Научная статья
Original article
УДК 550.8.056
DOI: 10.24412/2658-4964-2021-103696

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАЗБИЕНИЯ ГЕОФИЗИЧЕСКИХ КРИВЫХ НА ИНТЕРВАЛЫ

POSSIBILITIES OF USING MACHINE LEARNING METHODS FOR SOLVING THE PROBLEM OF DIVIDING GEOPHYSICAL CURVES INTO INTERVALS

Бобылев Кирилл Дмитриевич, Ведущий специалист ООО «Тюменский Нефтяной Научный Центр», Россия, г.Тюмень

Bobylev Kirill Dmitrievich, Leading Specialist LLC "Tyumen Oil Scientific Center", Russia, Tyumen

Аннотация: Задача разбиения геофизических кривых на интервалы является одной из составных частей петрофизического цикла обработки информации. Корректное выполнение данного этапа позволяет с большей точностью оценить геологический объем запасов изучаемого объекта (пласта). Данная задача, в силу отсутствия строгой формализации, может занимать определенное количество времени и ручного труда специалиста-петрофизика. Повышение производительности труда в контексте решения задачи разбиения на пропластки возможно путем привлечения алгоритмов машинного обучения. Переформулирование задачи с выведением основных параметров,

определяющих итоговый вариант разбиения, позволяет получить концептуально новый и стабильный алгоритм, применительно к решению данной проблемы. Предлагаемые для определения пользователю параметры отражают физические ограничения существующей аппаратуры записи геофизических кривых. Тестирование алгоритма проведено на синтетических и реальных данных. Результаты разбиения, выполненного решающим деревом, согласуются с разметкой выполненной экспертом. Выдвинуты предположения относительно особенностей работы решающего дерева в контексте разбиения задачи разбивки на пропластки. Сформулированы идеи для повышения устойчивости алгоритма. Предложенный альтернативный подход, основанный на алгоритме решающего дерева, прост в реализации и позволяет эффективно осуществлять разбиение геофизических кривых для последующего осреднения.

Abstract: The problem of dividing geophysical curves into intervals is one of the components of the petrophysical information processing cycle. Correct execution of this stage allows to estimate with greater accuracy the geological volume of reserves of the studied object (formation). This task, due to the lack of strict formalization, can take a certain amount of time and manual labor of a petrophysicist. Increasing labor productivity in the context of solving the problem of splitting into layers by using machine learning algorithms. Reformulation of the problem with the derivation of the main parameters that determine the final version of the partition, allows you to get a conceptually new and stable algorithm in relation to the solution of this problem. The parameters proposed for determination by the user reflect the physical limitations of the existing equipment for recording geophysical curves. The algorithm was tested on synthetic and real data. The results of the partitioning performed by the decision tree are consistent with the markup performed by the expert. Assumptions are made regarding the peculiarities of the decision tree operation in the context of splitting the partitioning problem into layers. Ideas for improving the stability of the algorithm are formulated. The proposed

alternative approach based on the decision tree algorithm is easy to implement and allows you to efficiently split geophysical curves for subsequent averaging.

Ключевые слова: петрофизика, машинное обучение, подсчет запасов, попластовая интерпретация

Key words: petrophysics, machine learning, calculation of reserves, per-layer interpretation

Введение

Современный подход к оценке запасов месторождений включает в себя множество этапов, различающихся по сложности и трудозатратам. Проведение сейсморазведочных работ, бурение скважин с извлечением керна для последующего анализа, построение петрофизической и геологической моделей – основные составляющие данной задачи. Одним из наиболее важных этапов является этап петрофизического моделирования, на котором аккумулируется большая часть ошибок финальной оценки.

Петрофизический этап, в свою очередь, состоящий также из множества подзадач, в качестве выходных параметров определяет осредненные, попластовые, значения пористости, нефтегазонасыщенности и проницаемости [1]. Корректное определение этих параметров, в частности, напрямую зависит от того насколько качественно было выполнено разделение на интервалы (пропластки).

Решение данной задачи в определенном проценте случаев выполняется профильным специалистом вручную (либо в полуавтоматическом режиме). Эта особенность не может не вносить дополнительную погрешность, связанную с известной долей субъективности, свойственной данной задаче, в виду слабой ее формализации. Также следует учитывать особенность регистрации геофизических кривых – разрешающую способность методов. Тот факт, что некоторые геофизические приборы, исходя из конструктивных особенностей, не могут однозначно регистрировать сигнал от пропластков

менее определенной мощности, накладывает дополнительное ограничение на задачу разбиения на интервалы [2].

Резюмируя вышесказанное: задача разбиение геофизической кривой на пропластки является слабоформализованной задачей с дополнительными ограничениями. Решение данной задачи возможно путем применения специальных методов обработки сигнала, коим, по сути, и являются регистрируемые геофизические кривые, однако самостоятельная настройка и, тем более, реализация такого рода алгоритма затруднительна без наличия опыта решения подобного рода задач. Как следствие, возникает необходимость в поиске алгоритма, доступного вне зависимости от используемого программного обеспечения и уровня компетенций специалиста в смежных областях. В общем случае, в схожих ситуациях методы машинного обучения зарекомендовали себя как надежные инструменты, позволяющие, в определенном смысле, решить задачу «в лоб».

