



МАШИННОЕ ОБУЧЕНИЕ В ЗАДАЧАХ СЕЙСМОФАЦИАЛЬНОГО АНАЛИЗА

Евгений Игоревич Корыткин¹, Георгий Михайлович Митрофанов^{2,✉}, Александр Максимович Камашев³

^{1,2,3}Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН,
630090, Новосибирск, просп. Акад. Коптюга, 3, Россия,

¹ООО «СахалинНИПИ нефти и газа», 693009, Южно-Сахалинск, ул. Амурская, 53, Россия,

^{2,3}Новосибирский государственный университет, 630090, Новосибирск, ул. Пирогова, 1, Россия,

²Новосибирский государственный технический университет, 630073, Новосибирск, пр. К. Маркса, 20, Россия,

¹evgeny_korytkin@mail.ru, <https://orcid.org/0009-0002-6185-6155>

²MitrofanovGM@ipgg.sbras.ru, <https://orcid.org/0000-0003-4203-2740>

³KamashevAM@ipgg.sbras.ru, <https://orcid.org/0000-0001-8368-1144>

Аннотация. Рассматриваются принципы машинного обучения и его использования при решении задач сейсмофациального анализа. Обсуждаются вопросы эффективности получаемых решений на качественном уровне, а также существенные моменты, влияющие на их эффективность. К последним относятся: качество данных, группы используемых атрибутов, особенности алгоритмов кластеризации. В качестве примера приводятся результаты, полученные по горизонтам, относящимся к бобриковско-турнейской толще нижнекаменноугольного возраста в юго-западной части Оренбургской области.

Ключевые слова: сейсморазведка 3D, классификация, кластеризация, сейсмический атрибут, метод «К-средних», метод Кохонена, байесовский классификатор, «обучение с учителем», выделение сейсмофации

Финансирование: исследование выполнено по плану базовых научно-исследовательских работ ИНГГ СО РАН (проект Минобрнауки РФ FWZZ-2022-0017).

Для цитирования: Корыткин Е.А., Митрофанов Г.М., Камашев А.М. Машинное обучение в задачах сейсмофациального анализа // Геофизические технологии. 2024. № 4. С. 50–63. doi:10.18303/2619-1563-2024-4-50.

MACHINE LEARNING METHODS IN SEISMIC FACIES ANALYSIS PROBLEMS

Evgeny I. Korytkin¹, Georgy M. Mitrofanov^{2,✉}, Aleksandr M. Kamashev³

^{1,2,3}Trofimuk Institute of Petroleum Geology and Geophysics, SB RAS, Koptug Ave., 3, Novosibirsk, 630090, Russia,

¹SakhalinNIPI Oil and Gas Ltd, Amurskaya Str., 53, Yuzhno-Sakhalinsk, 693009, Russia,

^{2,3}Novosibirsk State University, Pirogova Str., 1, Novosibirsk, 630090, Russia,

²Novosibirsk State Technical University, K. Marks Ave., 20, Novosibirsk, 630073, Russia,

¹evgeny_korytkin@mail.ru, <https://orcid.org/0009-0002-6185-6155>

²MitrofanovGM@ipgg.sbras.ru, <https://orcid.org/0000-0003-4203-2740>

³KamashevAM@ipgg.sbras.ru, <https://orcid.org/0000-0001-8368-1144>

Abstract. The principles of machine learning and its use in solving seismic facies analysis problems are considered. The issues of the efficiency of the obtained solutions at a qualitative level, as well as significant aspects influencing their

efficiency are discussed. The latter include: data quality, groups of attributes used, features of clustering algorithms. As an example, the results obtained for horizons related to the Bobrikovian–Tournasian strata of the Lower Carboniferous age in the southwestern part of the Orenburg region are given.

Keywords: 3D seismic exploration, classification, clustering, seismic attribute, K-means method, Kohonen method, Bayesian classifier, supervised learning, seismic facies extraction

Funding: The study was carried out as part of government assignment to the Russian Academy of Sciences in basic research, Project FWZZ-2022-0017.

For citation: Korytkin E.I., Mitrofanov G.M., Kamashev A.M. Machine learning methods in seismic facies analysis problems // Russian Journal of Geophysical Technologies. 2024. No. 4. P. 50–63. (In Russ.). doi:10.18303/2619-1563-2024-4-50.

ВВЕДЕНИЕ

В настоящее время сейсмofациальный анализ является важнейшим этапом качественной динамической интерпретации результатов сейсморазведочных работ. Основываясь на значительном опыте анализа сейсмических разрезов и расчетах различных сейсмических атрибутов, он позволяет выполнить идентификацию геологических объектов, проявляющихся в сейсмическом волновом поле. В результате появляется возможность определить детальное геологическое строение перспективных отложений и выделить сейсмические фации по схожести параметров волновых объектов. В связи с развитием методов машинного обучения геологи-интерпретаторы все чаще обращаются к алгоритмам кластеризации или классификации для решения задачи выделения фаций по сейсмическим данным. Обладая способностью к обучению на больших объемах данных и высокой точностью предсказаний, эти методы являются эффективным инструментом для анализа сложных объектов. В качестве критериев точности могут выступать результаты глубокого бурения, выполняемые в областях прогноза характеристик целевых геологических объектов.

Создание и развитие методов машинного обучения происходило параллельно с кибернетикой и компьютерами в 60–70 гг. двадцатого столетия [Нильсон, 1967; Дюран, Одел, 1977; Ту, Гонсалес, 1978; Вапник, 1979]. При рассмотрении задач классификации и распознавания образов появилась необходимость в использовании значительных вычислительных ресурсов, обеспечиваемыми мощными цифровыми системами. Геология, как и многие другие сложные науки, была одной из первых, у которой имелась высокая потребность в использовании соответствующих инструментов [Алексеев и др., 1988]. Но отсутствие достаточной формализации геологических понятий и объектов существенно ограничивало широкое практическое применение вычислительных средств при решении задач прогноза свойств геологических объектов. Поэтому важным этапом стало появление атрибутного анализа, связанного с сейсмическими исследованиями, и переход на его основе к описанию геологических объектов [Rijks, Jauffred, 1991]. За последние двадцать лет изданы сотни работ, демонстрирующие возможности использования методов машинного обучения при решении задач интерпретации. Обзоры соответствующих исследований можно найти в статьях [Zhao et al., 2015; Chopra, Marfurt, 2019; Owusu et al., 2024], а основы методов изложены в различных учебных пособиях [Чабан, 2004; Воронина и др., 2017; Мясников, 2023].

