



Обработка и подготовка данных наблюдений в интересах выделения особенностей динамики характеристик геоакустической эмиссии

Ю. И. Сенкевич*

Институт космофизических исследований и распространения радиоволн ДВО РАН,
684034, Камчатский край, Елизовский район, с. Паратунка, ул. Мира, 7, Россия

Аннотация. Деформация литосферного слоя под действием сейсмических процессов изменяет характеристики геоакустической эмиссии. Изучение особенностей динамики геоакустической эмиссии направлено на поиск признаков предсейсмических событий. Существует проблема качественной обработки сигналов геоакустической эмиссии и классификации получаемых результатов. Исследование направлено на поиск лучшей комбинации средств предобработки и кластеризации импульсного потока геоакустической эмиссии для выделения особенностей изменчивости характеристик такого сигнала. Обработываемые сигналы получены в ходе многолетних измерений в поверхностных слоях литосферы сейсмоактивного района Камчатского полуострова. Для выявления особенностей изменчивости характеристик сигналов геоакустической эмиссии осуществляется их структурно-лингвистическое преобразование в трехмерные отображения. Отображения обрабатываются, сравниваются и кластеризуются с применением сверточных нейронных сетей различной архитектуры. Лучший результат оценивается по трем подобранным критериям качества. Разработана методика поиска лучшего результата предобработки и кластеризации отображений. Представлен анализ результатов экспериментов.

Ключевые слова: обработка сигналов, распознавание образов, кластерный анализ, геоакустическая эмиссия, динамика характеристик сигнала, нейронные сети

Получение: 05.07.2024; Исправление: 06.08.2024; Принятие: 22.08.2024; Публикация онлайн: 25.08.2024

Для цитирования. Сенкевич Ю. И. Обработка и подготовка данных наблюдений в интересах выделения особенностей динамики характеристик геоакустической эмиссии // Вестник КРАУНЦ. Физ.-мат. науки. 2024. Т. 47. № 2. С. 75-94. EDN: ETVXVH. <https://doi.org/10.26117/2079-6641-2024-47-2-75-94>.

Финансирование. Работа выполнена за счет Государственного задания институте космофизических исследований и распространения радиоволн Дальневосточного отделения РАН (рег. № темы 124012300245-2).

Конкурирующие интересы. Конфликтов интересов в отношении авторства и публикации нет.

Авторский вклад и ответственность. Автор участвовал в написании статьи и полностью несет ответственность за предоставление окончательной версии статьи в печать.

***Корреспонденция:**  E-mail: senkevich@ikir.ru

Контент публикуется на условиях Creative Commons Attribution 4.0 International License



© Сенкевич Ю. И., 2024

© ИКИР ДВО РАН, 2024 (оригинал-макет, дизайн, составление)



Processing and preparation of observation data in the interests of highlighting the features of the dynamics of the characteristics of geoacoustic emission

*Y. I. Senkevich**

Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS,
684034, v. Paratunka, Mirnaya st., 7, Russia

Abstract. The lithospheric layer deformation under the action of seismic processes affects the characteristics of geoacoustic emission. The study of the geoacoustic emission dynamics is aimed at finding signs of pre-seismic events. There is a problem obtained for the high-quality processing of geoacoustic emission signals and the results classification. The study is aimed at finding the best combination of pre-processing and clustering tools for the pulse flow of geoacoustic emission to identify the features of the characteristics dynamics of such a signal. The processed signals were obtained during long-term measurements in the surface lithosphere layers of the seismically active region of the Kamchatka Peninsula. To identify the variability features of geoacoustic emission signals characteristics they are converted by sructurno-linguistic into a three-dimensional image. The images are processed, compared and clustered using convolutional neural networks of various architectures. The best result is assessed by three selected quality criteria. A technique has been developed for finding the best preprocessing and clustering result. The experimental result analysys are presented.

Key words: signal processing, pattern recognition, cluster analysis, geoacoustic emission, signal characteristics dynamics display, neural networks

Received: 05.07.2024; Revised: 06.08.2024; Accepted: 22.08.2024; First online: 25.08.2024

For citation. Senkevich Y. I. Processing and preparation of observation data in the interests of highlighting the features of the dynamics of the characteristics of geoacoustic emission. *Vestnik KRAUNC. Fiz.-mat. nauki.* 2024, 47: 2, 75-94. EDN: ETBXVH. <https://doi.org/10.26117/2079-6641-2024-47-2-75-94>.

Funding. The work was supported by the Institute of Cosmophysical Research and Radio Wave Propagation FAB RAS State Task (subject registration No. 124012300245-2).

Competing interests. There are no conflicts of interest regarding authorship and publication.

Contribution and Responsibility. The author participated in the writing of the article and is fully responsible for submitting the final version of the article to the press.

*Correspondence:  E-mail: senkevich@ikir.ru

The content is published under the terms of the Creative Commons Attribution 4.0 International License

© Senkevich Y. I., 2024

© Institute of Cosmophysical Research and Radio Wave Propagation, 2024 (original layout, design, compilation)



Введение

Одной из целей научных изысканий, проводимых лабораторией акустических исследований Института космофизических исследований и распространения радиоволн дальневосточного отделения Российской академии наук, состоит в поиске признаков предсейсмических событий [1, 2]. Изучается изменение физического состояния участка приповерхностного слоя литосферы Земли под действием локальных сейсмических явлений. Акцент исследования направлен на изучение связи физического состояния среды с характеристиками порождаемого ею сигнала геоакустической эмиссии (ГАЭ). В этой связи обработка сигналов ГАЭ приобретает ключевое значение в получении значимой информации для проведения исследований. Задача предобработки сигналов ГАЭ состоит в выделении групп фрагментов сигнала, которые имеют близкие между собой характеристики. Для выполнения такого поиска выполняется структурно-лингвистическое преобразование сигнала в трехмерное отображение [3]. Такие отображения отражают динамику характеристик сигнала на заданном временном масштабе времени. После выполнения преобразования задача дальнейшей предобработки сводится к сравнительному анализу полученных отображений и составлению групп по степени их сходства - кластерный анализ. Значимость, достоверность и полнота результата зависят от выполнения условий сочетаемости средств подготовки данных отображений со средствами выделения в группы по выбранным критериям сходства отображений. По этой причине успешное приближения к цели исследования прямо зависит от успешного решения задачи выбора алгоритма предобработки и алгоритма кластеризации отображений. Проведенные эксперименты показывают, что решение задачи выбора названных инструментов исследования требует отдельного подхода.

Постановка задачи исследования

Задача выделения аномалий сигнала ГАЭ более полно формулируется, как вычисление многообразия амплитудно-фазовых характеристик в импульсов потоке и формирование базового алфавита шаблонов типовых структур импульсов [4, 5]. Весь анализируемый сигнал разбивается на равные фрагменты, для которых вычисляются алфавиты и выстраиваются их распределения. Каждое полученное распределение элементов алфавита выкладывается на двумерную сетку, где по одной оси вносятся элементы алфавита по их размерности, а по другой — выстраиваются алфавиты, полученные на последовательно нарезанных фрагментах сигнала. В итоге формируется трехмерное изображение изменчивости состава алфавитов во времени, где третья ось представлена значениями частоты встречаемости элементов алфавита каждого фрагмента сигнала [6, 7]. Такое преобразование названо структурно-лингвистическим [21]. Сформированное отображение позволяет наблюдать динамику характеристики сигнала ГАЭ от одного временного фрагмента к другому. Анализируя динамику

характеристик на разных масштабах наблюдения сигнала ГАЭ, выделяются характерные отображения, которые указывают на возникающие аномалии изменения характеристик сигнала на фоне слабых отклонений на других временных интервалах. Визуально или технически можно найти аномалии, которые по своему характерному рисунку (паттерну), можно назвать близкими по контурам. Результат выполненного преобразования есть представление динамики выбранного параметра сигнала ГАЭ или - отображение. Как показала практика многочисленных экспериментов, проводимых в Лаборатории акустических исследований ИКИР ДВО РАН, значительные трудности составляет поиск методов определения степени сходства отображений. Опираясь на достоинства искусственного интеллекта с применением нейронных сетей (НС), которые доказали высокую эффективность их внедрения в практику цифровой обработки сигналов и распознавания образов, было принято решение использовать их возможности для формирования кластеров отображений. Эксперименты по применению нейронных сетей для анализа отображений показали серьезную зависимость результата сравнения отображений от настройки входных параметров в цепи согласованных алгоритмов предобработки и кластеризации сигналов ГАЭ. Таким образом, в настоящем исследовании выделение особенностей динамики характеристик геоакустической эмиссии сводятся к задаче поиска лучшей в комбинации последовательного применяемых алгоритмов предобработки данных с алгоритмами кластеризации, а также правильному выбору критериев оценки результатов проделанной обработки сигнала ГАЭ.

