

---

---

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

---

---

УДК 550.8.072  
DOI: 10.17586/0021-3454-2021-64-7-532-541

## ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗА ИЛИ ЗАМЕЩЕНИЯ НЕДОСТАЮЩИХ КАРОТАЖНЫХ ДАННЫХ

Р. Д. АХМЕТСАФИН<sup>1</sup>, Р. З. АХМЕТСАФИНА<sup>2</sup>

<sup>1</sup>Московский физико-технический институт, Инжиниринговый центр по полезным ископаемым,  
141700, г. Долгопрудный, Московская обл., Россия

<sup>2</sup>Национальный исследовательский университет „Высшая школа экономики“,  
101000, Москва, Россия  
E-mail: rakhmetsafina@hse.ru

Сопоставляются 9 методов машинного обучения (ANN, ANFIS, ELM, FM, SVM, GPR, RF, RT, k-NN) на примере прогнозирования данных акустического каротажа. Решение задачи регрессии при машинном обучении может применяться не только для прогноза геофизических полей, но и для замещения недостающих данных. Построенную кривую  $\Delta T(p)$  интервального времени  $P$ -волны можно рассматривать как результат прогноза, если позднее предполагается акустический каротаж; если же дополнительный акустический каротаж невозможен, то синтетическая кривая  $\Delta T(p)$  замещает каротажную при дальнейшей интерпретации. Наилучшие результаты по тестовому испытанию показал метод RF.

**Ключевые слова:** машинное обучение, задача регрессии, замещение недостающих данных, акустический каротаж

**Введение.** Актуальность замещения недостающих каротажных данных возрастает в связи с широким применением на месторождениях, в поздней стадии эксплуатации, гидродинамического и геомеханического моделирования. Это объясняется тем, что современный комплекс ГИС (геофизического исследования скважин), отбора и анализа каменного материала применительно к этим задачам претерпел значительные изменения по сравнению с комплексом 20—30-летней давности (особенно это касается верхней части разреза, где комплекс ГИС был минимальным). Применение простых эмпирических зависимостей для оценки физико-механических и фильтрационно-емкостных свойств по фактически имеющимся каротажным записям и каменному материалу не всегда позволяет построить качественные модели. Наиболее востребованным является замещение недостающих кривых плотностного и акустического каротажа — для оценки физико-механических свойств и порового давления при геомеханическом моделировании; нейтронного гамма-каротажа — для оценки пористости и литологического расчленения; каротажа потенциала собственной поляризации — для фациального анализа и др.

Современные методы машинного обучения (Machine Learning Algorithms — MLA) и анализа данных основаны на обучении на прецедентах. Они имеют открытые исходные коды и ориентированы на минимальное интерактивное вмешательство. Эти методы нашли широ-

кое применение в различных областях исследований для классификации и составления прогнозов, где детерминированные модели не могут быть четко сформулированы.

При использовании MLA предполагается, что каждый исследуемый объект описывается парой  $\langle \mathbf{x}, y \rangle$ , где  $\mathbf{x}$  — данные (многомерный вектор),  $y$  — целевое значение, метка. Необходимо найти функцию  $f(\mathbf{x}) \approx y$ . В зависимости от значений  $y$  MLA разделяется на две задачи:

1) классификация — значения  $y$  дискретны (принимают несколько заранее определенных значений или классов);

2) регрессия — значения  $y$  непрерывны (принимают любое значение из диапазона).

Большинство методов MLA способны решать обе задачи.

В геонауках [1] *задача классификации* может применяться для предсказания месторождений полезных ископаемых, предсказания аномальных проявлений в земной коре и природных процессов, стратиграфического и литологического расчленения, фациального анализа, корреляции разрезов скважин и пр. Классификация применяется как к исходным исследуемым объектам, при этом результатом являются обновленные (робастные или „классифицированные“) оценки значения  $y_{\text{class}} = f(\mathbf{x})$ , так и к новым объектам для прогноза  $y_{\text{class\_new}} = f(\mathbf{x}_{\text{new}})$ .