Значительное количество статей по соответствующим направлениям исследований опубликованы в зарубежных изданиях, что послужило поводом для подготовки настоящей работы на русском языке. В ней представлены принципы машинного обучения и возможности его эффективного применения при

решении задач фациального анализа. Основное внимание уделено двум подходам: обучению с учителем и без учителя. В рамках «обучения без учителя» проводилась кластеризация по сейсмическим атрибутам и форме сейсмического сигнала. «Обучение с учителем» основывалось на классификации по форме сейсмической записи с использованием информации из скважин. В качестве примера представлены результаты, полученные при определении перспективных зон распространения коллектора продуктивного терригенного пласта Б₂ в отложениях нижнего карбона.

МЕТОДЫ И АЛГОРИТМЫ

Направление искусственного интеллекта, известное как машинное обучение, получило широкое распространение во многих областях науки и промышленности, где требуется разработка и создание аналитических моделей, способных автоматически распознавать скрытые и ранее неизвестные закономерности в данных. Эти модели также могут самостоятельно обучаться и приобретать свойства, необходимые для распознавания имеющихся закономерностей. Ключевым компонентом машинного обучения является набор данных, предоставляемых исследователю. Он описывает объекты или процессы, подлежащие исследованию, отражает присущие им свойства и закономерности. Поэтому его называют обучающим множеством. Он может быть получен путем наблюдения объекта или процесса в прошлом или создан экспертом/аналитиком на основе гипотез, аналогий, личного опыта.

В процессе обучения модели, обучающие примеры, которые представляют наблюдения из обучающего множества, предъявляются ей последовательно. Этот процесс позволяет модели приобретать нужные свойства. Обучение модели является итеративной процедурой, где на каждом шаге корректируются параметры модели в соответствии с правилами, определяемыми алгоритмом обучения. Обучение продолжается до достижения моделью необходимой точности при выполнении преобразований (например, кластеризация, регрессия) в рамках методов машинного обучения с требуемой точностью. После успешного обучения и тестирования модель может быть использована для анализа рабочих данных. Наиболее широко используемыми технологиями машинного обучения являются [Воронина и др., 2017]: обучение без учителя и обучение с учителем.

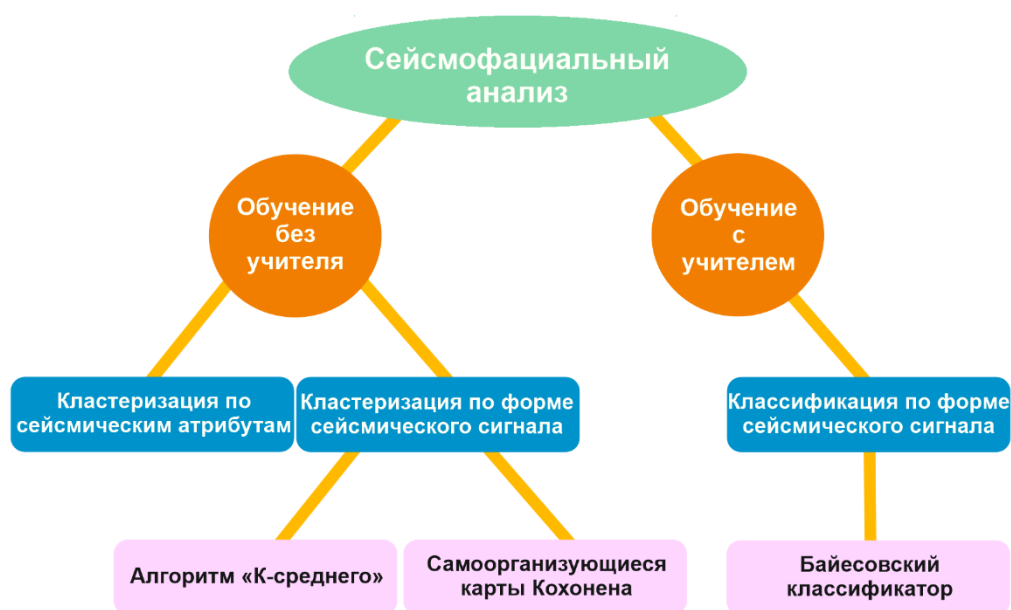


Рис. 1. Методы и алгоритмы, используемые при сейсмофациальном анализе.

В данной работе при сейсмофациальном анализе будет рассмотрено использование размеченных и неразмеченных данных в контексте обучения с учителем и без учителя. Особое внимание будет уделено двум масштабным задачам в области машинного обучения, которые широко применяются в сейсмофациальном анализе – кластеризация и классификация (рис. 1) [Coleou et al., 2003; Приезжев и др., 2014]. При этом кластеризация будет выполняться по сейсмическим атрибутам и по форме сейсмического сигнала. Во втором случае в качестве обучающей выборки используются скважинные данные.

Метод «K-средних» используется для кластеризации данных [MacQueen, 1967]. Он основан на разбиении векторного пространства на K заранее определенных кластеров и представляет собой итерационную процедуру, включающую следующие шаги, показанные на рис. 2. Вначале определяется необходимое число кластеров K . При его выборе могут использоваться различные критерии, в частности, эвристического типа, как в методе локтя, на что указывалось в работе [Митрофанов и др., 2024], но оно может определяться и другими соображениями, например, цветовой дифференциацией карт, как будет показано ниже.