Подготовка базы данных экспериментов

В настоящее время в Лаборатории акустических исследований накоплена архивная база измерений ГАЭ, полученных на Камчатке в рамках государственного задания ФГБУ Институт космофизических исследований и распространения радиоволн Дальневосточного отделения наук РФ: Государственное задание по теме (2021–2023 гг.) "Физические процессы в системе ближнего космоса и геосфер при солнечных и литосферных воздействиях регистрационный номер АААА-А21-121011290003-0. Исходные данные являются отображениями, сформированными в результате выполнения авторской программы структурно-лингвистической обработки и представления сигналов ГАЭ [8, 9]. Для работы с выбранным стандартом нейросети, был подготовлен датасет, представляющий растровые совокупности изображений в формате Portable Network Graphics (PNG). Датасет составлен из отображений, представляющих результаты измерений сигналов с 2017 по 2021 гг. Каждое отображение отражает динамику за 1 месяц. Общее число исходных данных — 12 месяцев * 5 лет = 60. Исходные отображения аппроксимируются до цветных картин размеров 64x64x3. В результате аугментации число отображений было доведено до 3000.

На рис. 1 выведен пример 15 случайно выбранных отображений из подготовленной серии отображений за период с 2017 по 2021 годы. В имени файла зашифрована дата - <год><месяц> отображения.

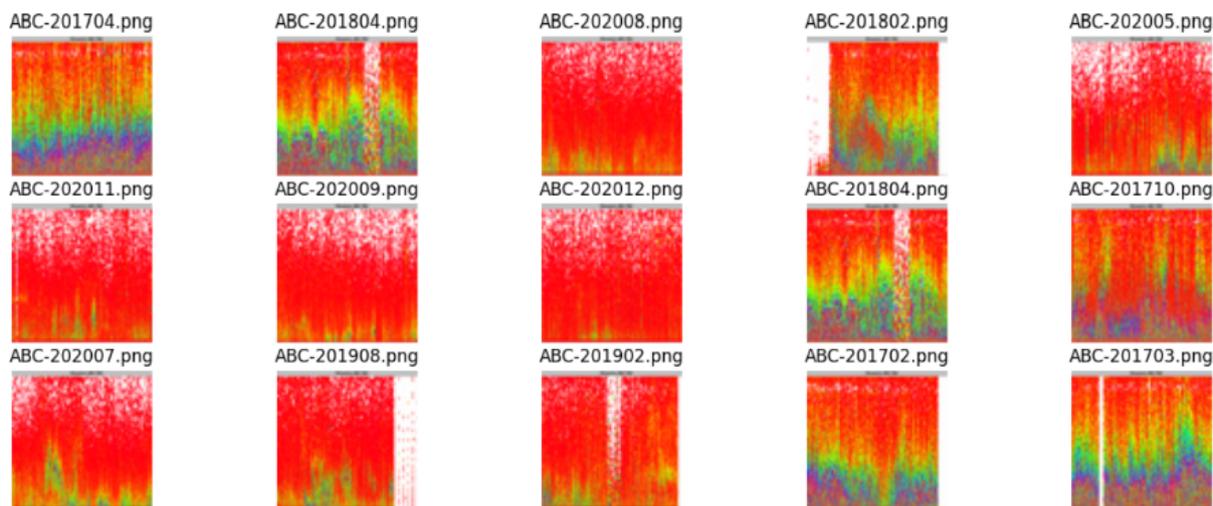


Рис. 1. Примеры отображений динамики характеристик геоакустической эмиссии
[Figure 1. Examples displays of the geoacoustic emission characteristics dynamics]

Выбор алгоритмов предобработки данных и алгоритмов кластеризации

Специфика полученных отображений состоит в отражении динамики характеристик сигнала, передаваемой изменениями палитры рисунка, что ассоциируется с представлениями о 'тепловых 3D картинах', где отсутствуют явно очерченные образы. Место объектов на картинах занимают градиенты цветности или уровней серого, которые несут основную информационную нагрузку. Этот факт стал еще одним аргументом в пользу принятия решения автором о поведении экспериментов с различными методами кластеризации, отличающихся по своей математической сущности (метрики) деления признакового пространства. Эксперименты с различными описаниями характеристиками сигналов и данных, направляемых на кластеризацию показывают, что результат распределения этих структур в фазовом пространстве, далеко не всегда соответствует представлениям экспериментатора об их правильном размещении. Как будет показано далее, результат кластеризации, имеющий лучший показатель оценки качества по заданному математическому критерию, оказывается парадоксальным с позиции субъективной оценки качества сформированных кластеров [10]. По этой причине было использовано четыре варианта алгоритмов кластеризации, реализующих различный подход к группированию свойств объектов. В пакет обработки попали следующие методы библиотеки Keros [11]:

- 'KMeans' (алгоритмом Ллойда) - метод K- средних использует алгоритм выбора центроидов, которые минимизируют инерцию или критерий суммы квадратов внутри кластера (минимум-дисперсионный подход);

- 'AgglomerativeClustering' (Агломеративная (иерархическая) кластеризация) - тип иерархической кластеризации, используемый для группировки объектов в кластеры на основе их сходства - минимизации евклидовой метрики;

- 'Birch' (алгоритма Берча) - протокол сбалансированных итераций и кластеризация с использованием иерархического метода;

- 'GaussianMixture' (модель гауссовой смеси, GMM) - GMM использует подход суждения о том, что вероятность, которая определяет окончательную категорию атрибуции, вычисляется через вероятность принадлежности к определённой категории. В качестве метрики близости объектов принимается вероятностная мера в отличие от предшествующих, где мера по умолчанию принята евклидовой.

Данные методы в программе запускаются с параметрами по очереди, с фиксированными глобальными параметрами используемых нейронных сетей. Это ограничение заметно сужает диапазон возможностей методики экспериментов. В данном случае важно было показать, как меняется результат в различных сочетаниях используемых алгоритмов предобработки и кластеризации. Многолетний опыт автора и опыт коллег РАН в этой области, [например, [12, 13]] по применению кластеризации указывает, что одним из фундаментальных положений в успехе распознавания групп объектов по их признакам является правильный выбор средств целевой обработки данных, предшествующих непосредственно процессу кластеризации, которые эти самые признаки выделяют. Выбор таких средств целевой предобработки является отдельной задачей и в большинстве случаев носит интуитивный характер, основанный на опыте работы с конкретными данными. Поэтому, применение различных методов предобработки отображений можно рассматривать как попытку найти (подобрать) оптимальный способ выделить характерные детали, с которыми далее будет работать метод кластеризации. Выделение таких деталей на практике задача неоднозначная в своей постановке. Трудно предсказать, какие детали отображения могут оказаться информационно значимыми. В конечном счете, эти аргументы и привели к необходимости проведения серии экспериментов с различными сочетаниями алгоритмов предобработки и кластеризации. Для экспериментов были выбраны следующие алгоритмы предобработки, которые широко используются в цифровой обработке сигналов:

- без применения НС - базовый вариант для понимания влияния препроцессинга на результат последующей кластеризации.

- с применением алгоритма, реализующего Метод главных компонент (РСА) [14] для концентрации информации. Задача алгоритма - снижение размерности в данных. Алгоритм широко используется в современных методах ИИ в качестве инструмента обучения без учителя.

- с применением НС VGG16 [15] для концентрации информации - сверточная НС. Задача алгоритма - выделение общих признаков распознавании отображений.

Использовался аппарат НС в той части, где она концентрирует признаковое пространство до подачи на первый полносвязный слой. Таким образом удалось полноценно использовать возможности предобработки отображений этим инструментом.

