

МЕТОДИКА ПОДГОТОВКИ ДАННЫХ ДЛЯ ЭКСПРЕСС-АНАЛИТИКИ КАРОТАЖНЫХ КРИВЫХ С ПОМОЩЬЮ РАСПРЕДЕЛЕННОГО ФРЕЙМВОРКА APACHE SPARK НА ПРИМЕРЕ КЛАССИФИКАЦИИ КРИВЫХ ПС КАРОТАЖЕЙ

М.А. Сальников, инженер ОИТ ИШИТР
С.Н. Ушаков, аспирант гр. А1-39.
Томский политехнический университет
E-mail: mas55@tpu.ru, wowman9721@gmail.com

Введение

Современное месторождение генерирует огромное количество данных, которое сложно, а порой даже невозможно, качественно проанализировать имеющимися на данный момент средствами. Это обуславливает потребность в привлечении средств распределенной обработки данных с помощью использования традиционных структур данных и инструментов программирования. Исходя из этого, для работы с данными был выбран распределенный фреймворк для анализа данных Apache Spark, который способен предоставить весьма обширный программный интерфейс для работы с таким языком программирования и работы с данными, как Python. Стоит также отметить, что благодаря механизму ленивых преобразований, Spark способен снизить общий объем вычислений, тем самым повысив производительность работы по анализу больших данных. Особенно ценным видится применение такого подхода в области поддержки бурения, где специалистам приходится принимать в реальном времени решения, имеющие огромные последствия.

Целью данной работы выступает создать методику подготовки такого рода данных для подгрузки в модели машинного обучения, которые смогут выступать помощниками в принятии технологических решений. В данной работе, за неимением данных каротажей при непосредственном бурении скважины, был представлен пайплайн для обработки данных кривых каротажа потенциала собственной поляризации (ПС) снятых со скважин действующего добывающего фонда, как наиболее доступного аналога. Целью алгоритма стояло автоматизировано подготовить и классифицировать данные каротажа согласно классификации типовых форм кривых по Муромцеву, которые представлены на рисунке 1. В выборке присутствовали кривые снятые со 166 скважин, а так же данные по интервалам нахождения пласта коллектора для каждой из скважин, заранее подготовленные геологом [1].

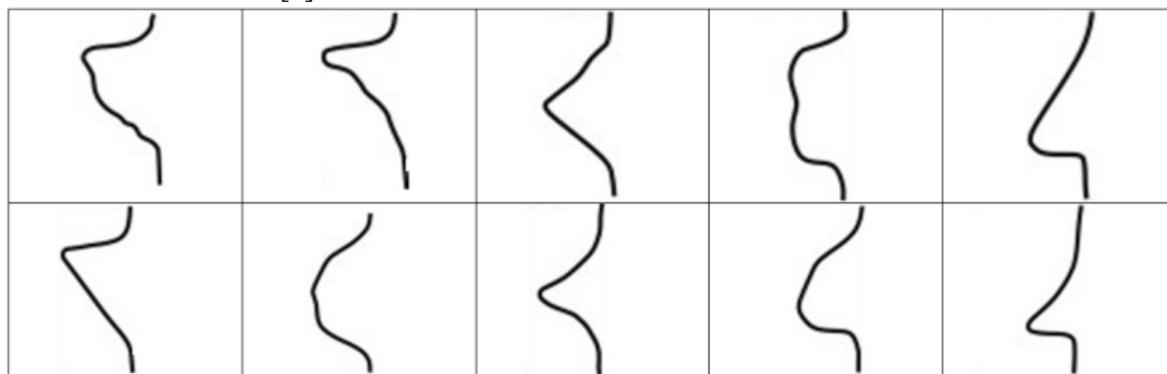


Рис. 1. Типовые кривые по Муромцеву

Процесс подготовки данных и их классификация

В процессе подготовки данных прежде всего необходимо было отсеять скважины, о положении целевого пласта которых данные отсутствуют. Также стоит отметить, что в предоставленных данных на некоторых промежутках значений глубины были выявлены отсутствующие значения. На этом основании можно сделать вывод о том, что на данных глубинах по каким-то причинам съемка не велась, вследствие чего было принято решение удалить эти значения. Для приведения заранее оцифрованных типовых кривых и реальных скважинных данных к общей системе координат был применен метод нормализации Min-Max Scaling.