После определения числа кластеров случайным образом выбираются K -наблюдений из исходного набора данных в качестве начальных центров кластеров. Для каждого из остальных наблюдений, входящих в исходный набор данных, определяется ближайший центр кластера, используя метрику Евклида или Махаланобиса при измерении расстояния [Хачумов, 2012]. Соответствующие метрики определяются выражениями:

$$\rho_E(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{\sum_{i=1}^n (x_i - \mu_i)^2},$$

$$\rho_M(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - \mu_i) \sum_{j=1}^n \text{cov}_{ij}^{-1} (x_j - \mu_j)},$$

где x_i – компоненты n -мерного вектора данных, μ_i – компоненты n -мерного вектора центра кластера, а cov_{ij}^{-1} – элементы обратной матрицы ковариаций для векторов соответствующего кластера. Если метрика Евклида определена для произвольных векторов, то метрика Махаланобиса относится к векторам, компоненты которых являются случайными величинами. Для независимых случайных величин матрица ковариаций имеет диагональный вид, и метрика Махаланобиса представляется выражением:

$$\rho_M(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{\sum_{i=1}^n (x_i - \mu_i)^2 \sigma_i^{-2}},$$

т. е. учитывает возможную неравномерность по компонентам у векторов, составляющих данный кластер.

Затем выполняется процесс формирования кластеров и уточнения их центров. Он состоит в следующем. Наблюдения, которые «притягиваются» определенным центром, объединяются в начальные кластеры. Для каждого кластера определяется центроид, представляющий собой центр тяжести кластера, т. е. вектор с компонентами, равными средним значениям соответствующих компонент наблюдений, входящих в кластер. Затем центр каждого кластера перемещается в его центроид, что приводит к появлению нового центра кластера. Процесс повторяется итерационно. Очевидно, что границы кластеров изменяются и их центры перемещаются на каждой итерации. В результате уменьшается расстояние между элементами внутри кластеров, и увеличиваются расстояния между кластерами.

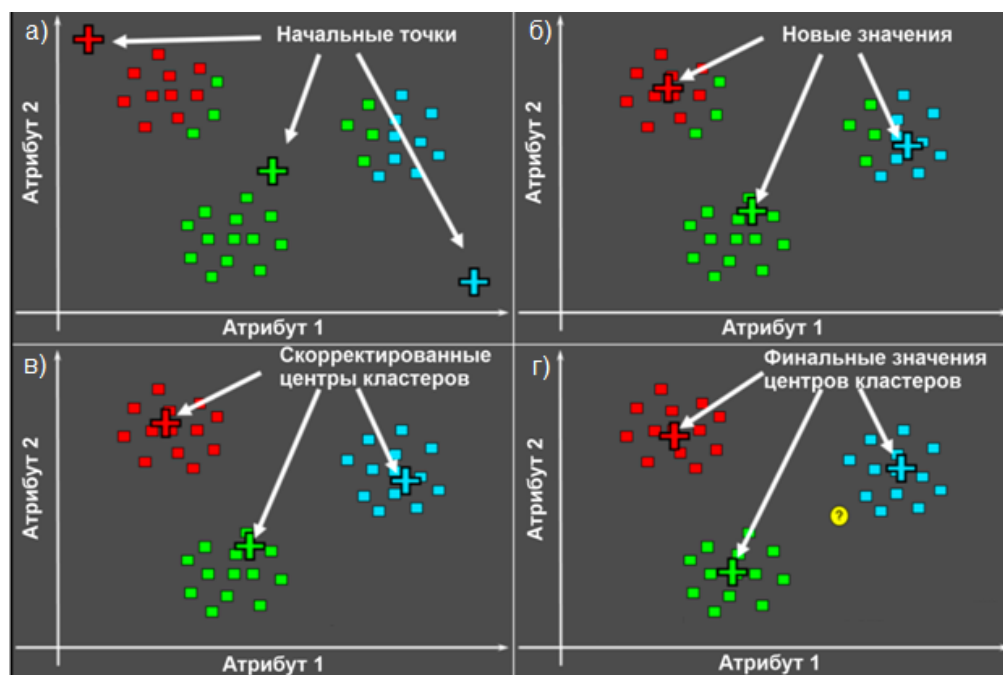


Рис. 2. Иллюстрация к методу «K-средних» (идея рисунка взята из [Zhao et al., 2015] с его модификацией).

На рисунке 2 показано выполнение процесса кластеризации по значениям двух атрибутов, которые являются компонентами векторов наблюдений, в три кластера. Первоначально выбираются три случайных или равноотстоящих, но различных вектора, служащих начальными центрами каждого кластера (рис. 2, а). Затем вычисляется расстояние Евклида или Махаланобиса между каждым вектором данных и каждым центром трех кластеров. Затем кодируется цветом или иным образом каждый вектор данных, принадлежащий кластеру с наименьшим расстоянием (рис. 2, б). Пересчитываются средние значения каждого кластера из ранее определенных векторов данных. Пересчитывается расстояние от каждого вектора до новых центров кластеров с определением принадлежности вектора кластеру по наименьшему расстоянию (рис. 2, в).

Процесс продолжается до тех пор, пока изменения в центрах кластеров не сойдутся в своих конечных местоположениях (рис. 2, г). Таким образом, остановка процесса кластеризации происходит, когда изменения в границах кластеров и местоположении центроидов прекращаются от итерации к итерации. Другими словами, на конечных итерациях в каждом кластере остается один и тот же набор наблюдений. Существует практическое подтверждение того, что предлагаемый алгоритм обычно находит стабильные кластеры после нескольких десятков итераций.

Следует отметить, что преимуществом данного алгоритма являются его скорость и простота реализации. Однако у него есть и недостатки, такие как: неопределенность в выборе начальных центров кластеров и необходимость изначально задавать количество кластеров, что может потребовать некоторую предварительную информацию о входных данных. Например, при проведении сейсмофацального анализа на новой площади необходимо иметь представление о количестве изучаемых фацальных обстановок, а также может потребоваться проведение предварительного атрибутивного анализа с целью охарактеризования получаемых значений атрибутов и их связей с целевыми геологическими объектами. При этом ключевым недостатком алгоритма является отсутствие гарантий

достижения глобального оптимума. В случае сложного набора данных результат напрямую зависит от начальных значений центроидов.

Еще одно замечание относится к выбору расстояний при формировании кластеров. В алгоритме «K-средних» классически используется расстояние Евклида. В алгоритме самоорганизующихся карт Кохонена [Kohonen, 1982] могут быть применены оба расстояния. В рассматриваемом ниже байесовском классификаторе используется расстояние Махаланобиса.