- с применением НС Автоэнкодер [16] для концентрации информации. Задача алгоритма - отсеять малоинформативные признаки, помехи, выбросы в данных. Алгоритм успешно используется в препроцессинге акустических и электрических сигналов, обучаются на неразмеченных данных с целью выявления внутреннего представления и эффективного кодирования информации

Выбор метрик оценки качества кластеризации

Для понимания получаемых результатов и возможности их корректного объективного сравнения необходимо было выбрать метрики оценки качества и достоверности алгоритмов кластеризации, которые наиболее подходили бы для работы с конкретным набором данных. Практика многочисленных исследований с применением кластеризации доказывает, что выбор определенной метрики зависит от конкретной задачи и особенностей данных. В некоторых случаях может оказаться эффективным использование метрик, основанных на измерении расстояний до центроидов или плоскостей, а также метрик, учитывающих структуру данных [17]. Таким образом, задача выбора и оптимизации метрик расстояний является открытой и активно исследуется в научных исследованиях. В данном исследовании использование различных метрик оценки качества выбранных методов предобработки и кластеризации позволило лучшим образом настраивать глобальные параметры рабочих алгоритмов НС. Для проведения экспериментов были выбраны следующие метрики оценки:

- Коэффициент силуэта [18] - позволяет измерить компактность и разделение кластеров. Он дает оценку в диапазоне от -1 до 1, где более высокое значение указывает на лучшее качество кластера. Положительное значение указывает на то, что образцы хорошо соответствуют соответствующим кластерам, а отрицательное значение предполагает, что они, возможно, были отнесены к неправильным кластерам.

- Индекс Калинского-Харабаша [19] - рассчитывает отношение дисперсии между кластерами к дисперсии внутри кластера. Более высокий индекс подразумевает плотные и хорошо разделенные кластеры.

- Индекс Данна [20] - оценивает компактность и разделение кластеров. Это отношение минимального межкластерного расстояния к максимальному внутрикластерному расстоянию. Более высокое значение указывает на лучшую кластеризацию.

Методика проведения экспериментов по комбинированию методов препроцессинга данных с методами кластеризации картин

В ходе экспериментов выполнялась программа циклической смены комбинаций методов препроцессинга и кластеризации, включая расчеты для заданной последовательности кластеров согласно Таблице 1.

Таблица 1

Программа выполнения циклической смены комбинаций методов препроцессинга и кластеризации [Program for performing cyclic change of combinations of preprocessing and clustering methods]

№	Метод предобработки изображений	Метод кластеризации изображений
1.1	Без использования нейросетевых методов (Без НС)	KMeans
1.2		AgglomerativeClustering
1.3		Birch
1.4		GaussianMixture
2.1	Метода главных компонент (PCA)	KMeans
2.2		AgglomerativeClustering
2.3		Birch
2.4		GaussianMixture
3.1	Сверточная НС vgg16	KMeans
3.2		AgglomerativeClustering
3.3		Birch
3.4		GaussianMixture
4.1	Сверточная НС Автоэнкодер	KMeans
4.2		AgglomerativeClustering
4.3		Birch
4.4		GaussianMixture

Эксперименты проводились сериями, каждая из которых объединяет результаты применения одного из алгоритмов предобработки отображений поочередно с четырьмя из алгоритмов, реализующих метод кластеризации. Результаты каждой серии экспериментов для каждой из четырех комбинации с определенным методом предобработки изображений представлены ниже двумя блоками графиков. Первый блок графиков составлен из отражений полученных групп характеристик отображений в нормированной фазовой плоскости. Такое представление позволяет наглядно продемонстрировать компоновку

анализируемых картин в созданных кластерах. Слева-направо по рядам схемы представлены результаты конкретного метода кластеризации с числом кластеров 2,3,4,5 и 6 соответственно. Сверху вниз по строкам схемы представлены применяемые методы кластеризации в соответствии порядку их написания в табл. 1. Серым цветом выделены результаты, для которых предобработка данных или кластеризация не могли быть выполнены корректно. Второй блок графиков каждой серии экспериментов представляет сводную схему графиков изменений показателей качества выполненного анализа в зависимости от числа задаваемых кластеров в функциях кластерного анализа. Слева направо по рядам схемы представлены используемые методы кластерного анализа согласно последовательности их написания в табл. 1. Сверху вниз по строкам схемы представлены результаты оценки качества анализа для каждого из показателей качества в последовательности: Коэффициент силуэта, Индекс Калинского-Харабаша и Индекс Данна, соответственно.

Результаты проведения экспериментов

Результаты обработки и кластеризации первой серии экспериментов представлены на рис. 2 и 3.

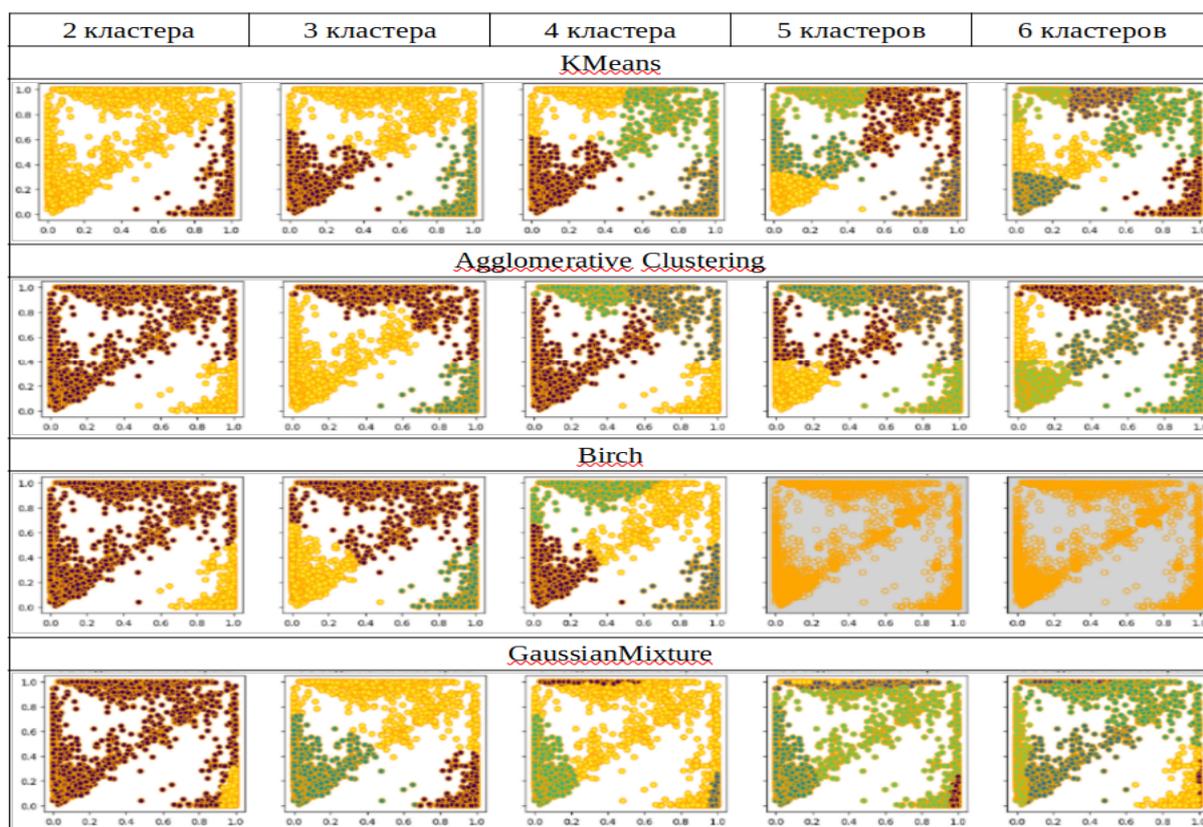


Рис. 2. Кластеризация в фазовом пространстве без выполнения предобработки отображений

[Figure 2. Clustering in phase space without preprocessing of mappings]

