

Решение задачи динамической интерпретации сейсмических данных при помощи методов машинного обучения

В. Р. Вокина^{1*}, А. С. Авдюков¹, А. А. Лесив¹, И. А. Крупкин¹, А. Н. Емельянов²

¹ООО «Тюменский нефтяной научный центр», Тюмень, Россия

²Тюменский индустриальный университет, Тюмень, Россия

*vrwokina@tnc.rosneft.ru

Аннотация. В статье рассматривается проблема динамической интерпретации сейсмических данных с использованием моделей машинного обучения Extremely Randomized Trees (Extra Trees), Gradient Boosting (GB) и Adaptive Boosting (AdaBoost) в применении к указанной задаче.

В статье проанализированы некоторые существующие решения поставленной задачи. Описано преимущество выбранных моделей машинного обучения и проведены исследования точности по метрике — среднеквадратическое отклонение от истинных значений. В процессе предварительного анализа исследований, проводимых на смежные темы, авторами данной статьи было выявлено, что вопрос динамической интерпретации и предсказания данных с использованием приведенных в статье методов машинного обучения не был освещен, что и стало основным объектом работы.

Далее формализовано применение упомянутых ранее моделей, описаны их особенности и преимущества применимо к решаемой задаче.

Исследованы несколько распространенных методов машинного обучения, позволяющих находить функциональные зависимости между входными параметрами, проведены вычислительные эксперименты для оценки их применимости и сравнительного анализа алгоритмов.

По результатам экспериментов был сделан вывод, что метод Extra Trees в большей мере подходит для практического применения относительно решаемой задачи, поскольку демонстрирует наиболее высокую точность подбора функциональной зависимости и динамической интерпретации.

Ключевые слова: машинное обучение, сейсмика, динамическая интерпретация, Extra Trees, Gradient Boosting, Adaptive Boosting, сейсмические атрибуты, пористость, карты

Для цитирования: Решение задачи динамической интерпретации сейсмических данных при помощи методов машинного обучения / В. Р. Вокина, А. С. Авдюков, А. А. Лесив [и др.]. – DOI 10.31660/0445-0108-2024-5-117-131 // Известия высших учебных заведений. Нефть и газ. – 2024. – № 5. – С. 117–131.

The solution of the task of dynamic interpretation of seismic data using machine learning methods

Victoria R. Vokina^{1*}, Alexey S. Avdyukov¹, Anastasia A. Lesiv¹, Igor A. Krupkin¹,
Andrey N. Emelyanov²

¹*Tyumen Petroleum Research Center LLC, Tyumen, Russia*

²*Industrial University of Tyumen, Tyumen, Russia*

**vrvokina@tnc.rosneft.ru*

Abstract: This article examines the problem of dynamically interpreting seismic data using machine learning models, which include Extremely Randomized Trees (Extra Trees), Gradient Boosting (GB), and Adaptive Boosting (AdaBoost) for the given problem.

The study analyzes some existing solutions of the problem and describes the advantages of these machine learning models. Accuracy is estimated using the root mean square error metric.

The authors found that dynamic interpretation and prediction of seismic data using these machine learning methods had not been extensively explored in research on related topics, which became the main focus of the study. The article formalizes the use of the mentioned models and highlights features and advantages for the given problem.

Several common machine learning methods were investigated to find functional relationships between input parameters. Computational experiments were conducted to evaluate their applicability and compare the algorithms.

The results show that the Extra Trees method is the most suitable for practical use for the given problem, as it demonstrates the highest accuracy in determining functional relationships and dynamic interpretation.

Keywords: machine learning, seismic data, dynamic interpretation, Extra Trees, Gradient Boosting, Adaptive Boosting, seismic attributes, porosity, maps

For citation: Vokina, V. R., Avdyukov, A. S., Lesiv, A. A., Krupkin, I. A., & Emelyanov, A. N. (2024). The solution of the task of dynamic interpretation of seismic data using machine-learning methods. *Oil and Gas Studies*, (5), pp. 117-131. (In Russian). DOI: 10.31660/0445-0108-2024-5-117-131

Введение

Проблема динамической интерпретации сейсмических данных относится к нетривиальным задачам цифровой обработки информации. В процессе обработки сейсмических данных пытаются выделить полезные типы волн и сохранить пропорциональность амплитуд и коэффициентов отражений. Кроме того, поиск более мелких и сложных ловушек, содержащих углеводороды, является востребованным направлением при разработке месторождений. Справиться с этими задачами помогает сейсморазведка и качественная обработка полученных данных.

Определение прогнозных значений фильтрационно-емкостных свойств (ФЕС) для коллекторов, содержащих в себе пластовые флюиды разного компонентного состава, является непростым технологическим процессом. Согласно классификации, приведенной в работе [1], прогноз сейсмических данных можно разделить на три этапа, первый из которых заключается в извлечении атрибутов сейсмической записи. На втором этапе производится поиск корреляционных связей между извлеченными параметрами. На заключительном этапе по найденным зависимостям прово-

дится построение и анализ карт, показавших лучшие статистические связи с искомыми геологическими параметрами.