Классификатор необходимо применять при добавлении данных (на рис. 2 это точка, обозначенная желтым цветом). Он требуется для определения, в какой кластер попадают данные [Чабан, 2004; Мясников, 2023]. Для решения указанной задачи нами применялся один из простых вероятностных алгоритмов – байесовский классификатор, который иногда называют «наивный». Он основан на применении теоремы Байеса и делает «наивное» предположение о независимости признаков классифицируемых объектов [Friedman et al., 1997; Verma et al., 2012]. Алгоритм этого классификатора масштабируем по числу признаков, и обладает сопоставимой точностью с другими популярными методами, такими как «машины опорных векторов» [Вьюгин, 2013].

Как и любой другой, наивный байесовский классификатор присваивает метки классов наблюдениям, которые представлены векторами признаков. Этот классификатор предполагает, что каждый признак независимо влияет на вероятность принадлежности наблюдения к классу. Он считает, что каждый из признаков независимо влияет на вероятность того, что наблюдаемый объект является искомым объектом, без учета возможных корреляций между его характеристиками.

В основе построения использованного классификатора лежит обучение с учителем. Вопреки нереалистичному предположению о независимости признаков, простые байесовские классификаторы успешно применяются в решении разнообразных практических задач [Wickramasinghe, Kalutarage, 2021]. Дополнительным плюсом данного метода является его способность обучаться на небольшом количестве примеров. Фактически, байесовский классификатор представляет собой модель, основанную на вероятностях. Вероятностный классификатор определяет класс с наибольшей условной вероятностью для заданного вектора признаков.

В рамках данной работы байесовский классификатор был усовершенствован. На вход для проведения классификации помимо обучающей выборки добавлялись карты априорной вероятности по каждому набору классов. В отличие от существующих программных разработок, где алгоритм автоматически формирует условные вероятности, нами по скважинным данным и керну определяется число литоклассов и задаются карты априорных вероятностей. Таким образом, пользователь осуществляет активное участие в работе алгоритма. При этом фиксируется количество классов и устраняются условные вероятности. Предлагаемое усовершенствование повышает точность результатов классификации, что подтверждается результатами бурения. В дальнейших работах будут приведены результаты по анализу конкретных скважин.

РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ МЕТОДОВ НА РЕАЛЬНЫХ ДАННЫХ

Материалами для сейсмофациального анализа выступали 3D сейсмические данные, полученные в одном из районов Оренбургской области. Целевым интервалом исследований являлись отражения, относящиеся к горизонтам бобриковско-турнейской толщи нижнего карбона и регистрируемые во

временном интервале $t_0=1.44\text{--}1.55$ с. На начальном этапе исследований были выполнены шаги, являющиеся стандартными для промышленных компаний.

Перед кластеризацией и классификацией по сейсмическим данным был проведен анализ данных бурения, ГИС по 36 глубоким скважинам, а также кернового материала в интервале бобриковских отложений по пяти глубоким скважинам. По керну были выделены восемь различных фаций. Пять из них являлись неколлектором, одна была коллектором с пониженными ФЭС и две обладали хорошими коллекторскими свойствами. В процессе дальнейшего исследования при сейсмическом масштабировании были определены три сейсмofации: коллектор, плохой коллектор и неколлектор.

Кластеризация по сейсмическим атрибутам. Она носила стандартный характер. Первоначально был выполнен атрибутный анализ по целевому пласту Б₂. Этот анализ включал в себя восемь различных атрибутов: А1 – мгновенная частота, А2 – мгновенная фаза, А3 – мгновенная амплитуда, А4 – относительный акустический импеданс, А5 – мгновенная амплитуда к мгновенной частоте, А6 – хаос, А7 – когерентность, А8 – градиент. Описание первых пяти атрибутов дано в работе [Митрофанов и др., 2024]. Три последних относятся к стратиграфическим, предназначенным для выделения сейсмических структур, связанных с системами осадконакопления или оценки латеральной изменчивости. Эти же атрибуты позволяют выделять локальные вариации картины сейсмических отражений, направленность пластов и литологических объектов. При расчете атрибутов проводился выбор положения оптимального временного окна длительностью 10 мс и осуществлялось построение карт сейсмических атрибутов (рис. 3, а).

В процессе анализа построенных карт, а также при последующей кластеризации и классификации было отмечено, что в аномальных зонах, находящихся в северо-западной и северо-восточной частях площади, отражения имеют более выраженный и непрерывный характер с увеличением амплитуд. Эти характеристики могли указывать на более песчанистый разрез. Такое предположение подтверждалось открытием небольшой залежи, которая в последующем не разрабатывалась.

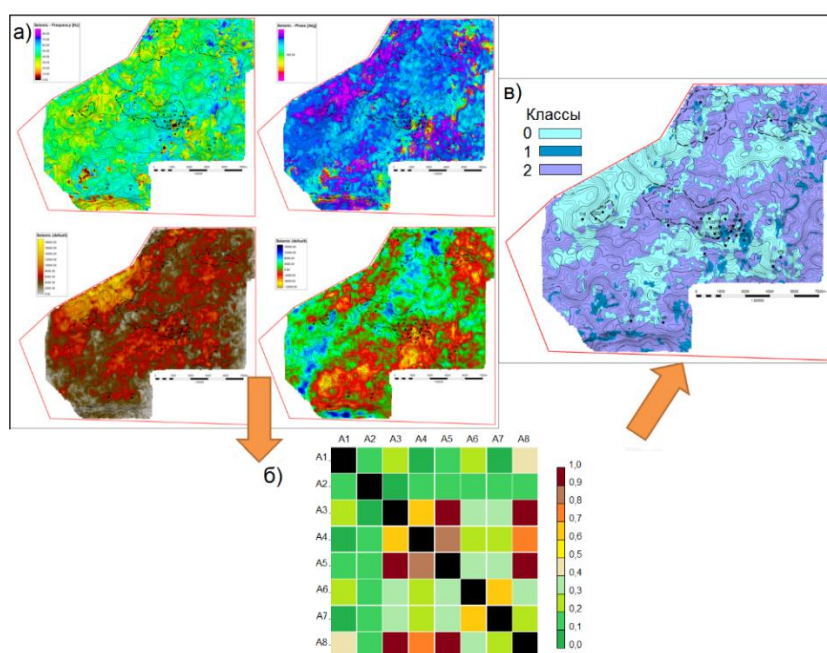


Рис. 3. Алгоритм кластеризации по сейсмическим атрибутам: а – примеры карт по пласту Б₂ (слева направо: мгновенной частоты, мгновенной фазы, мгновенной амплитуды, относительного акустического импеданса); б – корреляционная матрица по сейсмическим атрибутам; в – результирующая карта кластеров по отобранным атрибутам.

