

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧЕ ВОССТАНОВЛЕНИЯ КАРОТАЖНЫХ КРИВЫХ

*М.А. Сальников, инженер ОИТ ИШИТР,
Н.К. Вальман
Томский политехнический университет
E-mail: mas55@tpu.ru*

Введение

В настоящее время растет потребность в системах, способных не только выполнять однажды запрограммированную последовательность действий над заранее определенными данными, но и самостоятельно анализировать вновь полученную информацию, находить в ней закономерности, делать прогнозы и т.д. В этой области применения наилучшим образом зарекомендовали себя методы машинного обучения [1].

Целью данной работы является провести глубокую аналитику применимости методов машинного обучения в задаче восстановления каротажей нефтегазовых скважин на основании результатов других исследований на скважине [2]. Эксперимент был построен вокруг восстановления плотностного гамма-гамма каротажа, далее RHOV. Ценность данного подхода заключается в экономии средств на проведении геофизических исследований скважин, так как это является дорогостоящей операцией. Так же применить данные алгоритмы можно для восстановления ряда каротажных кривых в случае потери их из-за неполадок техники или человеческого фактора.

Описание алгоритма и процесс работы

Задачей работы стало определение наименований каротажей которые следует использовать при восстановлении RHOV, определение оптимального количества этих каротажей, а так же определение оптимальной архитектуры модели машинного обучения в применении к конкретно этой задаче [4]. Для выполнения первых двух задач был применен корреляционный анализ, результаты которого можно увидеть на рисунке 1. Итоговый выбор велся на основании двух критериев: физическая природа каротажа и степень его скоррелированности с RHOV. В итоге было решено провести 2 типа вычислительных экспериментов: с применением 8 и 5 признаков. Так в случае с 5 признаками были использованы RMED, RDEP, GR, PEF, SP каротажи, а в случае с 8 признаками к ним добавились еще и DRHO, LITHOLOGY, CONFIDENCE.

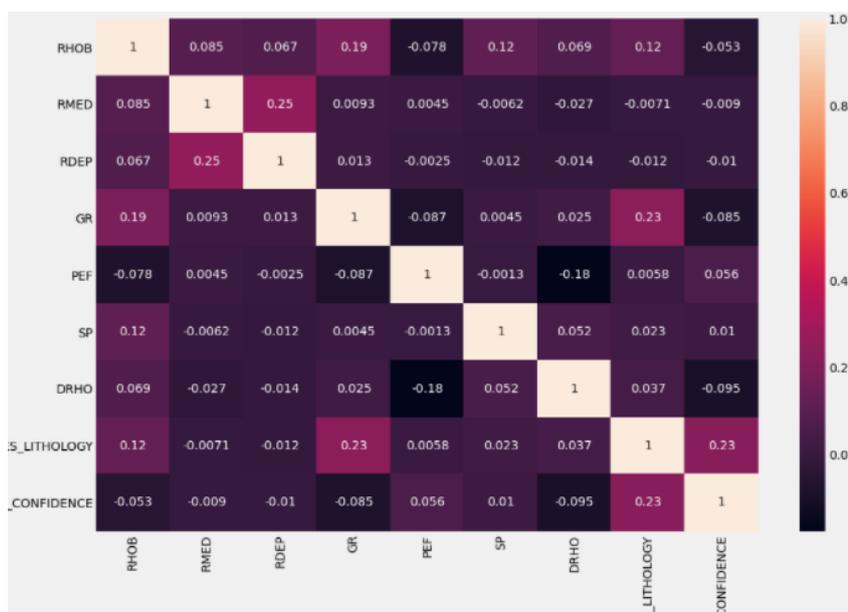


Рис. 1. Корреляционная матрица после фильтрации коррелированных значений для прогнозирования по 8 логарифмическим кривым.

Формирование массива данных для применения их в процессе обучения происходило с применением скользящего окна по каждой из выбранных кривых [5]. Так как часто значения могли быть близки и не нести качественной информации, а лишь “раздуть” итоговый датасет, было решено снова поставить два типа вычислительных экспериментов. В первом случае модели обучались на исходном датасете, то есть в исходном метрическом пространстве. Во втором случае было решено применить метод снижения размерности PCA (метод главных компонент). В результате было сохранено более 90% исходной дисперсии изначальной выборки, что можно считать качественным результатом [3]. Для нормализации данных применялся MinMaxScaler.

В качестве моделей машинного обучения было выбрано два зарекомендовавших себя метода: Random Forest и нейронная сеть. В качестве методов регуляризации в нейронных сетях было применено две методики: BatchNormalization и Dropout. Таким образом в общей сложности было обучено 12 моделей. Аналитика по их качеству представлена в таблице 1. Оценка алгоритмов производилась посредством среднего квадратичного отклонения.

Таблица 1. Сравнение качества результирующих моделей на основании MSE

Модель	Пространство	Регуляризация	Количество признаков	MSE на обучающей выборке	MSE на тестовой выборке
Random Forest	Исходное		5	0,149	0,149
			8	0,147	0,148
	PCA		5	0,169	0,170
			8	0,212	0,212
Нейронная сеть	Исходное	BatchNormalization	5	0,14699	0,148
			8	0,136	0,138
		Dropout	5	0,405	0,406
			8	0,264	0,264
	PCA	BatchNormalization	5	0,158	0,159
			8	0,218	0,218
		Dropout	5	0,180	0,181
			8	0,451	0,452

Заключение

В результате проведения тестирования можно сделать вывод о том, что алгоритмы являются работоспособными. Лучшим алгоритмом оказалось классическая нейронная сеть, в исходном метрическом пространстве с применением метода регуляризации BatchNormalization обученная с применением 8 признаков. Что выглядит логичным. Тем не менее можно заметить незначительную разницу между качеством нейронной сети и Random Forest при обучении на 5 признаках, при ощутимой разнице в трудоемкости и скорости обучения каждого из алгоритмов. Поэтому Random Forest вполне может выступать альтернативой применению нейронных сетей при недостатке данных. Применение PCA не оправдало себя. При его применении качество моделей ощутимо снижалось.

Список использованных источников

1. Шмидхубер, Дж. (1992). Алгоритм обучения временной сложности хранилища фиксированного размера для полностью повторяющихся непрерывно работающих сетей. Нейронные вычисления, 4(2), 243-248 с.
2. Шривастава Р. К., Грефф К. и Шмидхубер Дж. (2015). Сети автомобильных дорог. архив препринт arXiv: 1505.00387
3. Уильямс, Р. Дж., и Пэн, Дж. (1990). Эффективный алгоритм на основе градиента для онлайн-обучения рекуррентных сетевых траекторий. Нейронные вычисления, 2 (4), 490-501 с.
4. Уильямс, Р. Дж., и Зипсер, Д. (1995). Алгоритмы обучения на основе градиента для рекуррентных. Обратное распространение: теория, архитектура и приложения, 433 с.
5. Чжан, Д., Юньтянь, С. Н. Е. N., и Цзинь, М. Е. N. G. (2018). Генерация синтетических журналов скважин с помощью рекуррентных нейронных сетей. Разведка и разработка нефти, 45(4), 629-639 с.