В процессе интерпретации решаются две задачи.

1. Обратная кинематическая задача — определение границ геологических тел и скоростей по кинематическим параметрам волн.

2. Обратная динамическая задача — определение свойств пород.

После завершения трех этапов прогнозирования фильтрационно-емкостных свойств получается некоторый набор выходной информации.

Причем выходные данные имеют достоверность, пропорциональную соответствию свойств геологической среды и построенной математической модели, принимаемой при разработке конкретной динамической интерпретационной технологии.

Также в данной предметной области существует проблема отсутствия единственности решения при поиске взаимосвязи между сейсмическими данными и свойствами горных пород, поэтому предсказание петрофизических свойств на основе сейсмических данных — сложная, нетривиальная задача. В некоторых исследованиях предпринимались попытки определить искомую взаимосвязь с использованием контролируемых методов машинного обучения, но сталкивались с определенным рода проблемами [2]. В связи с этим актуальным является поиск других, альтернативных методов обработки и предсказания петрофизических свойств по данным сейсморазведки.

Похожие исследования по интерпретации сейсмических данных проводятся и другими учеными. В статьях поднимаются проблемы высокой вычислительной стоимости обучения, недостатка маркированных данных для обучения, сложности интерпретации результатов моделей.

Например, автоматизированная интерпретация сейсмических разломов являлась областью исследований для авторов статьи [3]. Авторы данной статьи отмечают появление относительно новых методов интерпретации сейсмических данных, основанных на глубоком обучении, которое показало многообещающие результаты. Однако на сегодняшний день эти методы не были разумно обобщены.

В некоторых других работах описываются используемые методы машинного и глубокого обучения, такие как случайный лес, опорные вектора, градиентный бустинг. В результате обученные модели достигали точности выше 85 %, что демонстрирует способность методов описывать сложно зависимые сейсмические данные [4, 5].

Точность и качество цифровой обработки информации играют неотъемлемую роль при дальнейшем моделировании физических процессов в пласте, поэтому важно подобрать такой метод машинного обучения, который будет иметь минимальную погрешность и позволит строить реальные карты распределений петрофизических величин.

На основании вышеизложенного целью данной работы является исследование применимости выбранных ансамблевых методов машинного обучения для решения задачи динамической интерпретации сейсмических данных на примере построения прогнозных распределений петрофизического параметра в определенном регионе.

Материалы и методы

В качестве исследуемого петрофизического параметра была выбрана пористость горных пород. Построение распределения общей пористости — карта, к которой стремились прийти авторы данной статьи.

Для проведения независимой оценки полученных результатов были запрошены 5 наборов некоторых модельных сейсмических данных — $Map_i(x, y)$, определяемых своими значениями на регулярной прямоугольной сетке. Каждая из карт содержит в себе две пространственные координаты и один из атрибутов волнового поля, рассчитанных для целевого интервала. Под атрибутами в нашем случае понимаются амплитуды и их производные, частоты и их производные, а также карта упругих характеристик.

Все массивы исходных данных контролируются набором неизвестных взаимосвязанных функций $F_j(x, y)$, из которых в нескольких точках (необязательно совпадающих с узлами сетки) заданы значения только одной функции $F_1(x, y)$. Исходные данные имеют между собой неизвестные корреляционные зависимости и образуют общую карту, позволяющую построить распределение пористости на основании сейсмических данных. На начальном этапе не было известно в каких зависимостях находятся известные карты и какая функция позволяет получить искомое распределение. Схематическая интерпретация поставленной задачи представлена на рисунке 1.

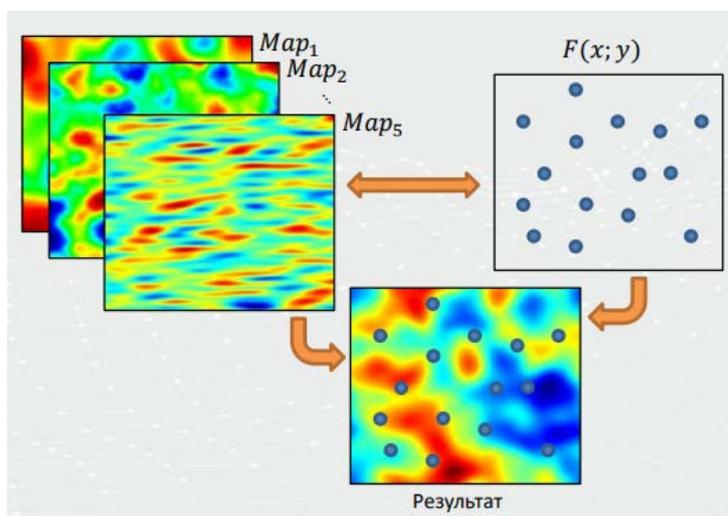


Рис. 1. Графическая интерпретация задачи

Исследование исходных данных установило, что каждый из 5 $Map_i(x, y)$ представляет собой сеточную функцию, узлы которой располагаются на одинаковом расстоянии относительно друг друга.