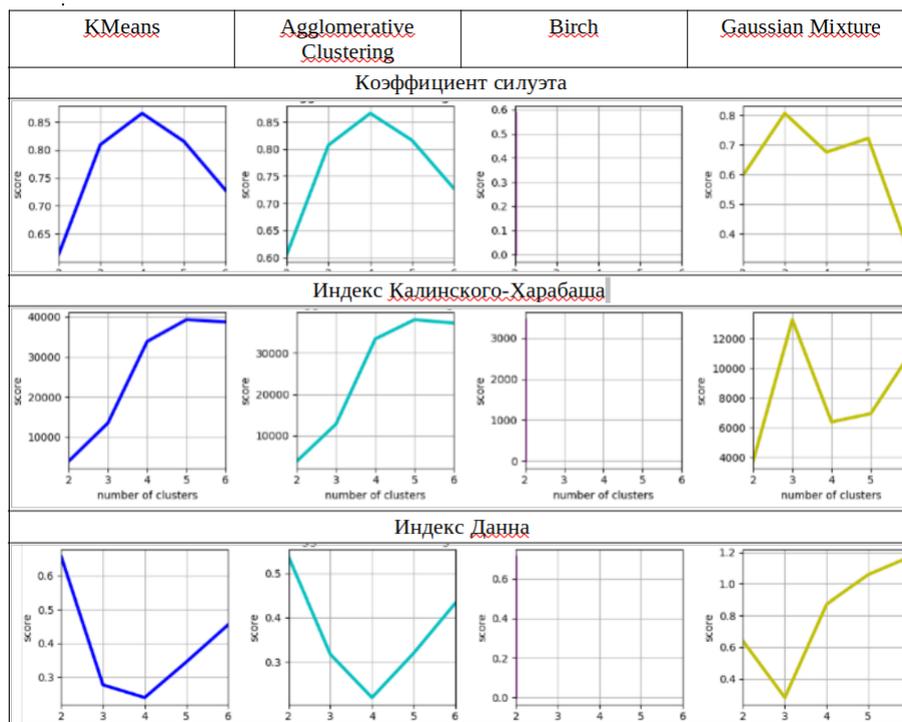


Рис. 3. Оценка качества кластеризации без выполнения предобработки отображений

[Figure 3. Assessing the clustering quality without performing preprocessing of mappings]

Применение апробированных методов непосредственно к подготовленному массиву картин указывают на достаточно слабое разделение данных на кластеры (лучшие значения контурных коэффициентов за серию из 10 экспериментов $\sim 0,76$ для метода KMeans и $\sim 0,74$ для метода AgglomerativeClustering). Практика экспериментов показывает, что эти значения могут снижаться до 0,42 в зависимости от порядка поступления отображений на обработку в перемешиваемом листинге имен входных файлов. Алгоритм Берча без предварительной обработки данных, как правило, выдает ошибку кластеризации, поскольку не может найти подходящих признаков для разделения отображений. Вероятностные критерии отбора метода GaussianMixture показывают наиболее низкие результаты оценки качества отбора картин.

Результаты проведения серии экспериментов с предобработкой методом главных компонент представлены на рис. 4 и 5.

Применение метода PCA, концентрирующего информацию в отображениях, вкупе с апробированными методами к подготовленному массиву отображений, позволило достичь хороших результатов разделения данных на кластеры (лучшие значения контурных к-тов за серию из 10 экспериментов $\sim 0,93$ для метода KMeans и $\sim 0,95$ для метода AgglomerativeClustering). Эти значения практически не снижались от эксперимента к эксперименту. Однако, есть видимые проблемы.

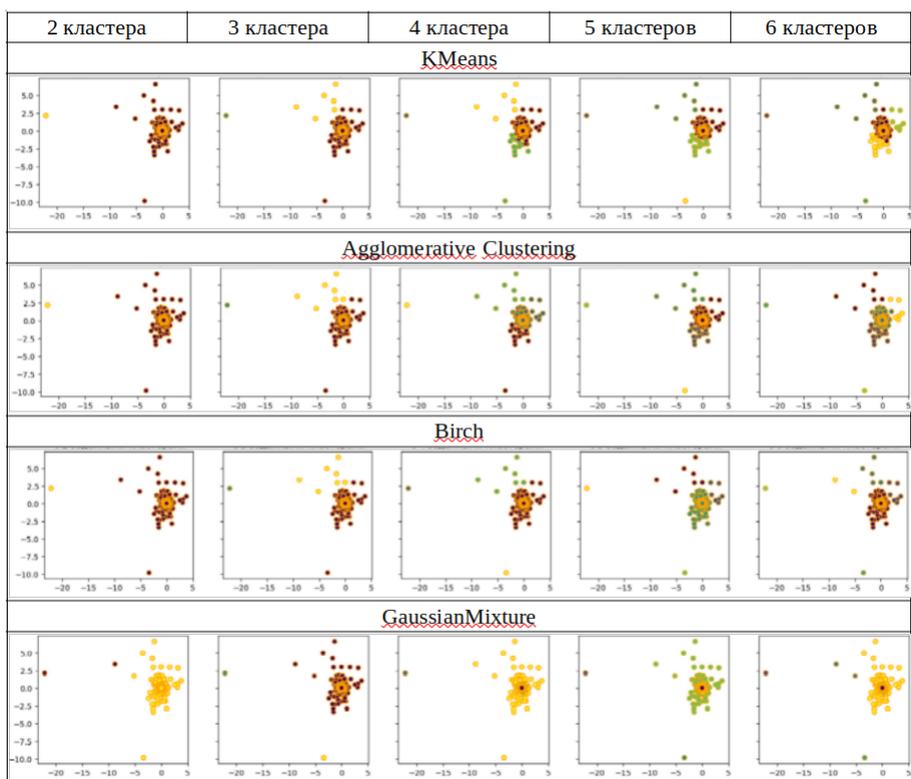


Рис. 4. Кластеризация в фазовом пространстве с применением предобработки методом главных компонент

[Figure 4. Clustering in phase space using principal component preprocessing]

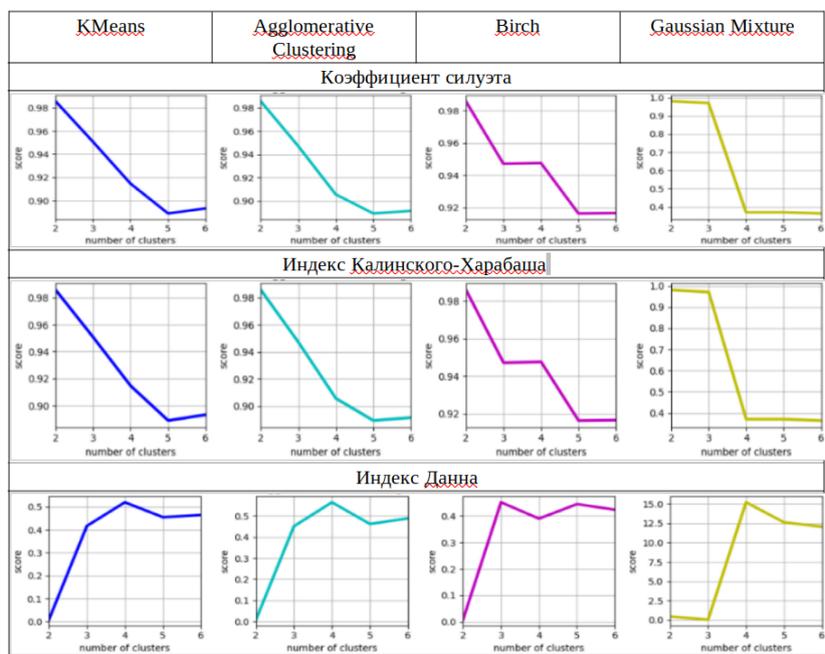


Рис. 5. Оценка качества с применением предобработки методом главных компонент

[Figure 5. Quality assessment using principal component analysis]

При внимательном рассмотрении картин распределения условных координат отличающих элементов обрабатываемых отображений, можно увидеть, что значение конструктивного k -та рассчитывается относительно единичных удаленных точек (редких отображений). Эти точки относятся к отображениям, которые по разным причинам отраженных на них деталей можно назвать выбросами. Что приводит на самом деле, по сути поиска, к слабому результату. Поэтому важно дополнительно визуально оценивать работу кластеризации, чтобы не войти в заблуждение при оценке значимости выбранного алгоритма предобработки. Отметим, что все методы показали одинаковый результат по числу кластеров с наивысшим рейтингом. Общая картина результатов позволяет увидеть область с высокой концентрацией объектов во всех экспериментах.