*Задача регрессии* в нотации MLA в геонауках может применяться для прогноза или замещения недостающей метеорологической, геолого-геофизической и промысловой информации ( $y_{\text{new}} = f(\mathbf{x}_{\text{new}})$ ).

Как в задаче классификации, так и в задаче регрессии предполагается, что построенная зависимость  $y = f(\mathbf{x})$  носит региональный характер.

В работе [2] для решения второй задачи — регрессии — применительно к прогнозу/замещению недостающих данных на примере акустического каротажа рассмотрены 3 метода MLA: SVM, RF и k-NN. В настоящей статье сопоставляются 9 методов MLA, наиболее часто применяемых в геонауках: ANN, ANFIS, FM, SVM, ELM, GPR, k-NN, RT и RF. В качестве примера рассматривается процесс прогнозирования данных акустического каротажа и приводятся результаты их сопоставления с имеющимися каротажными записями, а также с оценками, полученными методом линейной регрессии.

#### Тестируемые методы машинного обучения.

**ANN (Artificial Neural Networks)** — искусственные нейронные сети; **MLP (Multi Layer Perceptron)** — многослойный перцептрон. Наиболее распространенный подход к разработке непараметрической и нелинейной классификации/регрессии основан на методе ANN. Существует много различных типов ANN. Основными обрабатываемыми элементами искусственной нейронной сети прямого распространения (Feed-Forward Propagation Neural Network) [3, 4] являются нейроны. В нейронной сети блоки нейронов размещаются слоями и соединяются таким образом, что информация передается однонаправленно, от входных блоков — через блоки, расположенные на скрытых слоях, к блокам в выходном слое. Нейрон в основном выполняет линейную регрессию, за которой следует нелинейная функция. Нейроны разных слоев связаны между собой соответствующими связями (весами). Задача алгоритма обучения состоит в том, чтобы найти набор весов, который гарантирует, что для каждого входного вектора результирующий вектор из сети будет таким же или достаточно близким к выходному вектору. Применение метода в геонауках отражено в работах [5—8], прогноз акустического каротажа — в [9, 10].

**ANFIS (Adaptive Neuro-Fuzzy Inference System)** — искусственная нейронная сеть, основанная на нечеткой системе вывода Такаги — Сугено [11]. Метод был разработан в начале 1990-х гг. [12, 13]. Применение в геонауках — [8, 14], прогноз акустического каротажа — [15, 16].

**FM (Fuzzy Models)** — нечеткие модели прогнозирования. Метод разработан в начале 1990-х гг. [17] и предполагает три этапа обучения: 1) кластеризация по алгоритму Густафсона — Кесселя (Gustafson — Kessel) [18]; 2) построение по методу наименьших квадратов регрессионных моделей (как линейных, так и нелинейных) для каждого кластера; 3) построение

нечеткой системы вывода (Такаги — Сугено) для набора регрессионных моделей. Применение в геонауках — [19], прогноз акустического каротажа — [15, 20].

**SVM (*Support-Vector Machine, SVR Support-Vector Regression*)** — метод опорных векторов. Метод предложен в конце 1960-х гг. [21]. Первоначально он был разработан как метод линейной классификации, позднее обобщен для нелинейного классификатора и, наконец, был применен к регрессионным моделям. Основная идея метода заключается в трансформации входных объектов в некоторое многомерное пространство, где они могут быть линейно разделены на классы гиперплоскостями. Алгоритм формализован в работах [21, 22], применение в геонауках — [6, 7, 23], прогноз акустического каротажа — [2, 24].

**ELM (*Extreme Learning Machines*)** — экстремальное машинное обучение. Метод предложен в 2011 г. [25] и по сути является многослойной нейронной сетью прямого распространения. Отличие метода заключается в том, что узлами здесь являются вычислительные элементы, которые рассматриваются шире, чем классический нейрон ANN. Узлами в ELM на каждом слое могут быть как классические искусственные нейроны, так и подсети, образованные скрытыми узлами. Применение в геонауках — [26, 27], прогноз акустического каротажа — [28].