При отображении исходных данных на графике получается некоторое распределение значений, представленное на рисунке 2. На графике можно наблюдать выбросы, отмеченные красными точками и не попадающие на сеточную функцию. Для проведения этапов машинного обучения данные точки были исключены из массивов исходных данных.

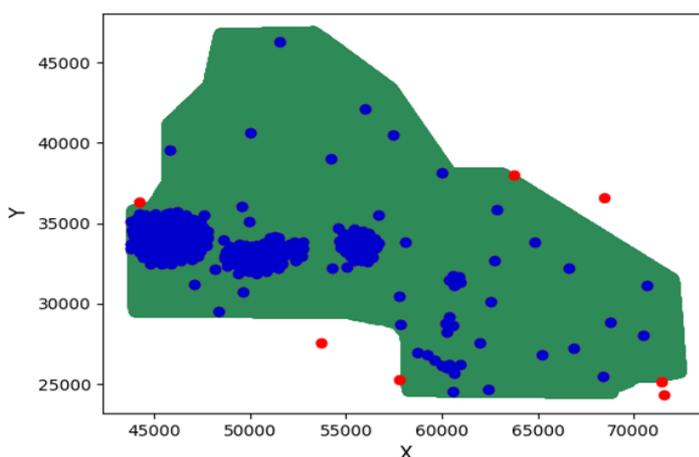


Рис. 2. Полигон значений

После проведения анализа и оценки данных была выбрана представленная рисунке 3 архитектура проекта.

На вход подается пять наборов с данными в виде $Map_i(x, y)$, где $i = 1, \dots, 5$. Далее для каждого набора с узлами строится своя модель обучения с целью поиска неизвестных функций $F_j(x, y)$ для каждой сетки. Затем полученные значения Z_i подаются в результирующую модель для обучения и нахождения целевой Z на основе ранее предсказанных величин (см. рис. 3).

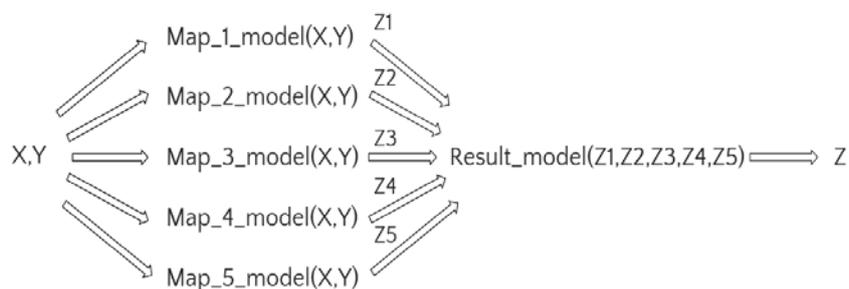


Рис. 3. Архитектура поиска неизвестной величины Z

Для реализации архитектуры были выбраны и протестированы три ансамблевых метода машинного обучения. Суть таких методов заключается в объединении нескольких базовых моделей в одну мощную для увеличения точности результата в сравнении с оценкой простой модели [6]. Такие методы устойчивы к шуму и выбросам в данных [7], почти не подвержены переобучению [8] и за счет усреднения прогнозов уменьшают дисперсию и смещение результата [9].

Исследуемые методы.

1. Адаптивный бустинг (Adaptive Boosting) [10] — одна из моделей бустинга [11], используемая для решения задач регрессии и классификации. Идея бустинга заключается в построении ансамбля базовых моделей, где каждая последующая модель стремится исправить ошибки предыдущих. Адаптация же состоит в использовании результатов предсказания ранее созданных базовых моделей при построении следующих моделей. В начале результаты предыдущих моделей корректируются за счет назначения больших весов данным, для которых был получен некорректный результат. Затем каждая последующая модель обучается на заново взвешенных данных от предыдущей модели. Таким образом следующая стремится уменьшить ошибку текущего ансамбля. Весь принцип работы для задачи регрессии математически описывается следующим образом (формула 1):

$$F(x) = \sum a_i \times h_i(x), \quad (1)$$

где a_i — вес i -ой простой модели в ансамбле, вычисляемый на каждом шаге по формуле $a_i = \frac{1}{2} \ln\left(\frac{1-error}{error}\right)$, где $error$ — ошибка i -ой простой модели на взвешенной выборке; $h_i(x)$ — предсказание i -ой простой модели в ансамбле.

