяя представленные точки и идентифицируя их как аномальные. Но в результате одноэтапной кластеризации алгоритмом Isolation Forest точность снизилась на 7 % (рис. 13).

В условиях постоянного увеличения количества признаков (и, следовательно, размерности) обрабатываемых данных традиционные одноэтапные алгоритмы кластеризации начинают допускать ошибки, классифицируя шумовые или аномальные данные как элементы нормальных кластеров, что в свою очередь ведет к увеличению процента ложных срабатываний и снижению общей точности кластеризации (табл. 1).

В отличие от этого, двухэтапный метод кластеризации предлагает альтернативный подход, который разбивает процесс на два этапа. На первом этапе осуществляется предварительная кластеризация, которая позволяет идентифицировать основные структуры

Количество шумовых точек: 78
Количество кластеров: 4

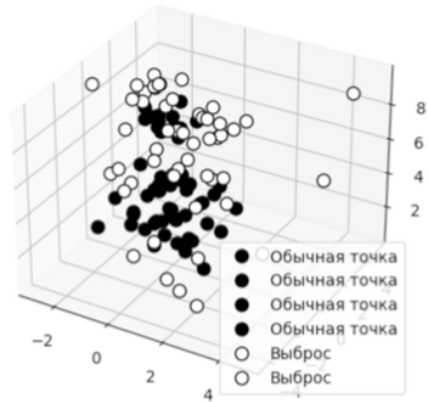
Кластеризация DBSCAN



а)

Количество шумовых точек: 53
Количество кластеров: 3

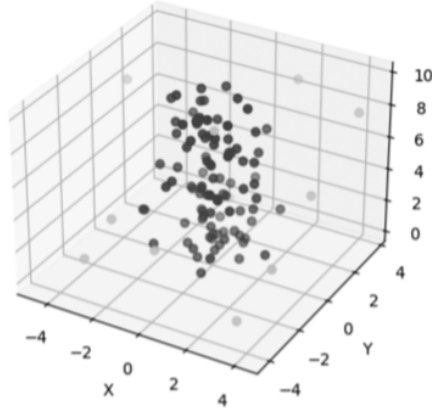
Кластеризация OPTICS



б)

Количество шумовых точек: 11
Количество кластеров: 2

Кластеризация Isolation Forest



в)

Рис. 12. Диаграммы рассеяния второго этапа кластеризации методами: DBSCAN (а); OPTICS (б) и Isolation Forest (в)

Таблица 1.

Точность кластеризации различными алгоритмами

Алгоритм	Количество признаков данных	Вид кластерной обработки	Точность
DBSCAN	2	двухэтапный	90%
OPTICS	2	двухэтапный	86%
Isolation Forest	2	двухэтапный	97%
DBSCAN	3	двухэтапный	75%
OPTICS	3	двухэтапный	72%
Isolation Forest	3	двухэтапный	97%
DBSCAN	3	одноэтапный	72%
OPTICS	3	одноэтапный	69%
Isolation Forest	3	одноэтапный	90%

Количество шумовых точек: 20
Количество кластеров: 2

Кластеризация Isolation Forest

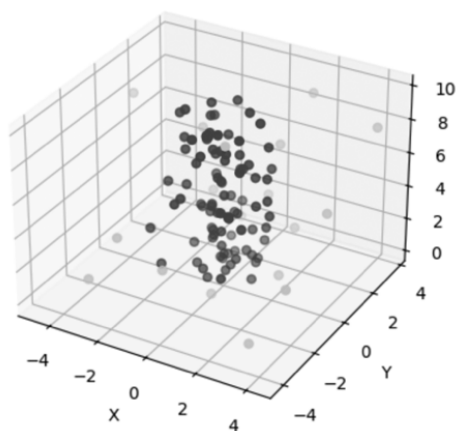


Рис. 13. Диаграмма рассеяния одноэтапной кластеризации алгоритмом Isolation Forest

в данных и определить некоторое количество шумовых элементов сразу. Этот этап также способствует сокращению объема данных, что делает возможным их последующий анализ с использованием более сложных алгоритмов кластеризации на втором этапе.