Результаты проведения серии экспериментов с предобработкой отображений в НС с архитектурой VGG16 представлены на рис. 6 и 7

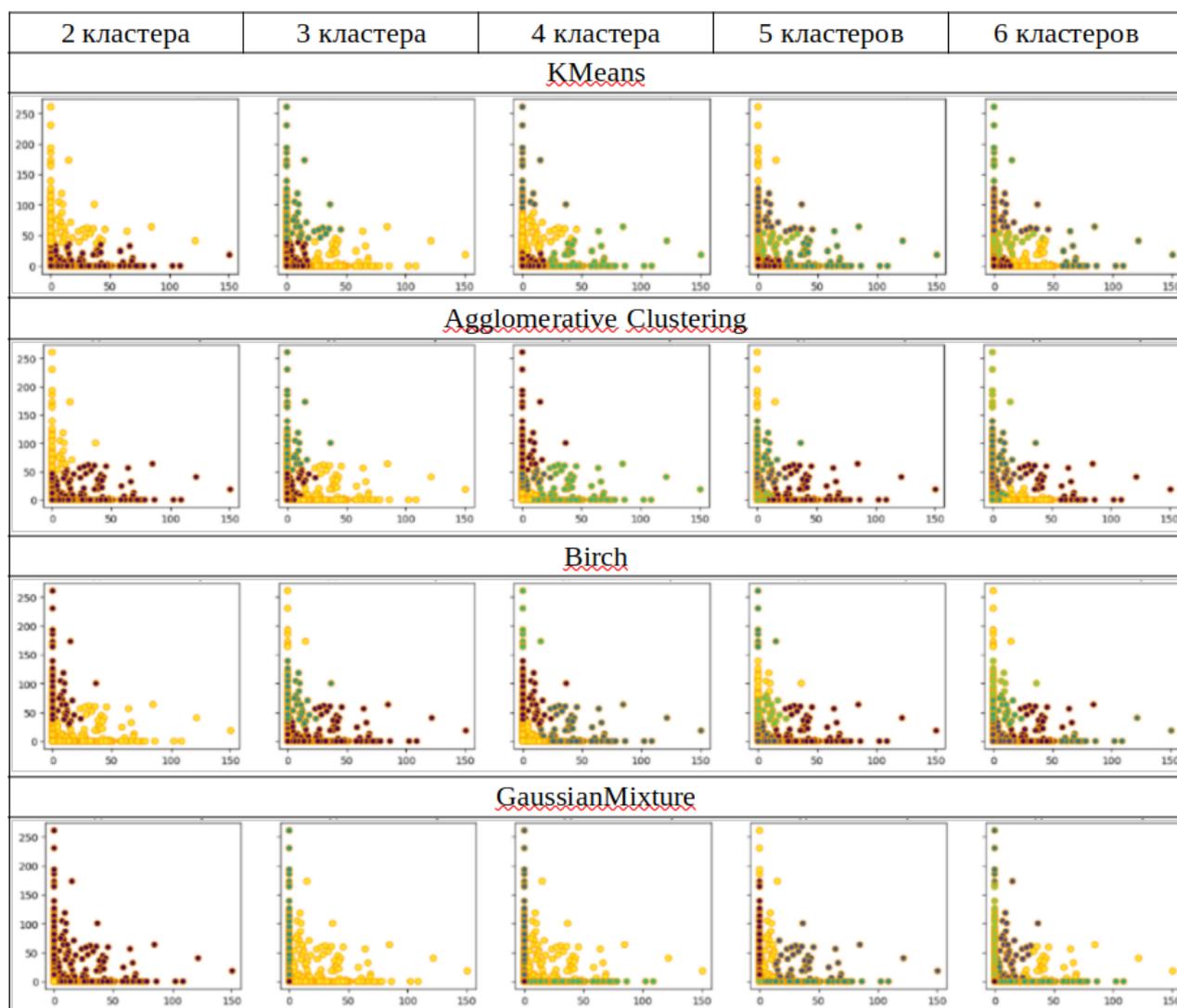


Рис. 6. Кластеризация с применением предобработки НС с архитектурой VGG16
[Figure 6. Clustering using NN preprocessing with VGG16 architecture]

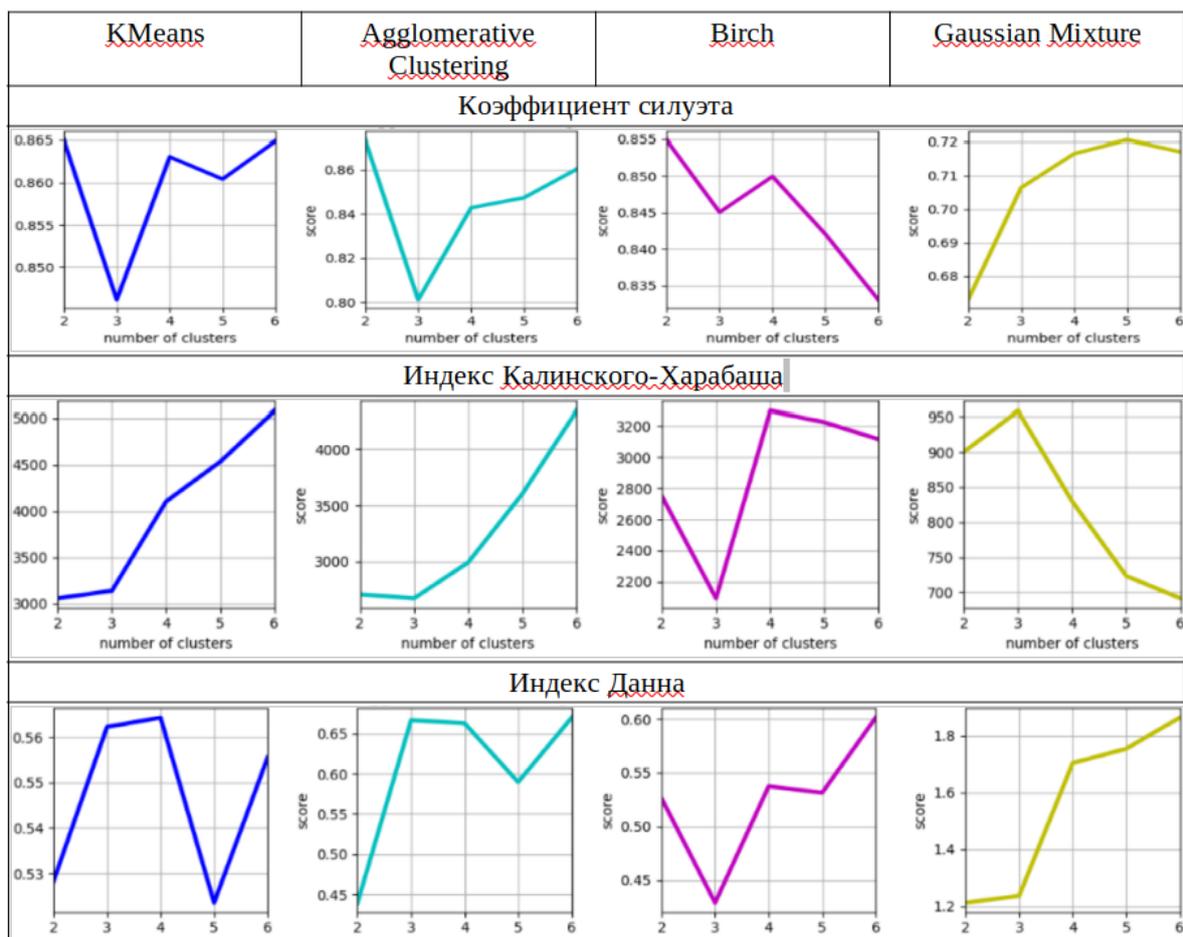


Рис. 7. Оценка качества с применением предобработки НС с архитектурой VGG16 [Figure 7. Quality assessment NN preprocessing with VGG16 architectures]