2. Градиентный бустинг (Gradient Boosting) [12], в основе метода все тот же бустинг, представляющий собой последовательную линейную комбинацию базовых моделей. В отличие от адаптивного бустинга, данный метод при пересчете функции потерь не присваивает неправильно предсказанным данным больший вес. Вместо этого он оптимизирует функцию потерь через поправку, которую вносит следующая базовая модель в композицию. Таким образом метод пытается сразу генерировать точные результаты, а не исправлять ошибки. Математически метод описывается следующим образом (формула 2):

$$a_N(x) = \sum_{m=1}^N \gamma_m b_m(x), \quad (2)$$

$$b_m(x) = \operatorname{argmin} \sum_{i=1}^l (b(x_i) - s_i)^2;$$

$$\gamma_m = \operatorname{argmin} \sum_{i=1}^l L(y_i, a_{m-1}(x_i) + \gamma b_m(x_i));$$

где $b_m(x)$ — новый базовый алгоритм; $a_N(x)$ — ансамбль из N алгоритмов; γ_m — коэффициент базового алгоритма; $L(y_i, a(x_i))$ — функция потерь, в нашем случае рассчитываемая по формуле $L(y_i, a(x_i)) = (y_i - a(x_i))^2$; s_i — антиградиент квадратичной функции потерь, рассчитываемый по формуле $s_i = 2(y_i - a(x_i))$.

3. Сверхслучайные деревья (Extremely Randomized Trees) [13]. Представляет собой ансамбль решающих деревьев и является модификацией алгоритма случайный лес (Random Forest) [14]. Данный метод отличается от случайного леса способом разбиения в вершинах деревьев. В классическом случайном лесе при построении каждого узла дерева происходит случайный выбор подмножества признаков, затем для каждого признака из этого подмножества вычисляются значения критерия информативности и выбирается тот признак и порог для него, при котором критерий достигает максимума. В сверхслучайных деревьях также производится случайный выбор подмножества признаков, но для каждого признака пороговое значение выбирается случайным образом, а не оптимизируется. Это позволяет уменьшить вычислительную сложность алгоритма и увеличить его скорость работы. После построения всех деревьев в ансамбле прогноз для нового объекта в задаче регрессии строится путем усреднения прогнозов отдельных деревьев. Принцип работы метода для задачи регрессии (формула 3) описывается как

$$f(X) = \frac{1}{N} \sum_{i=1}^N f_i(X), \quad (3)$$

где

$$f_i(X) = \begin{cases} c_i, X \in i \\ f_i^1, X_m < t_i; \\ f_i^2, X_m \geq t_i \end{cases}$$

$f_i(X)$ — функция, представляющая i -е дерево; c_i — среднее значение целевой переменной для всех объектов в листовом узле i ; X_m — m -й признак объекта X ; t_i — пороговое значение в узле i ; f_i^1 и f_i^2 — функции, представляющие поддеревья, образованные после деления узла i .

Для реализации выбранных моделей машинного обучения были использованы классы, представленные в библиотеке Scikit-learn [15].

Экспериментальное исследование

Эксперименты были проведены на полученных 5 наборах с данными. Проведение исследования осуществлялось на огромном количестве данных. Каждая таблица $Mar_i(x, y)$ содержала около 1,2 млн строк с данными.

Каждый набор был разделен на тренировочную и тестовую выборки в соотношении, где тренировочная выборка составила 70 %. В качестве метрики для сравнения рассматриваемых алгоритмов машинного обучения

используется корень из среднеквадратичной ошибки — Root Mean Squared Error (RMSE) (формула 4) [16]. Данная метрика используется для случаев, когда есть необходимость выявить большие ошибки и выбрать модель, которая дает наиболее точный результат. Формула для вычисления метрики:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4)$$

где y_i — истинное значение величины; \hat{y}_i — расчетное значение величины; N — количество данных.

У каждого алгоритма существуют свои гиперпараметры, изменение/подбор которых позволяет добиться лучшего результата по метрике качества обучения. Подбор оптимальных гиперпараметров происходил с помощью инструмента GridSearchCV [17]. Параметры модели с наименьшей оценкой RMSE считались наилучшими. Как видно из таблицы, по результатам экспериментов лучшей моделью оказалась Extra Trees продемонстрировав результат 1,168 % на тестовом наборе данных.