Затем была построена корреляционная матрица по картам (рис. 3, б) с целью исключения атрибутов с очень высокими коэффициентами корреляции (более 0.75) из процесса кластеризации, поскольку они не вносят дополнительной информации. В итоге было выбрано шесть атрибутов: A1, A2, A3, A4, A6, A7 и была рассчитана результирующая карта классов по сейсмическим атрибутам в интервале пласта Б₂ (рис. 3, в). В ходе тестирования по выбору количества классов опытным путем было выбрано не более трех, поскольку при большем количестве классов нарушается общность карты, и она становится недостаточно выраженной. Этими классами являлись: 0 – определяемый по картам мгновенной амплитуды и относительного акустического импеданса; 1 – определяемый по картам хаоса и когерентности; 2 – определяемый по картам мгновенной фазы и частоты.

При анализе результатов кластеризации по сейсмическим атрибутам отмечается наличие выраженного класса под номером 0, соотносимого с мгновенной амплитудой и относительным акустическим импедансом, выделяющей ранее выявленные аномалии на северо-западе и северо-востоке, приуроченные к залежам месторождений. Также выделяются зоны, выходящие за пределы залежей, что, возможно, может являться предпосылкой к наличию ловушек неструктурного типа, которые формировались в обстановках развития вдоль береговых баров и прибрежных валов. В то же время было отмечено, что устойчивых зависимостей кластеров с открытыми залежами и продуктивными скважинами не было обнаружено. Таким образом, решить задачу с выделением перспективных зон под постановку дальнейшего поисково-разведочного и эксплуатационного бурения представленным подходом не удалось.

Кластеризация по форме сейсмического сигнала. Форма сейсмических отражений зависит от процессов, формирующих целевые геологические горизонты, что позволяет проводить кластеризацию по форме сейсмического сигнала и выделять зоны схожих условий седиментации или зоны схожего интерференционного отклика. Для этого выделяются кластеры со схожими по форме сейсмическими трассами, в результате чего внутри каждого кластера имеются формы сигнала, имеющие минимальные различия друг с другом и не имеющие схожих характеристики с другими кластерами (рис. 4).

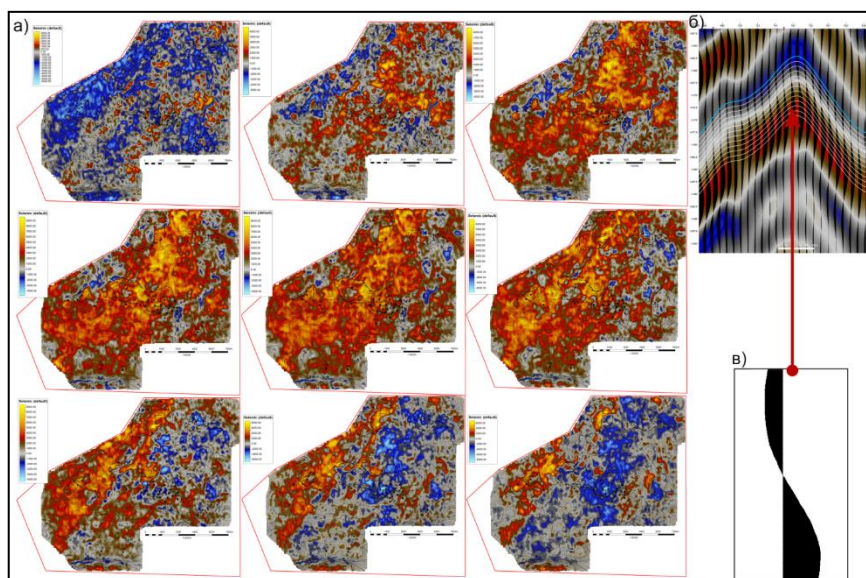


Рис. 4. Карты амплитуд по пласту Б₂, используемые для кластеризации по форме сигнала: а – примеры карт амплитуд в интервале пласта Б₂; б – фрагмент временного сейсмического разреза с нанесенными временными интервалами вдоль отражения от пласта Б₂; в – пример формы сейсмического сигнала.

При выполнении процесса кластеризации на сейсмических разрезах была выделена кровля целевого терригенного пласта Б₂ бобриковских отложений (на рис. 4, б выделена синим цветом), который характеризуется определенным геологическим строением. После чего пласт нарезается на множество пропластков с шагом 1 мс, производится извлечение амплитуды вдоль отражающего горизонта и изображается в виде карт. Было проведено составление 16 карт амплитуд на разных уровнях от кровли до подошвы пласта Б₂ (рис. 4, а).

Анализ всех полученных карт является весьма затруднительным и может привести к упущению важных деталей, таких как подобные или отличающиеся зоны. Поэтому использование методов кластеризации «К-средних» и самоорганизующихся карт Кохонена [Kohonen, 1982; Приезжев и др., 2014] позволяют уменьшить размерность задачи, объединить все карты в одну и выделить схожие зоны.

При реализации алгоритма кластеризации в рассматриваемом случае был задан поиск пяти кластеров. Центры кластеров были пересчитаны и перераспределены 85 раз по заданному количеству итераций. Итогом кластеризации с применением «К-средних» является карта, представленная на рис. 5, б). На рисунке 5а ниже центров кластеров представлены две случайные трассы, обозначенные черным и синим маркерами. Черный маркер находится в красной зоне, которая соответствует кластеру под номером 1. Синий маркер соответствует третьему центру кластера, который относится к желтой зоне на карте.