**GPR (*Gaussian Process Regression*)** — регрессия гауссовского процесса — еще один метод машинного обучения на основе ядер для задач нелинейной регрессии [29]. Метод GPR, как правило, формулируется и интерпретируется в байесовском контексте [30]. Функциональная зависимость, описывающая модель прогнозирования, строится итерационно. На каждой итерации формируется функциональная зависимость по исходной обучающей выборке, при этом результат предыдущей итерации используется для ограничения возможных форм результирующей зависимости. Начальная функциональная зависимость задается или генерируется случайным образом. Применение в геонауках — [29, 31], прогноз акустического каротажа — [9, 32].

**k-NN (*k-Nearest Neighbors*)** — метод  $k$ -ближайших соседей — простой алгоритм, который хранит все доступные случаи и предсказывает числовую цель на основе меры подобия (например, функции расстояния). Используется при статистической оценке и распознавании образов с начала 1970-х гг. как непараметрический метод. Основным принцип метода — объект  $\langle x, y \rangle$  относится к тому классу, которому принадлежит большинство из его соседей [33]. Применение в геонауках — [6, 34], прогноз акустического каротажа — [2].

**RT (*Regression Tree*)** — регрессия с помощью деревьев. Популярность деревьев решений (Decision Trees — DT) [35] связана с простотой, интерпретируемостью, малыми вычислительными затратами и возможностью графического представления. Дерево решений представляет собой набор иерархически организованных ограничений или условий, которые последовательно применяются от корня к конечному узлу или листу дерева. Применение в геонауках — [5, 7, 36].

**RF (*Random Forests*) и ELR (*Ensemble Learners Regression*)** — „случайный лес“, ансамблевые методы прогнозирования: бэггинг, бустинг. Алгоритм, сочетающий в себе две идеи: метод бэггинга Бреймана и метод случайных подпространств, применяется для решения задач классификации, регрессии и кластеризации. Основная идея заключается в использовании большого ансамбля решающих деревьев, каждое из которых само по себе дает невысокое качество классификации или регрессии, но за счет их большого количества результат получается удовлетворительным. Алгоритм описан в работах [33, 37, 38]. Применение в геонауках — [5—7], прогноз акустического каротажа — [2].

Все рассмотренные методы представлены открытыми кодами в программах MatLab, Python, Fortran или C++. Авторами обобщены эти коды и разработан единый интерфейс для удобного выбора объектов (кривых на геофизических планшетах), настройки каждого метода и сопоставления результатов. Ниже приведены результаты тестирования методов.

**Пример.** В качестве примера рассмотрены LAS-файлы пяти скважин из открытого доступа: [http://www.kgs.ku.edu/PRS/Scans/Log\\_Summary/index.html](http://www.kgs.ku.edu/PRS/Scans/Log_Summary/index.html) (Kansas Geological Survey, The University of Kansas, AUG.zip, 2019; табл. 1). Расположение скважин на карте — <https://maps.kgs.ku.edu/oilgas/index.html>.

Таблица 1

Номер скважины (№)	Название файла	Скважина	Территория	Относительные координаты скважины
1	1050385055.las	SCHNEIDER TRUST 13B #1	WILDCAT	664' FNL & 1402' FWL
2	1050385385.las	UHLAND NO. 1-16	ANTELOPE SPRINGS	2278' FSL & 1214' FWL
3	1050385491.las	SELTMANN NO. 1-12	START	610' FNL & 2440' FWL
4	1050385503.las	MUNN #2-12	SAWYER CANYON	1000' FSL & 1110' FEL
5	1050385520.las	JACKIE NO.1	WALNUT VALLEY	375' FNL & 1585' FEL

Примечание. FNL (from north line), FSL (from south line), FWL (from west line), FEL (from east line) — принятые в геофизике обозначения координат относительно геодезических линий.