Гиперпараметры и результаты получившихся моделей

Название алгоритма	Гиперпараметр	RMSE, %
Extremely Randomized Trees	n_estimators = 300 max_features = 1 min_samples_split = 2 min_samples_leaf = 1	1,168
Gradient Boosting	n_estimators = 200 learning_rate = 1.0 max_depth = 5	1,219
Adaptive Boosting	n_estimators = 100 learning_rate = 1,0 loss = exponential	1,342

Рассмотрим каждый гиперпараметр из таблицы подробнее:

- n_estimators — число деревьев в лесу. Чем больше деревьев, тем лучше качество модели, но требует больших мощностей для построения;
- max_features — число признаков при поиске лучшего разделения;
- min_samples_split — минимальное число объектов, необходимое для того, чтобы узел дерева мог расщепиться. Этот параметр тесно связан с максимальной глубиной дерева. При слишком высоком значении этого параметра глубина дерева не может увеличиться из-за невозможности дальнейшего деления;
- min_samples_leaf — минимальное число объектов в листьях. Чем меньше объектов в листьях, тем точнее, но сложнее модель;

- `learning_rate` — вес, применяемый к каждому базовому алгоритму. Чем выше значение, тем больше вклад каждого регрессора;
- `max_depth` — максимальная глубина дерева. Чем глубже дерево, тем лучше модель. Параметр должен быть подобран таким образом, чтобы все данные были распределены по дереву с минимально доступным разделением по листьям;
- `loss` — функция потерь, используемая для обновления весов после каждой итерации бустинга.

Исследование показало, что модель на основе адаптивного бустинга проявила себя хуже остальных, объяснить это можно подобранной через `GridSearchCV` экспоненциальной функцией потерь. Она измеряет вероятность правильного предсказания класса истинного значения. Это может быть эффективно для задач классификации, но менее подходящим для задач регрессии, где требуется предсказывать непрерывные значения. Модель на основе сверхслучайных деревьев наиболее оптимальна из-за случайного выбора пороговых значений для разделения узлов, что делает обучение менее чувствительным к шуму в данных [18]. Кроме того, алгоритм лучше справляется с нелинейными зависимостями за счет большого количества деревьев, тем самым охватывая большее количество условий.

Результаты и обсуждение

В ходе исследования решена задача динамической интерпретации сейсмических данных и построения прогнозной модели карты общей пористости с учетом особенностей задачи. Проведены вычислительные эксперименты для выявления наиболее подходящего метода. Спрогнозированные значения были сопоставлены с эталонными реальными данными.

Полученные в ходе вычислительных экспериментов результаты интерпретированы следующим образом.

1. Модель `Adaptive Boosting` имеет худший показатель метрики из всех, представленных в данной статье. Данный метод восприимчив к шумам в данных, соответственно, склонен к переобучению [19]. Кроме того, причиной низкого результата может быть неправильно подобранная функция потерь. Экспоненциальная функция потерь не учитывает различия в отклонениях и может быть менее устойчивой к выбросам.

2. Модель `Gradient Boosting` имеет значительно лучший показатель метрики, чем `Adaptive Boosting`. В основе градиентного бустинга заложена более сложная стратегия обучения — стохастический градиентный спуск, который позволяет более эффективно обучаться на больших объемах данных [20]. Также метод использует более сложные модели, такие как глубокие деревья решений, что позволяет лучше аппроксимировать сложные зависимости в данных.

3. Модель `Extremely Randomized Trees` показала себя наилучшим образом. Представленная модель менее чувствительна к шумам в данных

из-за случайности выбора значений в узлах. Более того, большое количество деревьев позволяет методу работать с нелинейными зависимостями и охватывать большее количество условий для нахождения аппроксимирующей функции.

4. Использование методов машинного обучения может успешно решать задачи динамической интерпретации сейсмических данных и построения карт петрофизических свойств, в нашем случае была получена карта общей пористости, практически идентичная реальной карте. Согласно метрике $RMSE = 1,168 \%$, это говорит о том, что прогнозное распределение пористости и распределение данного параметра на проверочной выборке практически совпадает.

В результате проведенного исследования Extremely Randomized Trees имеет наилучший показатель по выбранной метрике. На основании полученных данных была построена искомая карта распределения пористости. Сопоставление полученной при помощи машинного обучения карты с эталонной картой представлено на рисунке 4.