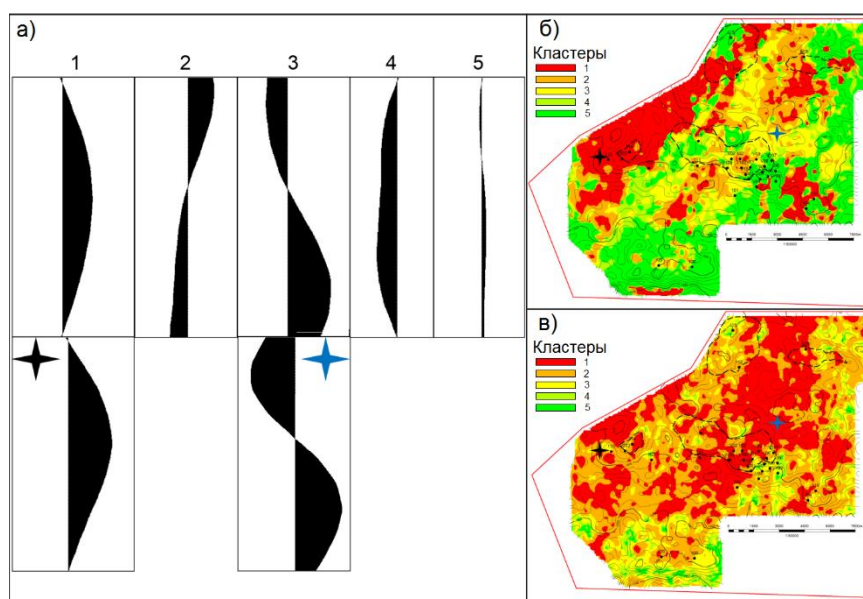


Рис. 5. Кластеризация по форме сейсмического сигнала: а – форма сейсмического сигнала центров кластеров и участки двух выбранных трасс; б – результирующая карта кластеров по алгоритму «К-средних»; в – результирующая карта по алгоритму самоорганизующихся карт Кохонена.

Последующее применение алгоритма самоорганизующихся карт Кохонена приводит нас к существенно отличающемуся результату (рис. 5, в). Путем детального сравнительного анализа полученных карт кластеров, основанных на форме сейсмического сигнала, можно сделать вывод, что ни одна из примененных методик сейсмофациального анализа, основанных на «обучении без учителя», не позволяет выявить устойчивую зависимость и совершить привязку ранее выявленных фаций и продуктивных зон коллекторов пласта Б₂ к определенному кластеру. Полученный вывод подтверждается ключевой особенностью метода «К-средних» (об отсутствии глобального оптимума).

В то же время в результате кластеризации были выявлены некоторые аномальные зоны в северо-западной части исследуемого участка. Но эти зоны имели различную классификацию при использовании метода «К-средних» и метода самоорганизующихся карт Кохонена.

Таким образом, стандартные подходы, используемые при сейсмофациальном анализе и основанные как на атрибутах, так и форме сигнала, оказались недостаточно эффективными в случае прогнозирования коллекторских свойств целевого горизонта. В связи с отсутствием наблюдаемой устойчивой зависимости, в дальнейшем использовался метод классификации «обучение с учителем» с применением наивного байесовского классификатора, использующего карты априорной вероятности (см. предыдущий раздел статьи).

Классификация по форме сейсмического сигнала с применением алгоритма байесовского классификатора. На предыдущих этапах используемые алгоритмы не имели исходной информации об имеющихся данных и опирались исключительно на внутреннюю структуру. Однако теперь машине предоставлялась априорная информация в виде скважинных данных с выделенными фациальными обстановками в интервале пласта Б₂.

Отметим, что методы машинного обучения отлично адаптируются под обучающую выборку и алгоритм способен безошибочно определять наличие или отсутствие искомых фаций и зон распространения коллектора. Однако у машинного обучения имеются проблемные моменты, такие как эффект переобучения и неустойчивость результатов. Для решения этих проблем в контексте обучения, входная выборка скважин была поделена на две части – обучающую и контрольную. При анализе сейсмических данных и получаемой карты классов скважинная информация по 26 скважинам выступала в роли учителя, а остальные 10 скважин были исключены и использовались в качестве контрольной выборки для проверки модели. Кроме того, на этапе интерпретации ГИС с распространением литофаций по скважинам в интервале пласта Б₂ проводился расчет априорных вероятностей каждого литокласса с дальнейшим построением карт априорных вероятностей (рис. 6). В зонах отсутствия скважинной информации вероятности распространялись согласно геологическим представлениям развития литотипов на участке исследования.

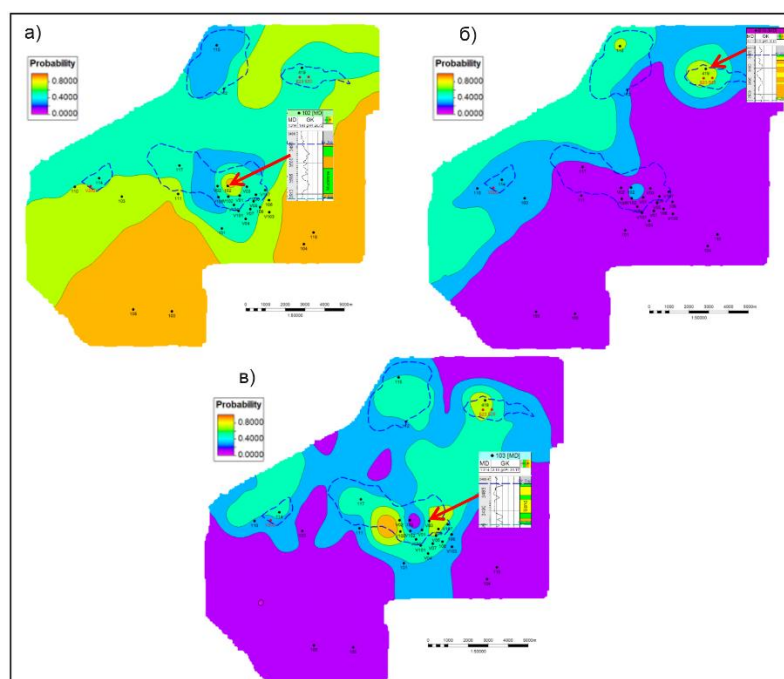


Рис. 6. Карты априорной вероятности: а – неколлектора; б – коллектора с ухудшенными ФЕС; в – коллектора.

В ходе валидации на основе многократных прогнозов фаций при исключении и включении разных скважин проводилась оценка, насколько увеличивалась корреляция, уменьшалась ошибка и проводилась оптимизация обучения. На выходе помимо результирующей карты литоклассов могут быть построены карты апостериорной вероятности распространения каждого класса. При последующем совмещении карты классов с картой вероятности можно оценить зоны наиболее вероятного распространения каждого класса фаций по пласту Б₂. Результаты соответствующих исследований будут представлены в отдельной работе.