Целевое значение для обучения моделей — кривая интервального времени  $P$ -волны, в отечественной научной литературе обычно обозначается как  $\Delta T(P)$  и измеряется в микросекундах на метр, мкс/м (в зарубежной литературе обозначается как  $DT(P)$  или *slowness*, единицы измерения usec/ft). Обучение проводилось по десяти кривым: GR (Gamma Ray — гамма-каротаж), RHOV (Rho Bulk — плотностной гамма-гамма каротаж), RHOC (Rho Corrected — скорректированная плотность), RILD (Resistivity Induction Log, Deep — индукционный каротаж, удельное сопротивление в дальней зоне), RILM (Resistivity Induction Log, Medium — индукционный каротаж, удельное сопротивление в средней зоне), RLL3 (Resistivity Laterolog 3 — электромагнитный каротаж, удельное сопротивление в ближней зоне), Rxo/Rt (Flushed Zone Resistivity / Resistivity — сопротивление зоны проникновения/сопротивление пласта), SP (Spontaneous Potential — потенциал собственной поляризации), MCAL (Monmitored Caliper — профилемер), MN (Micro Normal Resistivity — микронормальное сопротивление). Каротажные кривые, отмеченные подчеркиванием, представлены на рис. 1 для скважины №1 (по оси ординат —  $h$  — глубина скважины).

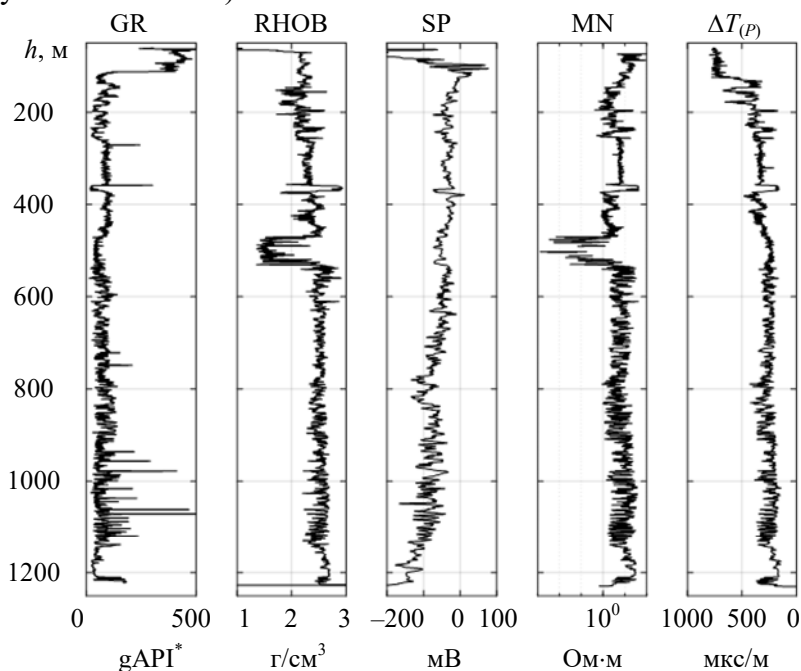


Рис. 1

\* gAPI — единицы измерения для кривой гамма-каротажа.

В процессе обучения модели прогнозирования  $\Delta T_{(P)}$  последовательно использованы данные по одной (№ 1), двум (№ 1÷2), трем (№ 1÷3), четырем (№ 1÷4) и пяти (№ 1÷5) скважинам. Среднеквадратическая ошибка (СКО) прогноза и каротажных значений  $\Delta T_{(P)}$  для каждой скважины и рассматриваемых методов представлена в табл. 2. Как LR обозначена линейная регрессия по методу наименьших квадратов; в скобках указан вычислительный индекс — время выполнения на компьютере относительно LR.