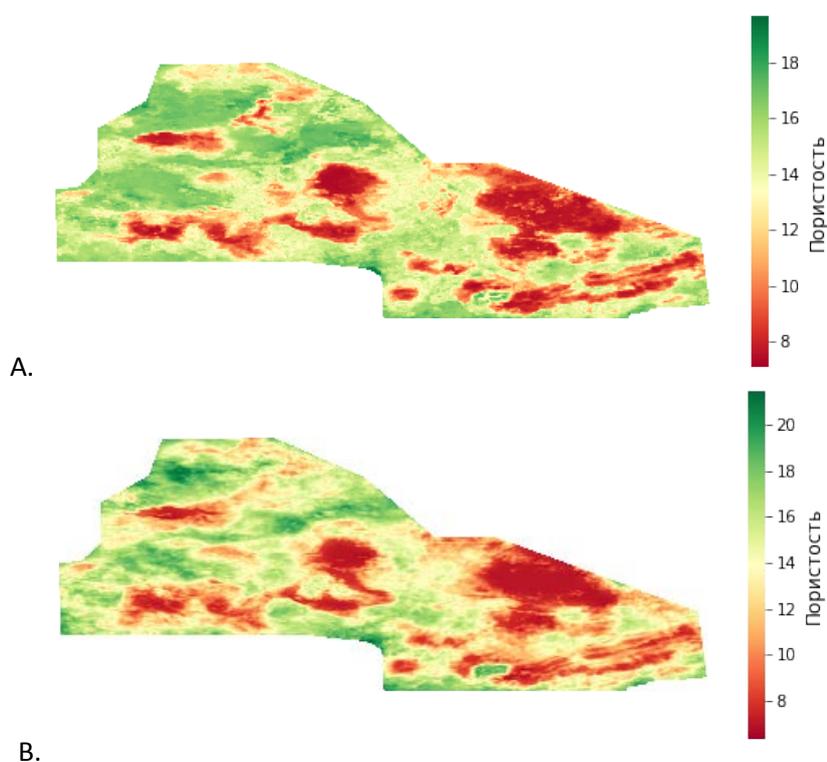


Рис. 4. Итоговое распределение пористости:
А — Extremely Randomized Trees, В — эталонная карта

Можно заметить, что полученное нами распределение с достаточно высокой точностью повторяет эталонное распределение выбранного петрофизического параметра — пористости.

Выводы

Исследование и развитие моделей и методов, предназначенных для решения задачи динамической интерпретации сейсмических данных, а также построение прогнозных карт, характеризующих петрофизические свойства горных пород, представляют практический и научно-исследовательский интерес и являются актуальными и в настоящее время.

В процессе анализа исходных данных авторы статьи исключили все не попадающие в контур сейсмических данных значения, однако проведение кластеризации не посчитали необходимым. Предположительно данная процедура поспособствует повышению точности расчета и улучшению показателя метрики. В соответствии с приведенным заключением поставлена цель в дополнительном исследовании методики и моделей машинного обучения, позволяющих повысить точность проведенного исследования.

По результатам анализа предметной области были рассмотрены несколько методов машинного обучения, позволяющих производить прогноз указанного параметра. По итогам вычислительных экспериментов авторами сделаны выводы, что из рассмотренных моделей машинного обучения метод Extremely Randomized Trees в большей мере подходит для практического применения в связи с тем, что имеет лучшие показатели по метрике, позволяющей оценить точность решения.

Кроме того, на основании построенного прогнозного распределения петрофизического параметра в определенном регионе и сопоставления полученного результата с эталонными значениями можно сделать вывод, что выбранные ансамблевые методы машинного обучения применимы для решения задачи динамической интерпретации сейсмических данных, в частности метод Extremely Randomized Trees, имеющий данные по метрике $RMSE = 1,168 \%$.

Дальнейшие перспективы решения задачи динамической интерпретации и построения прогнозных карт петрофизических свойств заключаются в более детальном изучении исходных данных, проведении кластеризации и рассмотрении другого набора моделей машинного обучения с целью повышения расчетной и прогнозной точности.

Список источников

1. Аmani, М. М. М. Применение методов геостатистики в факторно-регрессионном прогнозировании пористости коллекторов по сейсмическим атрибутам / М. М. М. Атамани. – Текст : непосредственный // Молодежь и современные информационные технологии : сборник трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых, 3–7 декабря 2018 года, Томск. – Томск : Томский политехнический университет, 2019. – С. 156–157.