Предобработка отображений с использованием детализации аппаратом сверточных слоев модели к исходному массиву отображений показала хороший результат (лучшие значения контурных к-тов за серию из 10 экспериментов ~ 0,87 для метода KMeans и ~ 0,86 для метода AgglomerativeClustering). С небольшим преимуществом в рейтинге лидирует метод Берча ~ 087+. В случае применения методов кластеризации к предобработанным данным показывает на графиках распределение кластеров практически в ортогональной проекции на плоскости по отношению друг-друга, что и определяет такой высокий показатель контурного коэффициента. Такой результат можно объяснить тем, что архитектура VGG16 в части касающейся предобработки отображений, во многом напоминает алгоритм метода главных компонент, но при этом сохраняет баланс соотношений числа отображений между n кластерами порядка $1/n$ по частям. Интересное решение представлено по результатам оценки метода GaussianMixture, где наблюдается тенденция повышения контурного коэффициента пропорционально числу разбиений на кластеры. Это можно объяснить тем, что использование сверточных слоев в архитектуре модели VGG16, в конечном итоге приводят к повышению вероятностных связей по законам нормального распределения

выделенных характеристик отображений. Следует отметить еще одну особенность представленных графиков контурного к-та - это появление локальных минимумов и максимумов, которые можно интерпретировать как ситуацию, когда массив отображений «плохо бьется» на связанные группы объектов, и, когда появляются некоторые связанные подгруппы, соответственно.

Результаты проведения серии экспериментов с предобработкой отображений в НС с архитектурой — Кодер/Декодер представлены на рис. 8 и 9

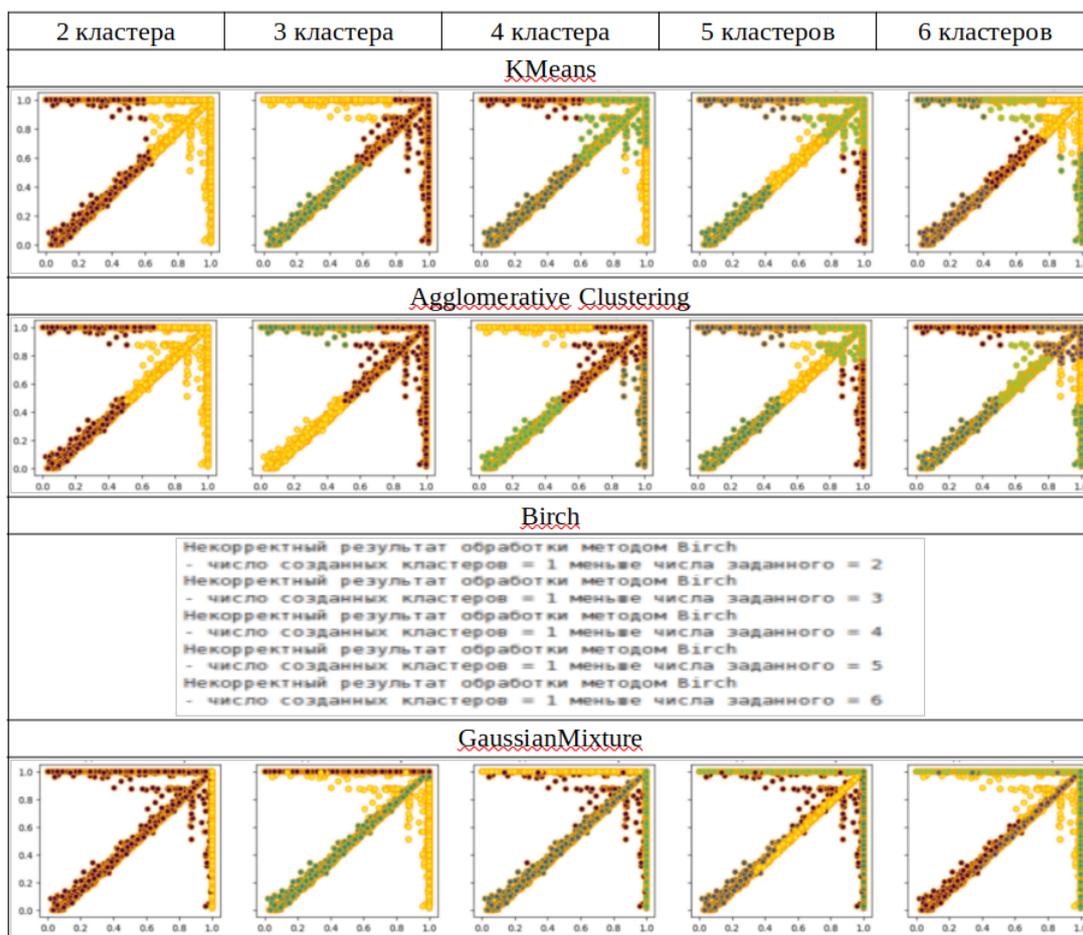


Рис. 8. Кластеризация с применением предобработки НС с архитектурой — Кодер/Декодер

[Figure 8. Clustering using NN preprocessing with Encoder/Decoder architecture]

Применение НС с архитектурой Кодер/Декодер показал интересный результат с позиции размещения меток отображений в фазовой плоскости. Кластеры сконцентрированы по осям и диагонали фазовой плоскости и оценка кластеризации по индексу Данна является наивысшей - 0,98. При этом отсутствуют выбросы, что ожидаемо в силу идеи обработки НС с архитектурой Кодер/Декодер, которая устраняет большую часть "шума" на возвращаемых отображениях. То есть кластеры имеют лучшие показатели отношения минимального межкластерного расстояния к максимальному внутрикластерному расстоянию. Такой результат явно можно назвать лучшим.

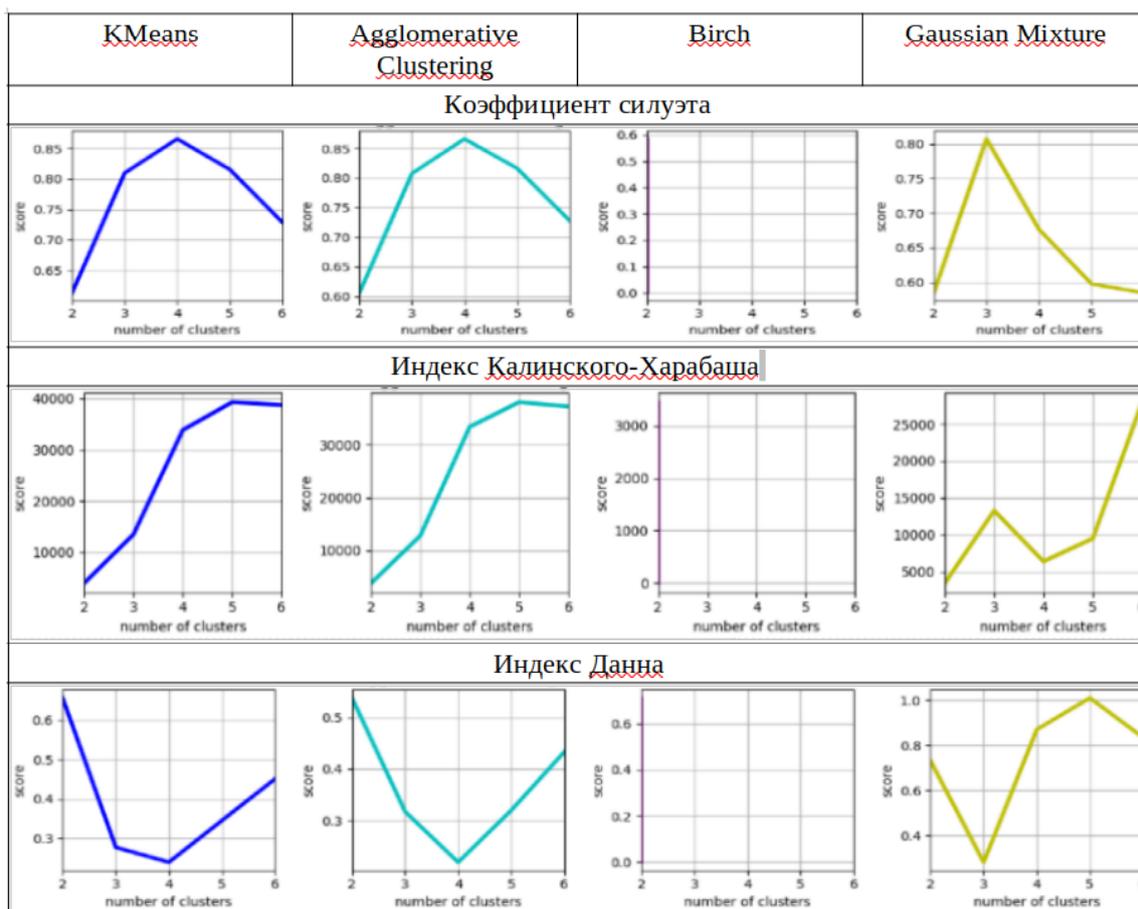


Рис. 9. Оценка качества с применением предобработки НС с архитектурой — Кодер/Декодер
 [Figure 9. Quality assessment using NN preprocessing with Encoder/Decoder architecture]