Таблица 2

Номера скважин	Скважина					Скважина				
	№ 1	№ 2	№ 3	№ 4	№ 5	№ 1	№ 2	№ 3	№ 4	№ 5
LR (вычислительный индекс = 1)						ELM (1.7)				
1÷1	287.28	142.33	58.26	89.29	175.45	125.76	287.34	171.48	268.86	127.81
1÷2	291.94	109.35	55.04	58.31	165.12	153.73	80.90	127.71	177.96	50.86
1÷3	291.86	110.52	52.67	60.86	157.34	151.32	74.66	48.60	196.75	44.59
1÷4	296.71	100.74	55.61	45.23	158.24	183.33	66.63	53.16	75.93	56.55
1÷5	301.97	96.42	51.54	45.75	124.84	159.23	62.15	40.03	60.54	26.65
ANFIS (17.6)						SVM (9.3)				
1÷1	287.28	142.33	58.26	89.29	175.45	301.83	118.56	53.24	53.90	161.62
1÷2	185.70	82.78	57.01	43.19	75.82	79.01	15.65	34.19	22.03	159.49
1÷3	155.40	63.45	51.52	36.96	70.61	86.07	17.57	11.13	20.58	141.91
1÷4	181.29	80.14	53.62	29.18	95.28	95.44	17.86	11.66	10.85	131.18
1÷5	154.60	62.12	51.04	27.71	54.09	98.05	17.47	11.73	10.56	6.54
ANN (277.1)						FM (25.0)				
1÷1	85.09	72.61	35.57	33.51	45.52	29.28	255.17	114.60	32.99	211.67
1÷2	92.01	49.27	24.96	28.49	37.22	35.90	13.36	25.24	17.96	24.54
1÷3	85.27	31.82	21.80	29.56	45.99	36.86	13.48	11.28	15.95	29.39
1÷4	95.69	45.26	26.70	20.95	40.20	38.18	14.03	11.71	8.95	28.13
1÷5	95.10	36.08	20.57	17.51	18.45	41.25	14.64	11.95	9.51	9.93
GPR (61.9)						k-NN (95.8)				
1÷1	19.89	155.03	84.85	56.43	148.32	0.00	174.36	73.41	84.36	68.00
1÷2	19.37	6.12	45.01	29.79	95.35	0.00	0.00	76.80	60.42	53.69
1÷3	22.79	8.07	7.20	25.05	68.15	0.00	0.00	0.00	55.28	41.71
1÷4	26.41	9.57	8.42	8.31	108.82	0.00	0.00	0.00	0.00	44.35
1÷5	26.68	10.08	8.67	8.31	4.85	0.00	0.00	0.00	0.00	0.00
RT (1.2)						RF (11.5)				
1÷1	6.78	95.25	48.55	39.81	211.98	9.95	50.29	22.46	18.87	40.37
1÷2	6.40	3.10	47.30	34.88	39.58	9.85	5.17	18.75	12.38	40.24
1÷3	5.59	3.05	2.40	31.35	35.98	9.80	5.35	2.94	11.06	35.89
1÷4	5.99	3.18	2.21	1.94	39.58	10.11	5.25	2.98	2.41	32.61
1÷5	5.96	3.02	2.49	2.18	1.66	9.99	5.43	2.95	2.39	1.92
ELR bagging (8.4)						ELR least-squares boosting (4.0)				
1÷1	9.61	48.14	22.65	19.54	40.18	13.45	83.75	51.82	52.42	84.20
1÷2	10.02	5.15	19.36	12.69	32.96	17.00	11.44	40.86	29.49	73.64
1÷3	9.93	5.28	2.83	11.71	31.64	17.34	11.88	10.59	26.71	61.34
1÷4	9.92	5.38	2.97	2.41	33.05	18.82	12.04	10.61	8.87	38.80
1÷5	9.99	5.40	2.90	2.38	1.95	18.83	11.45	9.96	8.71	7.86

На рис. 2 представлены результаты сопоставления прогнозируемых и зарегистрированных значений  $\Delta T_{(P)}$  для метода RF по пяти скважинам при обучении (модели прогнозирования) только по одной скважине №1 (соответствующие СКО приведены в табл. 2).