2. Tarantola, A. Inverse problem theory and methods for model parameter estimation / A. Tarantola. – DOI 10.1137/1.9780898717921. – Direct text. – Philadelphia : Society for industrial and applied mathematics, 2005. – 342 p.
3. Successful leveraging of image processing and machine learning in seismic structural interpretation : A review / Z. Wang, H. Di, M. A. Shafiq [et al.]. – DOI 10.1190/tle37060451.1. – Direct text // The Leading Edge. – 2018. – Vol. 37, Issue 6. – P. 451–461.
4. Obinnaya Chikezie Victor, N. Automated Seismic Interpretation : Machine Learning Technologies are Being used to Develop Automated Seismic Interpretation to Identify Geological Features, Such as Faults and Stratigraphic Horizons / N. Obinnaya Chikezie Victor, L. Oghenechodja Daniel. – DOI 10.51483/IJAIML.3.2.2023.74-98. – Direct text // International Journal of Artificial Intelligence and Machine Learning. – 2023. – Vol. 3, Issue 2. – P. 74–98.
5. Integration of Multiple Bayesian Optimized Machine Learning Techniques and Conventional Well Logs for Accurate Prediction of Porosity in Carbonate Reservoirs / S. Alatefi, R. Abdel Azim, A. Alkouh, G. Hamada. – DOI 10.3390/pr11051339. – Direct text // Processes. – 2023. – Vol. 11, Issue 5. – P. 1339–1361.
6. Dietterich, T. G. Ensemble Methods in Machine Learning / T. G. Dietterich. – DOI 10.1007/3-540-45014-9_1. – Direct text // International Workshop on Multiple Classifier Systems. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2000. – P. 1–15.
7. Polikar, R. Ensemble based systems in decision making / R. Polikar. – DOI 10.1109/MCAS.2006.1688199. – Direct text // IEEE Circuits and Systems Magazine. – 2006. – Vol. 6, Issue 3. – P. 21–45.
8. Opitz, D. Popular ensemble methods : An empirical study / D. Opitz, R. Maclin. – DOI 10.1613/jair.614. – Direct text // Journal of Artificial Intelligence Research. – 1999. – Vol. 11. – P. 169–198.
9. Bühlmann, P. Boosting with the L_2 Loss : regression and classification / P. Bühlmann, B. Yu. – DOI 10.1198/016214503000125. – Direct text // Journal of the American Statistical Association. – 2003. – Vol. 98, Issue 462. – P. 324–339.
10. Freund, Y. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting / Y. Freund, R. E. Schapire. – DOI 10.1006/jcss.1997.1504. – Direct text // Journal of Computer and System Sciences. – 1997. – Vol. 55, Issue 1. – P. 119–139.
11. CatBoost : unbiased boosting with categorical features / L. Prokhorenkova, G. Gusev, A. Vorobev [et al.]. – DOI 10.48550/arXiv.1706.09516. – Direct text // In Advances in Neural Information Processing Systems. – 2018. – Vol. 31. – P. 6638–6648.
12. Friedman, J. H. Greedy Function Approximation : A Gradient Boosting Machine / J. H. Friedman. – DOI 10.1214/aos/1013203451. – Direct text // The Annals of Statistics. – 2001. – Vol. 1. – P. 1189–1235.
13. Geurts, P. Extremely randomized trees / P. Geurts, D. Ernst, L. Wehenkel. – DOI 10.1007/s10994-006-6226-1. – Direct text // Machine learning. – 2006. – Vol. 63. – P. 3–42.
14. Breiman, L. Random Forests / L. Breiman. – DOI 10.1023/A:1010933404324. – Direct text // Machine Learning. – 2001. – Vol. 45. – P. 5–32.
15. Scikit-learn : Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort [et al.]. – DOI 10.5555/1953048.2078195. – Direct text // Journal of Machine Learning research. – 2011. – Vol. 12. – P. 2825–2830.

16. The Elements of Statistical Learning : Data Mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman. – Direct text // New York : Springer. – 2009. – Vol. 2. – P. 1–758.

17. GridSearchCV. – Text : electronic // Scikit-learn : website. – URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (accessed: 06.02.2024).

18. Biau, G. A random forest guided tour / G. Biau, E. Scornet. – DOI 10.1007/s11749-016-0481-7. – Direct text // Test. – 2016. – Vol. 25. – P. 197–227.

19. Long, P. M. Random classification noise defeats all convex potential boosters / P. M. Long, R. A. Servedio. – DOI 10.1007/s10994-009-5165-zs. – Direct text // Mach Learn. – 2010. – Vol. 78. – P. 287–304.

20. Natekin, A. Gradient boosting machines, a tutorial / A. Natekin, A. Knoll. – Text : electronic // Frontiers in Neurorobotics. – 2013. – Vol. 7, Issue 21. – URL: <https://doi.org/10.3389/fnbot.2013.00021>.

References

1. Amani, M. M. M. (2019). Primenenie metodov geostatistiki v faktorno-regressionnom prognozirovanii poristosti kollektorov po seysmicheskim atributam. Molodezh' i sovremennye informatsionnye tekhnologii: sbornik trudov XVI Mezhdunarodnoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh. Tomsk, TPU Publ., pp. 156-157. (In Russian).

2. Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. Philadelphia, Society for industrial and applied mathematics, 342 p. (In English). DOI: 10.1137/1.9780898717921

3. Wang, Z., Di, H., Shafiq, M. A., Alaudah, Y., & AlRegib, G. (2018). Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. The Leading Edge, 37(6), pp. 451-461. (In English). DOI: 10.1190/tle37060451.1

4. Obinnaya Chikezie Victor, N., & Oghenechodja Daniel, L. (2023). Automated Seismic Interpretation: Machine Learning Technologies are Being Used to Develop Automated Seismic Interpretation to Identify Geological Features, Such as Faults and Stratigraphic Horizons, 3(2), pp. 74-98. (In English). DOI: 10.51483/IJAIML.3.2.2023.74-98

5. Alatefi, S., Abdel Azim, R., Alkouh, A., & Hamada, G. (2023). Integration of multiple bayesian optimized machine learning techniques and conventional well logs for accurate prediction of porosity in carbonate reservoirs. Processes, 11(5), pp.1339-1361. (In English). DOI: 10.3390/pr11051339