Анализ полученных результатов исследования

Для получения статистически обоснованных результатов было проведена серия экспериментов с их предварительной оценкой по выше представленной методике. Общее количество повторений экспериментов с каждой из комбинаций методов предобработки и кластеризации составило 60. Решение обосновано необходимостью многократного просмотра и сравнительного анализа полученных результатов множественных экспериментов для сопоставления обнаруженных групп (кластеров) с попыткой выявления их физического смысла. Показатели оценок результатов экспериментов от серии к серии колебались в незначительных пределах. Однако вывод по результатам кластеризации с использованием различных подходов к анализу получаемых кластеров не может быть сделан однозначно. Об этом можно судить из графиков сравнения критериев оценки качества результатов кластеризации, представленных на рис. 10.

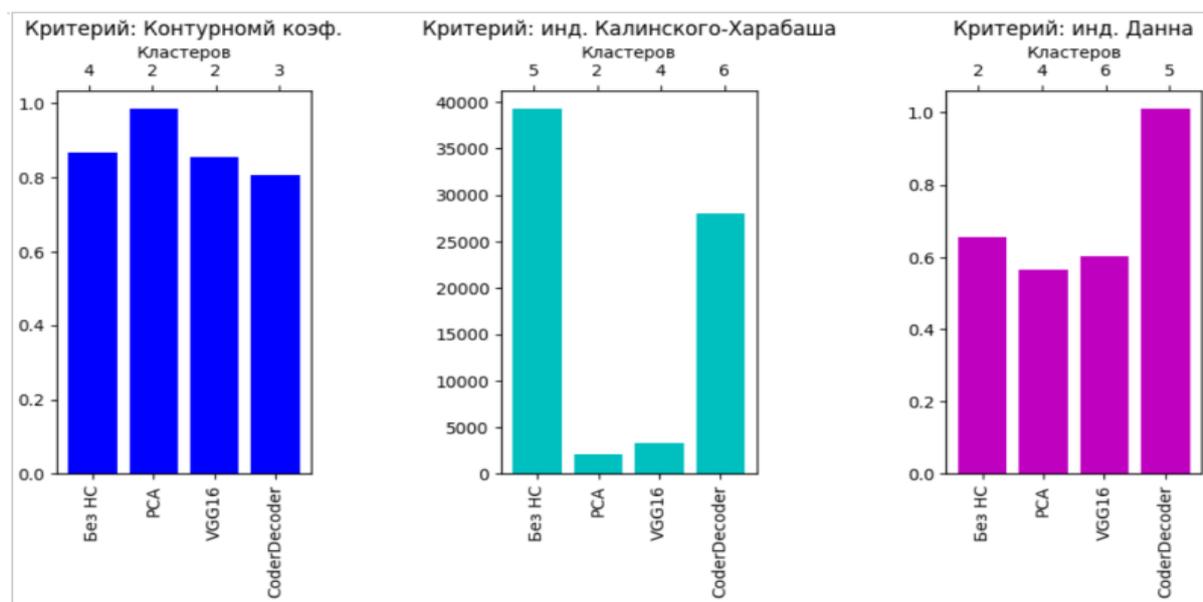


Рис. 10. Оценка качества с применением предобработки НС с архитектурой — Кодер/Декодер
 [Figure 10. Quality assessment using NN preprocessing with Encoder/Decoder architecture]

Заклучение

Анализ динамики характеристик сигнала геоакустической эмиссии затруднен в силу нестационарности потока составляющего его разнородных импульсов, что вызывает необходимость выполнения предварительной обработки и преобразования его в форматы, доступные для применения апробированных инструментов анализа данных. Одним из предлагаемых методов такого преобразования является символическое описание импульсов на основании их амплитудно-фазовой структуры и формирования алфавитов и формирования матрицы отображения динамики характеристик сигнала ГАЭ - структурно-лингвистическое преобразование.

Поиск специфических особенностей динамики характеристик сигнала геоакустической эмиссии, которые отражают влияние на сигнал множества неопределенных факторов, которые оказывают воздействие на генерирующую сигнал среду можно решать, используя современные достижения искусственного интеллекта на основе нейронных сетей. Среди наиболее подходящих нейронных сетей для распознавания образов выделяются сверточные нейронные сети (CNN), которые эффективно показали свои возможности при проведении экспериментов по кластеризации подготовленного датасета картин на базе множества отображений динамики характеристик сигнала геоакустической эмиссии.

В ходе серий экспериментов выяснилось, что невозможно однозначно априори определить лучшую комбинацию методов препроцессинга и кластеризации,

поскольку различные подходы к оценке качества результата кластеризации на выходе значительно расходятся по значениям. Поэтому в ходе исследования следует проводить эксперименты, позволяющие подобрать лучшее сочетание инструментов обработки и анализа, выполняя циклическую смену комбинаций методов препроцессинга и кластеризации. Выбор метода предобработки определяющим образом влияя на результат кластеризации в цепи комбинаций выбираемых методов.

Проведенное исследование позволило отработать экспериментальную методику выделения выделения групп картин, с близкими структурными показателями динамики характеристик геоакустической эмиссии.

Аббревиатуры

ГАЭ Геоакустическая эмиссия
НС нейронная сеть

Список литературы

1. Марапулец Ю. В. и др. *Комплексный анализ акустических и электромагнитных сигналов для оценки уровня сейсмической опасности*, Учебное пособие для студентов, обучающихся по специальности 011200 – геофизика., Т. 2 (Сейсмология). Владивосток: Дальнаука, 2020. 120 с.
2. Гапеев М. И., Марапулец Ю. В. Моделирование зон относительных сдвиговых деформаций перед сильными землетрясениями на Камчатке, произошедшими в период 2018-2021 гг, *Вестник КРАУНЦ. Физ.-мат. науки.*, 2021. Т. 37, № 4, С. 53-66 DOI: 10.26117/2079-6641-2021-37-4-53-66.
3. Senkevich Yu. I. Search for Hidden Patterns in Acoustic and Electromagnetic Pulse Signals, *IEEE*, 2020 DOI: 10.1109/SCM50615.2020.9198754.
4. Сенкевич Ю.И., Луковенкова О.О., Солодчук А.А. Методика формирования Реестра геофизических сигналов на примере сигналов геоакустической эмиссии, *Геосистемы переходных зон*, 2018. Т. 2, С. 409-418 DOI: 10.30730/2541-8912.2018.2.4.409-418.
5. Сенкевич Ю.И. и др. Способ обнаружения комплексного предвестника землетрясений, пат. 2758582, Рос. Федерация G01V 11/00, G01V 1/00, G01V 3/12 (ФГБУН ИКИР ДВО РАН. № 2020138668; 26.11.2020, Бюл. № 31. 9 с.).
6. Сенкевич Ю. И., Мищенко М. А. Методика оценки состояния приповерхностных осадочных пород по результатам наблюдений динамических характеристик геоакустической эмиссии, *Вестник КРАУНЦ. Физ.-мат. науки*, 2023. Т. 45, № 4, С. 109-121, DOI: 10.26117/2079-6641-2023-45-4-109-121.
7. Мищенко М.А., Сенкевич Ю.И., Щербина А.О. Современные методы обработки и анализа импульсных геофизических сигналов, *Вестник КРАУНЦ. Физ.-мат. науки*, 2022. Т. 41, № 4, С. 120-136 DOI: 10/26117/2079-6641.
8. Senkevich Yu., Marapulets Yu., Lukovenkova O., Solodchuk A. Technique of informative features selection in geoacoustic emission signals, *SPIIRAS*, 2022. vol. 18, no. 5, pp. 1066-1092 DOI: 10.1051/e3sconf/201912702005.
9. Senkevich Yu. I. Auto clustering of the variety of pulse signals based on their symbolic description, 2019. vol. 19 / E3S Web of Conferences, pp. 14.
10. Yang Z., Algesheimer R., Tessone C. J. *A Comparative Analysis of Community Detection Algorithms on Artificial Networks* <https://www.nature.com/articles/srep30750>.
11. *Keras library* <https://keras.io/why-this-name-keras>.
12. Марков Л.С. *Теоретико-методологические основы кластерного подхода*. Новосибирск: ИЭОПП СО РАН, 2015. 300 с.
13. Альсова О.К. Алгоритмы кластеризации разнотипных данных на примере решения медицинской задачи, 2014. Т. 6, № 37, С. 156-169 <https://doi.org/10.15622/sp.37.10>.