Постановка задачи

Попытаемся формализовать постановку задачи. Пусть нам задана кривая **X**, представленная шкалой глубин $\mathbf{D = \{d_i\}_{i=1...N}}$ с постоянным шагом **s** и значениями кривой для каждой глубины $\mathbf{\{X_i\} = \{X(d_i)\}}$. Нам необходимо построить разбиение шкалы глубин **D**, а именно найти вектор индексов $\mathbf{K = \{k_j\}_{j=1...M}}$ такие, чтобы соблюдались условия:

- $\min\left(d_{k_{j+1}} - d_{k_j}\right) \geq w_{min}$
- $\sum_{j=0}^{M-1} \frac{d_{k_{j+1}} - d_{k_j}}{M-1} \approx w_{avg}$
- $d_{k_1} = d_1$
- $d_{k_M} = d_N$

где w_{min}, w_{avg} – минимальная и средняя толщина прослоев.

Существует множество способов построить такое разбиение, предлагаемый подход основан на применении алгоритма DecisionTreeRegressor из популярной библиотеки машинного обучения sklearn [5]. Применение этого алгоритма позволяет строить аппроксимации кривой **X**

кусочно-постоянными функциями [3]. Входные параметры (w_{min}, w_{avg}) модифицируются с учетом шага квантования и общего числа точек в кривой и задаются в алгоритме решающего дерева, как максимальное число листьев (1) и минимальное количество точек для разбиения (2).

$$mlf = \frac{N \cdot s}{w_{avg}} \quad (1)$$

$$mss = \frac{2 \cdot w_{min}}{s} \quad (2)$$

Эти параметры будут являться основными настройками алгоритма разбиения на пропластки и будут присутствовать в любой реализации алгоритма решающего дерева в независимости от выбранной библиотеки машинного обучения [4].

Тестирование алгоритма

Наибольший интерес представляет собой тестирование алгоритма в случаях затруднительных даже для эксперта-интерпретатора. С учетом вышеобозначенного формировалась тестовая выборка данных.

Тестирование алгоритма проводилось на синтетических и реальных данных. В качестве синтетических данных использовалась сгенерированная модель разреза с известными свойствами и сигнал, полученный конволюцией с добавлением шума; реальные данные представлены записанными в интервале частого переслаивания геологических пропластков геофизическими методами разного вертикального разрешения. Результаты работы представлены на примере некоторых из использованных в тестовом наборе кривых на рисунках 1 и 2.

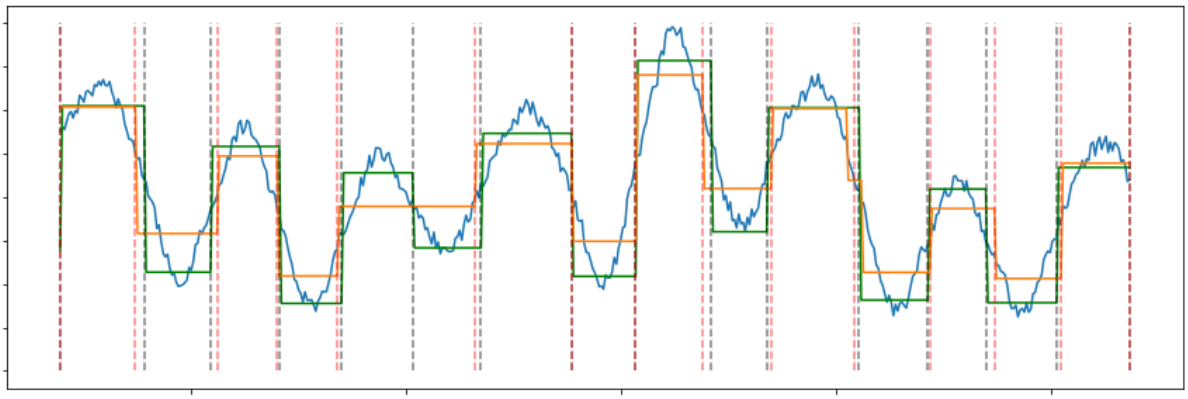

Рисунок 1. Синтетические данные. Синяя кривая – имитация сигнал прибора, зеленая кривая – истинные свойства модели, оранжевая – кусочно-постоянная функция алгоритма, красный пунктир – границы выделенные алгоритмом, черный пунктир – «реальные» границы