Заключение

Применение карты Кохонена на первом этапе анализа данных позволяет визуализировать и изучить их структуру, когда требуется общее представление без детального разделения на кластеры. После применения карты Кохонена для достижения более точного разделения данных на кластеры и выделения в них аномальных точек необходимо интегрировать традиционные алгоритмы кластеризации, которые проведут дополнительный анализ формируемых участков и выявят аномалии, которые могли быть упущены ранее.

Результаты, полученные на втором этапе, показали, что эффективность различных алгоритмов варьировалась в зависимости от размерности данных. Алгоритмы, такие как Isolation Forest и DBSCAN, могут требовать минимальной настройки параметров для достижения приемлемых результатов, в то время как OPTICS может требовать более глубокого понимания предварительных условий, чтобы добиться точного выделения кластеров.

Таким образом, двухэтапный метод кластеризации не только оптимизирует вычислительную нагрузку, но и улучшает конечные результаты кластеризации, обеспечивая более надежный и устойчивый анализ данных. Это особенно важно в контексте больших данных и сложных многомерных наборов, где традиционные подходы могут оказаться недостаточными.

Литература

1. Бухарев Д.А., Соколов А.Н., Рагозин А.Н. Применение иерархического кластерного анализа для кластеризации данных информационных процессов АСУ ТП, подвергающихся воздействию кибератак / Д.А. Бухарев, А.Н. Соколов, А.Н. Рагозин // Вестник УрФО. Безопасность в информационной сфере, 2023, Т. 1, №47. С. 59-68.
2. Митин Г.В., Панов А.В. Модификация алгоритма DBSCAN с использованием гибридных подходов к определению границ кластеров для обработки потоковых данных // Электронный научный журнал «ИТ-Стандарт», 2023, выпуск № 4, С. 36-57.
3. Мангутова Е.А. Обзор современных алгоритмов кластеризации данных / Мангутова Е.А., Гончаров А.С.; Томский политехнический университет, ОИТ // Молодежь и современные информационные технологии: сборник трудов XX Международной научно-практической конференции студентов, аспирантов и молодых учёных, 20-22 марта 2023 г., г. Томск. Томск: Изд-во ТПУ, 2023. С. 242-243.
4. Xia, Y., Wang D. A Survey on Density-Based Clustering Algorithms // International Journal of Computer Applications, 2015. 120(2). pp. 21-28.
5. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest // Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008. pp. 413-422.
6. Шкодырев В.П., Обзор методов обнаружения аномалий в потоках данных / В.П. Шкодырев, К.И. Ягафаров, В.А. Баштовенко, Е.Э. Ильина; Proceedings of the Second Conference on Software Engineering and Information Management, Saint Petersburg, Russia, April 21, 2017, p. 7.
7. Xu Rui, Wunsch D.C. Survey of Clustering Algorithms. Neural Networks, // IEEE Transactions 16(3). 2005. pp. 645-678. 10.1109/TNN.2005.845141.
8. Kohonen T. Self-Organizing Maps // Berlin: Springer-Verlag. 2001. p. 502.
9. Hodge, V.J. and Austin, J. A survey of outlier detection methodologies // Artificial Intelligence Review, (2004). 22 (2). pp. 85-126.
10. Hoaglin D.C. Volume 16: How to Detect and Handle Outliers. 2013. p. 77.
11. Гундина М.А., Богдан П.С., Южновская О.В. Особенности процесса определения количества аномальных значений при обработке измерительной информации // Вестник Белорусско-Российского университета. 2024, №2(83). URL: <https://cyberleninka.ru/article/n/osobennosti-protsessa-opredeleniya-kolichestva-anomalnyh-znacheniy-pri-obrabotke-izmeritelnoy-informatsii> (дата обращения: 31.10.2024).
12. Шадрин А.В. Визуализация трехмерных карт Кохонена с гексагональной решеткой // Цифровая обработка сигналов. 2015, №2. С. 23-27.