На рис. 3, а в качестве интегральной оценки результатов прогнозов по обучению моделей представлены усредненные значения СКО по каждой скважине (см. табл. 2); на рис. 3, б приведены значения итогового (по всем скважинам) относительного показателя погрешности

(δ) прогнозирования (для LR значение показателя равно 1) по результатам проведенного тестирования.

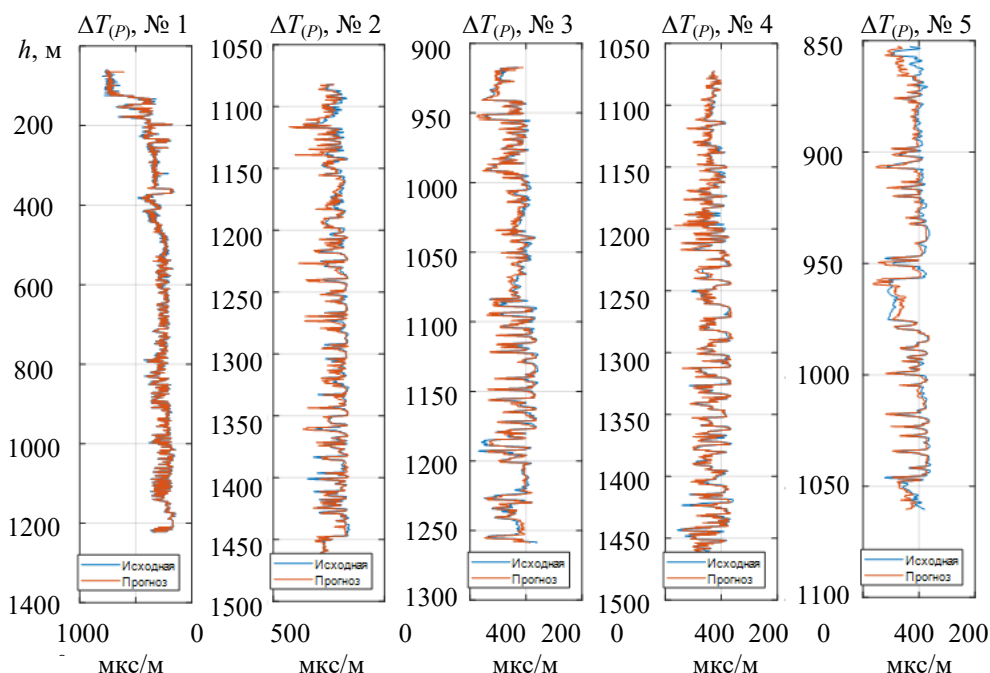


Рис. 2

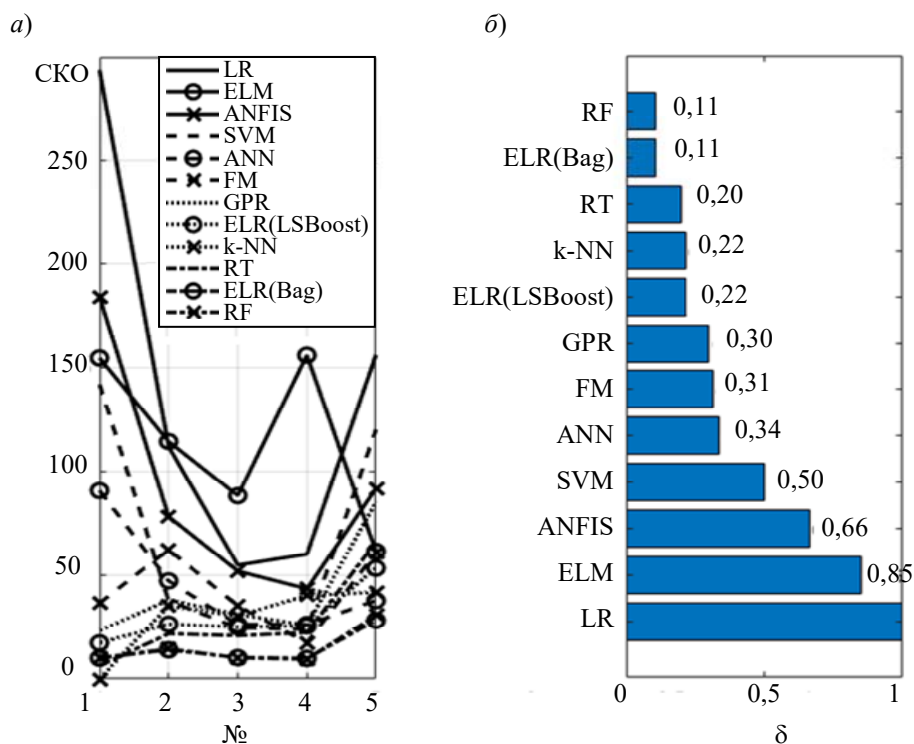


Рис. 3

**Заключение.** Метод МЛА — эффективное средство накопления и хранения знаний, позволяющее решать задачи интерпретации информации в условиях неопределенностей, таких как недостаток данных, или формализованных зависимостей. Для этих алгоритмов характерна одна особенность — они не всегда дают повторяющийся результат при одинаковых данных (прогноз при каждом цикле обучения и последующем расчете может не повторяться). Это обусловлено тем, что решение зависит от начального приближения, генерируемого

случайным образом. В „интерполяции“ различия решений (прогнозов) практически не заметны, а в „экстраполяции“ могут быть существенными.

Вычислительные затраты доступных реализаций MLA могут отличаться в сотни раз, а ошибки прогноза — в десятки раз. Причем наиболее затратные MLA не гарантируют высокую точность прогноза.

По результатам тестирования доступных реализаций MLA (Matlab Deep Learning Toolbox; Matlab Fuzzy Logic Toolbox; Matlab Statistics and Machine Learning Toolbox; Extreme Learning Machine for Classification and Regression; Fuzzy Modelling and Identification Toolbox) для решения задачи замещения недостающих каротажных данных сформирован следующий рейтинг методов: 1) RF (ELR(bagging)), 2) RT, 3) k-NN, 4) GPR, 5) FM, 6) ANN, 7) SVM, 8) ANFIS, 9) ELM. Неожиданным является последнее место ELM (по трем из пяти скважин совокупный прогноз оказался даже хуже линейной регрессии, см. рис. 3, а). Возможно, реализация метода, представленная в работе [28], не самая удачная.