6. Dietterich, T. G. (2000). Ensemble methods in machine learning. In International workshop on multiple classifier systems, pp. 1-15. Berlin, Heidelberg, Springer Berlin Heidelberg. (In English). DOI: 10.1007/3-540-45014-9_1

7. Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and systems magazine, 6(3), pp. 21-45. (In English). DOI: 10.1109/MCAS.2006.1688199

8. Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. Journal of artificial intelligence research, 11, pp.169-198. (In English). DOI: 10.1613/jair.614

9. Bühlmann, P., & Yu, B. (2003). Boosting with the L₂ loss: regression and classification. *Journal of the American Statistical Association*, 98(462), pp. 324-339. (In English). DOI: 10.1198/016214503000125
10. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), pp. 119-139. (In English). DOI: 10.1006/jcss.1997.1504
11. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, pp. 6638-6648. (In English). DOI: 10.48550/arXiv.1706.09516
12. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 1, pp. 1189-1235. (In English). DOI: 10.1214/aos/1013203451
13. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees // *Machine learning*, 63, pp. 3-42. (In English). DOI: 10.1007/s10994-006-6226-1
14. Breiman, L. (2001). Random Forests. *Machine Learning*, 45, pp. 5-32. (In English). DOI: 10.1023/A:1010933404324
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 12, pp. 2825-2830. (In English). DOI: 10.5555/1953048.2078195
16. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer, 2, pp. 1-758. (In English).
17. GridSearchCV. Scikit-learn. (In English). Available at: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
18. Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, pp. 197-227. (In English). DOI: 10.1007/s11749-016-0481-7
19. Long, P. M., & Servedio, R. A. (2010). Random classification noise defeats all convex potential boosters. *Mach Learn*, 78, pp. 287-304. (In English). DOI: 10.1007/s10994-009-5165-zs
20. Natekin, A., & Knoll, A. (2013). Gradient Boosting Machines, a Tutorial. *Frontiers in Neurorobotics*, 7(21). (In English). DOI: 10.3389/fnbot.2013.00021

Информация об авторах / Information about the authors

<p>Вокина Виктория Руслановна, <i>специалист, Управление развития интеллектуальных технологий, ООО «Тюменский нефтяной научный центр», г. Тюмень, магистрант, vrvokina@tmmc.rosneft.ru, ORCID: https://orcid.org/0000-0002-9651-1758</i></p>	<p>Victoria R. Vokina, <i>specialist, Intelligent Systems Development Department, Tyumen Petroleum Research Center LLC, Master Student, vrvokina@tmmc.rosneft.ru ; ORCID: https://orcid.org/0000-0002-9651-1758</i></p>
--	--

Авдюков Алексей Сергеевич, специалист, Управление развития интеллектуальных технологий, ООО «Тюменский нефтяной научный центр», г. Тюмень, магистрант, ORCID: <https://orcid.org/0009-0009-5125-7379>

Лесив Анастасия Александровна, специалист, Управление развития интеллектуальных технологий, ООО «Тюменский нефтяной научный центр», г. Тюмень, магистрант, ORCID: <https://orcid.org/0009-0007-6897-488X>

Крупкин Игорь Андреевич, специалист, Управление развития интеллектуальных технологий, ООО «Тюменский нефтяной научный центр», г. Тюмень, магистрант, ORCID: <https://orcid.org/0009-0003-9482-929X>

Емельянов Андрей Николаевич, доцент, Тюменский индустриальный университет, ВИШ EG, базовая кафедра ООО «Тюменский нефтяной научный центр», г. Тюмень, ORCID: <https://orcid.org/0009-0008-4153-6174>

Alexey S. Avdyukov, specialist, Intelligent Systems Development Department, Tyumen Petroleum Research Center LLC; Master Student, ORCID: <https://orcid.org/0009-0009-5125-7379>

Anastasia A. Lesiv, specialist, Intelligent Systems Development Department, Tyumen Petroleum Research Center LLC; Master Student, ORCID: <https://orcid.org/0009-0007-6897-488X>

Igor A. Krupkin, specialist, Intelligent Systems Development Department, Tyumen Petroleum Research Center LLC, Master Student, ORCID: <https://orcid.org/0009-0003-9482-929X>

Andrey N. Emelyanov, Associate Professor, Industrial University of Tyumen, EG HES, Basic Department of Tyumen Petroleum Research Center LLC, ORCID: <https://orcid.org/0009-0008-4153-6174>

Статья поступила в редакцию 05.06.2024; одобрена после рецензирования 12.07.2024; принята к публикации 22.07.2024.

The article was submitted 05.06.2024; approved after reviewing 12.07.2024; accepted for publication 22.07.2024.