14. Milewski P. *PCA decomposition and Keras neural network*, <https://www.kaggle.com/code/pmmilewski/pca-decomposition-and-keras-neural-network>.
15. *Keras 3 API documentation / Keras Applications / VGG16 and VGG19* <https://keras.io/api/applications/vgg/>.
16. *The Keras Blog* <https://blog.keras.io/building-autoencoders-in-keras.html>.
17. *AskPython* <https://www.askpython.com/python-modules/keras-metrics>.
18. *GeekforGeek* <https://www.geeksforgeeks.org/clustering-performance-evaluation-in-scikit-learn>.
19. *Calinski harabasz score scikit-learn developers* <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.calinski-harabasz-score.html>.
20. *GeekforGeek* <https://www.geeksforgeeks.org/dunn-index-and-db-index-cluster-validity-indices-set-1/>.
21. Senkevich Yu. I, Duke V. A., Mishchenko M. A., Solodchuk A. A. Information approach to the analysis of acoustic and electromagnetic signals, *E3S Web of Conferences*, 2017. vol. 20, no. 02012, pp. 9 DOI:10.1051/e3sconf/2017200201.

Информация об авторе



Сенкевич Юрий Игоревич ✉ – доктор технических наук, доцент, ведущий научный сотрудник, Лаборатории акустических исследований, Институт космофизических исследований и распространения радиоволн ДВО РАН, Паратунка, Россия,  ORCID 0000-0003-0875-6112.

References

- [1] Marapulets Yu.V., et al. Kompleksnyy analiz akusticheskikh i elektromagnitnykh signalov dlya otsenki urovnya seysmicheskoy opasnosti [Integrated analysis of acoustic and electromagnetic signals to assess the level of seismic hazard]. Vladivostok: Dal'nauka, 2020. – 120 p.,(In Russian).
- [2] Gapeev M. I., Marapulets Yu. V. Modeling of relative shear deformation zones before strong earthquakes in Kamchatka from 2018-2021. Vestnik KRAUNC. Fiz.-mat. nauki. 2021. vol. 37. No 4. pp. 53-66. DOI: 10.26117/2079-6641-2021-37-4-53-66 (In Russian).
- [3] Senkevich Yu. I. Search for Hidden Patterns in Acoustic and Electromagnetic Pulse Signals. IEEE, 2020. DOI: 10.1109/SCM50615.2020.9198754.
- [4] Senkevich Yu. I., Lukovenkova O. O., Solodchuk A. A. Metodika formirovaniya Reestra geofizicheskikh signalov na primere signalov geoakusticheskoy emissii. Geosistemy perekhodnykh zon. 2018. vol. 2. no. 4. pp. 409-418. DOI: 10.30730/2541-8912.2018.2.4.409-418 (In Russian).
- [5] Sposob obnaruzheniya kompleksnogo predvestnika zemletryaseniy: pat № 2758582 Ros. Federatsiya: G01V 11/00, G01V 1/00, G01V 3/12 / Senkevich Yu.I., Marapulets Yu.V., Lukovenkova O.O., Solodchuk A.A., Mishchenko M.A., Malkin E.I., Gapeev M.I.; zayavitel' i patentoobladatel' FGBUN IKIR DVO RAN. № 2020138668; zayavl. 26.11.2020; opubl. 29.10.2021, Byul. № 31. 9 p. (In Russian).
- [6] Senkevich Yu. I., Mishchenko M. A. Method for estimation of near-surface sedimentary rock state based on the results of observations of geoacoustic emission dynamic characteristics. Vestnik KRAUNC. Fiz.-mat. nauki. 2023. vol. 45, no. 4. pp. 109-121. DOI: 10.26117/2079-6641-2023-45-4-109-121. (In Russian).
- [7] Mishchenko M. A., Senkevich Yu. I., Shcherbina A. O. Modern methods of processing and analysis of geophysical pulse signals. Vestnik KRAUNC. Fiz.-mat. nauki. 2022. vol. 41. no. 4. pp. 120-136. DOI: 10/26117/2079-6641 (In Russian).
- [8] Senkevich Yu., Marapulets Yu., Lukovenkova O., Solodchuk A. Technique of informative features selection in geoacoustic emission signals. SPIIRAS Proceedings. 2019, no. 18(5), pp. 1066-1092. DOI: 10.15622/sp.2019.18.5.1066-1092 (In Russian).
- [9] Senkevich Yu. I. Auto clustering of the variety of pulse signals based on their symbolic description. E3S Web of Conferences. vol. 127. 2019. 14. DOI: 10.1051/e3sconf/201912702005
- [10] Performance Evaluation Metrics GeekforGeek, <https://www.geeksforgeeks.org/clustering-performance-evaluation-in-scikit-learn>
- [11] Keras library documentation <https://keras.io/why-this-name-keras>.
- [12] Markov L. S. Teoretiko-metodologicheskie osnovy klaster'nogo podkhoda [Theoretical and methodological foundations of the cluster approach]. Novosibirsk: IEOPP SO RAN, 2015. 300 p. (In Russian).
- [13] Al'sova, O. K. Algorithms for Clustering of a Heterogeneous Data on the Example of Solution of the Medical Task, SPIIRAS Proceedings. 2014. vol. 6, No 37, pp. 156-169 DOI: 10.15622/sp.37.10. (In Russian.).
- [14] Milewski P. PCA decomposition and Keras neural network, <https://www.kaggle.com/code/pmmilewski/pca-decomposition-and-keras-neural-network>
- [15] VGG16 and VGG19 Keras 3 API documentation / Keras Applications / VGG16 and VGG19 <https://keras.io/api/applications/vgg/>
- [16] Building Autoencoders in Keras The Keras Blog <https://blog.keras.io/building-autoencoders-in-keras.htm>

- [17] Keras Metrics – A Complete Guide AskPython <https://www.askpython.com/python-modules/keras-metrics>.
- [18] Performance Evaluation Metrics GeekforGeek <https://www.geeksforgeeks.org/clustering-performance-evalua>
- [19] Calinski harabasz score scikit-learn developers <https://scikit-learn.org/stable/modules/generated/sklearn>
- [20] Dunn index and DB index – Cluster Validity indices GeekforGeek <https://www.geeksforgeeks.org/dunn-index-and-db-index-cluster-validity-indices-set-1/>
- [21] Senkevich Yu. I , Duke V. A., Mishchenko M. A., Solodchuk A. A. Information approach to the analysis of acoustic and electromagnetic signals. E3S Web of Conferences. 2017. vol. 20. 02012. 9. DOI: 10.1051/e3sconf/2017200201

Information about the author



Senkevich Yury Igorevich ✉ – D. Sci. (Tech),
Docent, Leading Researcher, Laboratory of Acoustic
Research, Institute of Cosmophysical and Radio
Wave Propagation FEB RAS, Paratunka, Russia,
ORCID 0000-0003-0875-6112.