ЗАКЛЮЧЕНИЕ

При выполнении сейсмофациального анализа активно используются методы машинного обучения. Они являются эффективным инструментом при обработке больших объемов данных, позволяя выявлять скрытые закономерности, что повышает точность предсказания обстановок осадконакопления. Алгоритмы кластеризации позволяют разделить данные на кластеры, определить типичные группы фаций по их свойствам и отобразить пространственное распределение различных типов фаций на месторождении. Но в некоторых случаях не удается находить устойчивые зависимости между кластерами и уже открытыми залежами месторождений и продуктивных скважин. Тогда целесообразно воспользоваться алгоритмами байесовского классификатора.

В отличие от атрибутного анализа, кластеризации и классификации с применением алгоритмов *K*-средних и самоорганизующихся карт Кохонена, где применяется два типа данных: сейсмическая и скважинная информация; при использовании байесовского классификатора добавляется третий тип данных – карты априорных вероятностей. Благодаря этому удастся минимизировать ошибку классификации, определить границы развития фаций и выделить новые перспективные объекты, прогнозируя зоны распространения коллекторов. Также отмечается выявление уверенных зависимостей пробуренных продуктивных скважин с доказанной нефтегазоносностью с выделенными литофациями по результатам классификации, чего другими алгоритмами добиться не удалось.

Недостатком байесовского классификатора является необходимость обеспечения представительной выборки по скважинам с проведением интерпретации по данным ГИС. Также требуется значительная информация о распространении литофаций на площади исследования, поскольку некорректно заданные априорные вероятности могут сильно исказить итоговый результат классификации. В дальнейшем предполагается подробно рассмотреть результаты классификации, полученные с применением байесовского классификатора, и сопоставить их с результатами последующего эксплуатационного бурения.

СПИСОК ИСТОЧНИКОВ

Алексеев А.С., Пяткин В.П., Дементьев В.Н., Алсынбаев К.С., Ангельский А.Н., Болычевский А.Б., Гилев К.А., Горбунов Б.А., Забелин В.А., Зятькова Л.К., Кирейтов В.Р., Клименко А.Д., Кондауров М.И., Копылов А.И., Кузин Г.И., Ладыжец В.С., Нифонтов В.П., Осипов Н.А., Райгель В.И., Репков В.В., Селиванов А.Н., Спектор А.А., Стахеев А.В., Юдин А.Н. Автоматизированная обработка изображений природных комплексов Сибири. Новосибирск: Наука, 1988. 222 с.

Вапник В.Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. 448 с.

Воронина В.В., Михеев А.В., Ярушкина Н.Г., Святков К.В. Теория и практика машинного обучения: Учебное пособие. Ульяновск: УлГТУ, 2017. 70 с.

- Вьюгин В.В.** Математические основы теории машинного обучения и прогнозирования. М.: МЦНМО, 2013. 387 с.
- Дюран Б., Одел П.** Кластерный анализ / Пер. с англ. М.: Статистика, 1977. 128 с.
- Митрофанов Г.М., Коваленко И.А., Корыткин Е.И.** Применение сейсмофациального анализа с использованием атрибутов для прогноза коллекторов // Геофизические технологии. 2024. № 3. С. 33–45. doi:10.18303/2619-1563-2024-3-33.
- Мясников В.В.** Распознавание образов и машинное обучение. Основные подходы: Учебное пособие. Самара: Изд-во Самарского университета, 2023. 196 с.
- Нильсон Н.** Обучающиеся машины / Пер. с англ. М.: Мир, 1967. 180 с.
- Приезжев И.И., Солоха Е.В., Манрал С.** Фациальный анализ по форме сейсмического сигнала // Геофизика. 2014. № 1. С. 63–67.
- Ту Дж., Гонсалес Р.** Принципы распознавания образов / Пер. с англ. М.: Мир, 1978. 411 с.
- Чабан Л.Н.** Теория и алгоритмы распознавания образов: Учебное пособие. М.: МИИГАИК, 2004. 290 с.
- Хачумов М.В.** Расстояния, метрики и кластерный анализ // Искусственный интеллект и принятие решений. 2012. № 1. С. 81–89.
- Chopra S., Marfurt K.J.** Unsupervised machine learning applications for seismic facies classification // Proceeding of the 7th Unconventional Resources Technology Conference. Denver, 2019. Paper URTEC-2019-557-MS. doi:10.15530/urtec-2019-557.
- Coleou T., Poupon M., Azbel K.** Unsupervised seismic facies classification: A review and comparison of techniques and implementation // The Leading Edge. 2003. Vol. 22 (10). P. 942–953. doi:10.1190/1.1623635.
- Friedman N., Geiger D., Goldszmidt M.** Bayesian network classifiers // Machine Learning. 1997. Vol. 29 (2–3). P. 131–163. doi:10.1023/A:1007465528199.
- Kohonen T.** Self-organized formation of topologically correct feature maps // Biological Cybernetics. 1982. Vol. 43. P. 59–69. doi:10.1007/BF00337288.
- MacQueen J.** Some methods for classification and analysis of multivariate observations // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California, Los Angeles LA USA, 1967. P. 281–297.
- Owusu B.A., Boateng C.D., Asare V.S., Danuor S.K., Adenutsi C.D., Quayе J.A.** Seismic facies analysis using machine learning techniques: a review and case study // Earth Science Informatics. 2024. Vol. 17. P. 3899–3924. doi:10.1007/s12145-024-01395-3.
- Rijks E.J.H., Jauffred J.C.E.M.** Attribute extraction: an important application in any detailed 3-D interpretation study // The Leading Edge. 1991. Vol. 10 (9). P. 11–19. doi:10.1190/1.1436837.
- Verma S., Roy A., Perez R., Marfurt K.J.** Mapping high frackability and high TOC zones in the Barnett Shale: Supervised Probabilistic Neural Network vs. unsupervised multi-attribute Kohonen SOM // SEG Technical Program Expanded Abstracts. 2012. P. 1–5. doi:10.1190/segam2012-1494.1.
- Wickramasinghe I., Kalutarage H.** Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation // Soft Computing. 2021. Vol. 25 (3). P. 2277–2293. doi:10.1007/s00500-020-05297-6.
- Zhao T., Jayaram V., Roy A., Marfurt K.J.** A comparison of classification techniques for seismic facies recognition // Interpretation. 2015. Vol. 3 (4). P. SAE29–SAE58. doi:10.1190/INT-2015-0044.1.