References

1. Buharev D.A., Sokolov A.N., Ragozin A.N. Primenenie ierarkhicheskogo klasterного analiza dlya klasterizacii dannykh informacionnykh processov ASU TP, podvergayushchikhsya vozdejstviyu kiberatak / D.A. Bukharev, A.N. Sokolov, A.N. Ragozin // Vestnik URFO. Bezopasnost' v informacionnoj sfere, 2023, T. 1, № 47. S. 59-68.
2. Mitin G.V., Panov A.V. Modifikaciya algoritma DBSCAN s ispol'zovaniem gibridnykh podkhodov k opredeleniyu granic klasterov dlya obrabotki potokovykh dannykh // Ehlektronnyj nauchnyj zhurnal «IT-Standart», 2023, vypusk № 4, S. 36-57.
3. Mangutova E.A. Obzor sovremennykh algoritmov klasterizacii dannykh / Mangutova E.A., Goncharov A.S.; Tomskij politekhnicheskij universitet, OIT // Molodezh' i sovremennye informacionnye tekhnologii: sbornik trudov XX Mezhdunarodnoj nauchno-prakticheskoy konferencii studentov, aspirantov i molodykh uchonykh, 20-22 marta 2023 g., g. Tomsk. — Tomsk: Izd-vo TPU, 2023. — S. 242-243.9. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest // Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008. pp. 413-422.
4. Xia, Y., Wang D. A Survey on Density-Based Clustering Algorithms // International Journal of Computer Applications, 2015. 120(2). pp. 21-28.
5. Liu F.T., Ting K.M., Zhou Z.H. Isolation Forest // Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008. pp. 413-422.
6. Shkodyrev V.P., Obzor metodov obnaruzheniya anomalij v potokakh dannykh / V.P. Shkodyrev, K.I. Yagafarov, V.A. Bashovenko, E.E. Il'ina; Proceedings of the Second Conference on Software Engineering and Information Management, Saint Petersburg, Russia, April 21, 2017, p. 7.
7. Xu Rui, Wunsch D.C. Survey of Clustering Algorithms. Neural Networks, // IEEE Transactions 16(3). 2005. pp. 645-678. 10.1109/TNN.2005.845141.
8. Kohonen T. Self-Organizing Maps // Berlin: Springer-Verlag. 2001. p. 502.

9. Hodge, V.J. and Austin, J. A survey of outlier detection methodologies // Artificial Intelligence Review, (2004). 22 (2). pp. 85-126.
10. Hoaglin D.C. Volume 16: How to Detect and Handle Outliers. 2013. p. 77.
11. Gundina M.A., Bogdan P.S., Yuhnovskaya O.V. Osobennosti processa opredeleniya kolichestva anomal'nykh znachenij pri obrabotke izmeritel'noj informacii // Vestnik Belorussko-Rossijskogo universiteta. 2024. №2 (83). URL: <https://cyberleninka.ru/article/n/osobennosti-protsessa-opredeleniya-kolichestva-anomalnyh-znacheniy-pri-obrabotke-izmeritel'noj-informatsii> (data obrashcheniya: 31.10.2024).
12. Shadrin A.V. Vizualizaciya trekhmernykh kart Kokhonena s geksonal'noj reshetkoj // Cifrovaya obrabotka signalov. 2015, №2. S. 23-27.
-

Плетенкова Анастасия Дмитриевна, аспирант кафедры защита информации, ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: pletenkovaad@susu.ru

Соколов Александр Николаевич, кандидат технических наук, доцент, заведующий кафедрой защиты информации ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: sokolovan@susu.ru

Pletenkova Anastasia Dmitrievna, post-graduate student of the Information Security, Federal State Autonomous Educational Institution of Higher Education "South Ural State University (national research university)". 454080, Chelyabinsk, Lenina avenue, 76. E-mail: pletenkovaad@susu.ru

Sokolov Alexander Nikolayevich, Candidate of Technical Sciences, Associate Professor, Head of Department of Information Security, Federal State Autonomous Educational Institution of Higher Education "South Ural State University (national research university)". 454080, Chelyabinsk, Lenina avenue, 76. E-mail: sokolovan@susu.ru.