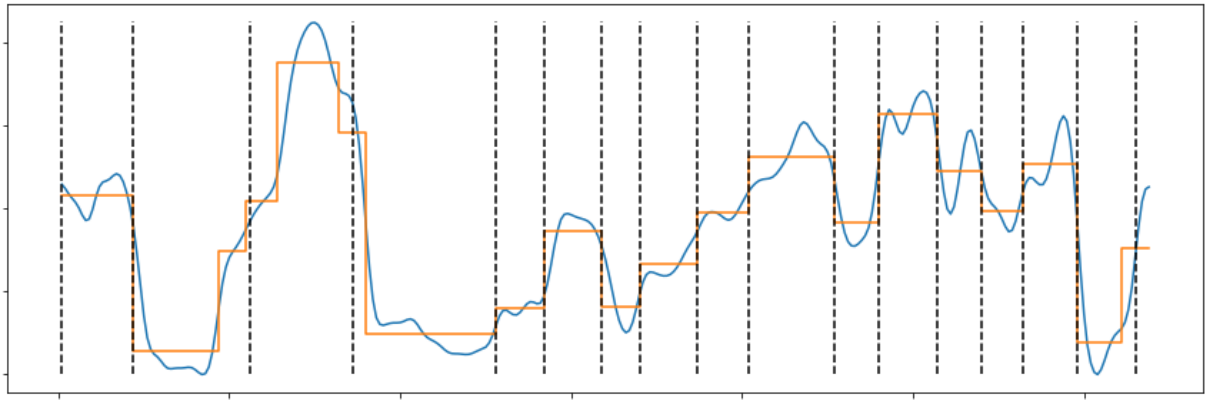

Рисунок 2. Реальные данные. Синяя кривая – геофизический метод, оранжевая кривая – кусочно-постоянная функция алгоритма, пунктир – разметка эксперта

Для некоторых интервалов можно отметить пропуски границ, что предположительно связано с особенностями разбиения дерева и наличием функциональной связи средней и минимальной мощности пропластка, являющимися входными параметрами алгоритма. В общем же случае, стоит отметить хорошее качество разбиения данных (среднее расхождение на реальных данных составило менее 0.1м). Выполненное профильными специалистами тестирование на большем массиве реальных данных различной конфигурации подтверждает качество работы алгоритма: пользователю в большинстве случаев приходится корректировать результаты работы для 0-10% процентов осредненных интервалов [6]. Данный метод применим и для

разбивки на пропластки по нескольким кривым одновременно: результаты работы алгоритма по каждой кривой в таком случае учитываются совместно.

Вывод

Предложенный алгоритм является альтернативой существующим решениям и однозначной заменой ручному выделению пропластков. Стоит отметить, что, как и любой другой алгоритм, позволяющий автоматизировать решение петрофизической задачи, он предполагает контроль выходных данных профильным специалистом. Результаты тестирования на реальных и синтетических данных отражают перспективность данного метода, а также удовлетворительную сходимость с результатами экспертной оценки. Реализация данного алгоритма доступна любому специалисту, без какого-ибо опыта в машинном обучении или же обработке сигналов, путем простого импортирования необходимой конфигурации из бесплатных библиотек (например sklearn).

В качестве основного направления улучшения алгоритма можно выделить: переопределение числа и сути входных параметров с целью избавления от косвенно коррелированных величин. В то время как минимальная мощность пропластков определяется однозначно, средняя мощность может быть весьма неоднозначной и требовать предварительного качественного анализа всего набора геофизических кривых.

Использованные источники

1. Латышова М.Г., Мартынов В.Г., Соколова Т.Ф. Практическое руководство по интерпретации методов ГИС. 2007г. 327с
2. Петерсилье В.И., Пороскун В.И., Яценко Г.Г. Методические рекомендации по подсчету геологических запасов нефти и газа объемным методов. – Москва-Тверь, 2003, 259с.
3. Quinlan, J. R. "Induction of decision trees" . Machine Learning, 1986: 81–106
4. Biggs, D; De Ville, B; Suen, E. A method of choosing multiway partitions for classification and decision trees". Journal of Applied Statistics, 1991, 18: 49–62

5. L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984.
6. Akinshin, A.V., Dmitrievskiy, M.V., Kantemirov, Y.D., Bobylev, K.D. 2021. “Automate Well Logging Data Interpretation with Digital Assistant”, SPE-208424-MS.

Used sources

1. Latyshova M.G., Martynov V.G., Sokolova T.F. A practical guide to interpreting well logging methods. 2007 327s
2. Petersilye V.I., Poroskun V.I., Yatsenko G.G. Methodical recommendations for calculating geological reserves of oil and gas by volumetric methods. - Moscow-Tver, 2003, 259s.
3. Quinlan, J. R. "Induction of decision trees". Machine Learning, 1986: 81-106
4. Biggs, D; De Ville, B; Suen, E. A method of choosing multiway partitions for classification and decision trees. "Journal of Applied Statistics, 1991, 18: 49–62
5. L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees,” Wadsworth, Belmont, CA, 1984.
6. Akinshin, A.V., Dmitrievskiy, M.V., Kantemirov, Y.D., Bobylev, K.D. 2021. “Automate Well Logging Data Interpretation with Digital Assistant,” SPE-208424-MS.

© Бобылев К.Д., 2021 Научно-образовательный журнал для студентов и преподавателей «StudNet» №12/2021.

Для цитирования: Бобылев К.Д., ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАЗБИЕНИЯ ГЕОФИЗИЧЕСКИХ КРИВЫХ НА ИНТЕРВАЛЫ// Научно-образовательный журнал для студентов и преподавателей «StudNet» №12/2021.