REFERENCES

- Alekseev A.S., Pyatkin V.P., Dementiev V.N., Alsynbaev K.S., Angelsky A.N., Bolychevsky A.B., Gilev B.A., Gorbunov B.A., Zabelin V.A., Zyat'kova L.K., Kireitov V.P., Klimenko A.D., Kondauron M.I., Kopylov A.I., G.I. Kuzin, Ladyzhets V.S., Nifontov V.P., Osipov N.A., Raigel' V.I., Repkov V.V., Selivanov A.N., Spektor A.A., Stakheev A.V., Yudin A.N.** Automated image processing of natural complexes of Siberia (In Russ.). Nauka, Novosibirsk, 1988. 222 p.
- Chaban L.N.** Theory and algorithms of pattern recognition: a tutorial (In Russ.). MIIGAIK, Moscow, 2004. 290 p.
- Chopra S., Marfurt K.J.** Unsupervised machine learning applications for seismic facies classification // Proceeding of the 7th Unconventional Resources Technology Conference. Denver, 2019. Paper URTEC-2019-557-MS. doi:10.15530/urtec-2019-557.
- Coleou T., Poupon M., Azbel K.** Unsupervised seismic facies classification: A review and comparison of techniques and implementation // The Leading Edge. 2003. Vol. 22 (10). P. 942–953. doi:10.1190/1.1623635.
- Durand B., Odel P.** Cluster analysis (In Russ.). Moscow, 1977. 128 p.
- Friedman N., Geiger D., Goldszmidt M.** Bayesian network classifiers // Machine Learning. 1997. Vol. 29 (2–3). P. 131–163. doi:10.1023/A:1007465528199.
- Khachumov M.V.** Distances, metrics and cluster analysis // Scientific and Technical Information Processing. 2012. Vol. 39 (6). P. 310–316. doi:10.3103/S0147688212060020.
- Kohonen T.** Self-organized formation of topologically correct feature maps // Biological Cybernetics. 1982. Vol. 43. P. 59–69. doi:10.1007/BF00337288.
- MacQueen J.** Some methods for classification and analysis of multivariate observations // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California, Los Angeles LA USA, 1967. P. 281–297.
- Mitrofanov G.M., Kovalenko I.A., Korytkin E.I.** Application of seismic facies analysis using attributes for reservoir prediction // Russian Journal of Geophysical Technologies. 2024. No. 3. P. 33–45. (In Russ.). doi:10.18303/2619-1563-2024-3-33
- Myasnikov V.V.** Pattern recognition and machine learning. Basic approaches: a tutorial (In Russ.). Samara: Samara University Publishing House, 2023. 196p.
- Nilson N.** Learning machines (In Russ.). Mir, Moscow, 1967. 180 p.
- Owusu B.A., Boateng C.D., Asare V.S., Danuor S.K., Adenutsi C.D., Quaye J.A.** Seismic facies analysis using machine learning techniques: a review and case study // Earth Science Informatics. 2024. Vol. 17. P. 3899–3924. doi:10.1007/s12145-024-01395-3.
- Priezzhev I.I., Solokha E.V., Manral S.** Facies analysis based on 3D seismic waveform (In Russ.) // Geofizika. 2014. No. 1. P. 63–67.
- Rijks E.J.H., Jauffred J.C.E.M.** Attribute extraction: an important application in any detailed 3-D interpretation study // The Leading Edge. 1991. Vol. 10 (9). P. 11–19. doi:10.1190/1.1436837.
- Tu J., Gonzalez R.** Principles of pattern recognition (In Russ.). Mir, Moscow, 1978. 411p.
- Vapnik V.N.** Restoration of dependencies based on empirical data (In Russ.). Nauka, Moscow, 1979. 448 p.
- Verma S., Roy A., Perez R., Marfurt K.J.** Mapping high frackability and high TOC zones in the Barnett Shale: Supervised Probabilistic Neural Network vs. unsupervised multi-attribute Kohonen SOM // SEG Technical Program Expanded Abstracts. 2012. P. 1–5. doi:10.1190/segam2012-1494.1.

Voronina V.V., Mikheev A.V., Yarushkina N.G., Svyatov K.V. Theory and practice of machine learning: a tutorial (In Russ.). UISTU, Ulyanovsk, 2017. 70 p.

Vyugin V.V. Mathematical foundations of the theory of machine learning and forecasting (In Russ.). Moscow, 2013. 387 p.

Wickramasinghe I., Kalutarage H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation // Soft Computing. 2021. Vol. 25 (3). P. 2277–2293. doi:10.1007/s00500-020-05297-6.

Zhao T., Jayaram V., Roy A., Marfurt K.J. A comparison of classification techniques for seismic facies recognition // Interpretation. 2015. Vol. 3 (4). P. SAE29–SAE58. doi:10.1190/INT-2015-0044.1.

ИНФОРМАЦИЯ ОБ АВТОРАХ

КОРЫТКИН Евгений Игоревич – аспирант Института нефтегазовой геологии и геофизики СО РАН, начальник отдела ГРП ООО «СахалинНИПИ нефти и газа». Основные научные интересы: развитие технологий прогноза коллекторских свойств целевых горизонтов.

МИТРОФАНОВ Георгий Михайлович – доктор физико-математических наук, главный научный сотрудник лаборатории динамических проблем сейсмики Института нефтегазовой геологии и геофизики СО РАН. Основные научные интересы: разработка методов и технологий обработки и интерпретации сейсмических данных, решение обратных задач геофизики.

КАМАШЕВ Александр Максимович – младший научный сотрудник лаборатории динамических проблем сейсмики Института нефтегазовой геологии и геофизики СО РАН. Основные научные интересы: разработка алгоритмов и программ машинного обучения и нейронных сетей.

*Статья поступила в редакцию 30 декабря 2024 г.,
одобрена после рецензирования 18 февраля 2025 г.,
принята к публикации 20 февраля 2025 г.*