Следующей проблемой работы с данными ПС каротажа явилась его высокая зашумленность, характерная данным электромагнитной природы. Так как следующим шагом в обработке данных

следует линейная интерполяция, высокий уровень шума мог негативно отразиться на результате, и получившаяся кривая плохо бы отражала поведение исходной. Для устранения этой проблемы и сглаживания кривой был применен метод спектрального анализа. Согласно этому методу кривая каротажа преобразуется в частотное представление с помощью быстрого преобразования Фурье (БПФ). В частотном представлении шумы детектируются и удаляются, после чего с помощью того же БПФ кривая восстанавливается уже без них. Линейная интерполяция с равными интервалами по глубине позволила привести все кривые к общей размерности, практически не утратив форму кривой [2].

Для того, чтобы обучить модель-классификатор, нам необходимо подать в нее в качестве учителя заранее оцифрованные типовые тренды. В качестве классификатора выбрана модификация алгоритма k-nearest neighbors с возможностью задания радиуса, в пределах которого находятся «соседи». Радиус был подобран экспериментальным путем на основе данных ПС каротажей целевого пласта, к которым были добавлены кривые, не относящиеся к интервалу пласта-коллектора и соответственно ни к одному из десяти классов. Таким образом, если алгоритм никак не классифицировал «фальшивые» кривые, то считалось, что радиус был подобран корректно и модель можно считать валидной. Работа итогового алгоритма на нескольких семплах из выборки представлена на рисунке 2.

В качестве метрики для классификации была использована DTW (dynamic time warping) благодаря своей способности уделять внимание непосредственно форме кривой [3].

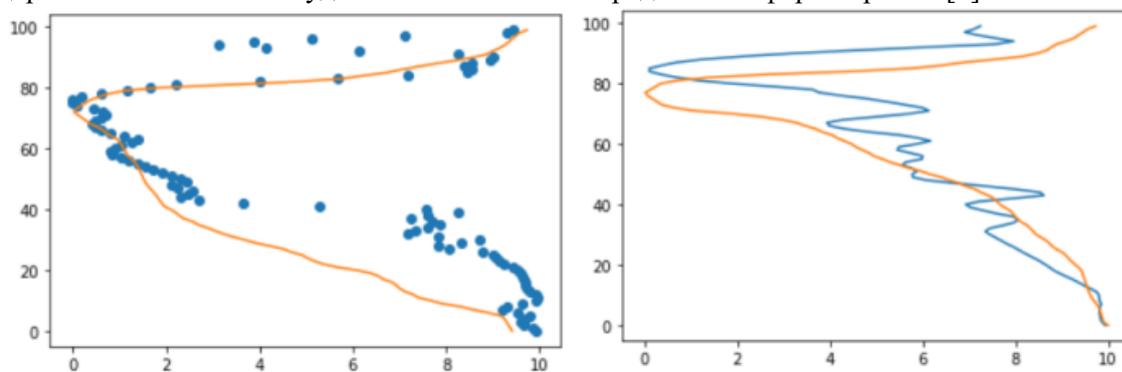


Рис. 2. Работа алгоритма классификации (оранжевым - типовая кривая, синим - реальные данные)

Заключение

Результатом работы служит разработанная методика работы с каротажными данными, которая позволяет проводить вычисления на большом объеме исходных данных благодаря фреймворку Spark. Таким образом, такой подход может обеспечить экспресс-аналитику по актуальным данным в сжатые сроки используя модели машинного обучения, разгружая специалиста и помогая ему принимать взвешенные и быстрые решения. По результатам работы классификации также видно, что алгоритм весьма неплохо улавливает тренды пространственного распределения пород-коллекторов, что свидетельствует о качественной подготовке данных.

Список использованных источников

1. Machine Learning Clustering of Reservoir Heterogeneity with Petrophysical and Production Data / D. V. Konoshonkin, G. Yu. Shishaev, I. V. Matveev [et al.] // SPE Europec featured at 82nd EAGE Conference and Exhibition Conferences Proceedings OnePetro, Amsterdam, The Netherlands, December 8-11, 2019. — Dallas: SPE, 2020. — [10 p.]
2. Классификация кривых ПС каротажа / С. Н. Ушаков, М. А. Сальников, А. Ю. Кайда; науч. рук. Е. И. Губин // Наука и практика: проектная деятельность - от идеи до внедрения материалы IX региональной научно-практической конференции, Томск, 2020. — Томск: ТУСУР, 2020. — [С. 71-73]
3. Cuturi, M. Blondel, M. 2017 Soft-DTW: a Differentiable Loss Function for Time-Series. ICML 2017.