### СПИСОК ЛИТЕРАТУРЫ

1. *Caté A., Perozzi L., Gloaguen E., Blouin M.* Machine learning as a tool for geologists // *The Leading Edge*. 2017. Vol. 36, N 3. P. 215—219.
2. *Pandey S., Saraiya R.* Prediction of sonic log data using machine learning regression methods // *Marine Acquisition Workshop*. 2018.
3. *Mas J. F., Flores J. J.* The application of artificial neural networks to the analysis of remotely sensed data // *Intern. Journal of Remote Sensing*. 2008. Vol. 29, N 3. P. 617—663.
4. *Rumelhart D. E., Hinton G. E., Williams R. J.* Learning representations by back-propagating errors // *Nature*. 1986. Vol. 323, N 6088. P. 533—536.
5. *Akinnikawe O., Lyne S., Roberts J.* Synthetic well log generation using machine learning techniques // *Unconventional Resources Technology Conf. Proc., Houston, Texas, USA, 23—25 July 2018; Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers*, 2018. P. 1507—1522.
6. *Cracknell M. J., Reading A. M.* Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information // *Computers & Geosciences*. 2014. Vol. 63. P. 22—33.
7. *Rodriguez-Galiano V., Sanchez-Castillo M., Chica-Olmo M., Chica-Rivas M. J. O. G. R.* Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines // *Ore Geology Reviews*. 2015. Vol. 71. P. 804—818.
8. *Yilmaz I., Kaynar O.* Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils // *Expert Systems with Applications*. 2011. Vol. 38, N 5. P. 5958—5966.
9. *Gupta I., Devegowda D., Jayaram V., Rai C., Sondergeld C.* Machine learning regressors and their metrics to predict synthetic sonic and mechanical properties // *Interpretation*. 2019. Vol. 7, N 3. P. 1—56.
10. *Onalo D., Oloruntobi O., Adedigba S., Khan F., James L., Butt S.* Dynamic data driven sonic well log model for formation evaluation // *J. of Petroleum Science and Engineering*. 2019. Vol. 175. P. 1049—1062.
11. *Takagi T., Sugeno M.* Fuzzy identification of systems and its applications to modeling and control // *IEEE Trans. on Systems, Man and Cybernetics*. 1985. N 1. P. 116—132.
12. *Jang J. S. R.* ANFIS: adaptive-network-based fuzzy inference system // *IEEE Trans. on Systems, Man, and Cybernetics*. 1993. Vol. 23, N 3. P. 665—685.
13. *Karaboga D., Kaya E.* Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey // *Artificial Intelligence Review*. 2019. Vol. 52, N 4. P. 2263—2293.
14. *Cabalar A. F., Cevik A., Gokceoglu C.* Some applications of adaptive neuro-fuzzy inference system (ANFIS) in geotechnical engineering // *Computers and Geotechnics*. 2012. Vol. 40. P. 14—33.

15. Ахметсафин Р. Д., Ахметсафина Р. З. Практическое применение нечетких моделей для интерпретации ГИС // Геофизический вестник. 2016. № 1. С. 11—14.
16. Basarir H. Prediction of rock mass P wave velocity using blasthole drilling information // Intern. Journal of Mining, Reclamation and Environment. 2019. Vol. 33, N 1. P. 61—74.
17. Babuška R. Fuzzy Modeling for Control. Springer Science & Business Media, 2012. Vol. 12.
18. Gustafson D. E., Kessel W. C. Fuzzy clustering with a fuzzy covariance matrix // IEEE Conf. on Decision and Control Including the 17th Symp. on Adaptive Processes. 1978. Vol. 17. P. 761—766.
19. Grima A. M., Babuška R. Fuzzy model for the prediction of unconfined compressive strength of rock samples // Intern. Journal of Rock Mechanics and Mining Sciences. 1999. Vol. 36, N 3. P. 339—349.
20. Kieu D. T., Keping A., Kitzig M. C. Prediction of sonic velocities from other borehole data: An example from the Keivitsa mine site, northern Finland // Geophysical Prospecting. 2018. Vol. 66, N 9. P. 1667—1683.
21. Vapnik V. The Nature of Statistical Learning Theory. Springer Science & Business Media, 2013.
22. Chang C. C., Lin C. J. LIBSVM: A library for support vector machines // ACM Trans. on Intelligent Systems and Technology (TIST). 2011. Vol. 2, N 3. P. 27.
23. Al-Anazi A. F., Gates I. D. Support vector regression to predict porosity and permeability: effect of sample size // Computers & Geosciences. 2012. Vol. 39. P. 64—76.
24. Bagheripour P., Gholami A., Asoodeh M., Vaezzadeh-Asadi M. Support vector regression based determination of shear wave velocity // J. of Petroleum Science and Engineering. 2015. Vol. 125. P. 95—99.
25. Huang G. B., Zhou H., Ding X., Zhang R. Extreme learning machine for regression and multiclass classification // IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics). 2011. Vol. 42, N 2. P. 513—529.
26. Olatunji S. O., Selamat A., Raheem A. A. A. Modeling permeability prediction using extreme learning machines // 4th Asia Intern. Conf. on Mathematical/Analytical Modelling and Computer Simulation, IEEE. 2010. P. 29—33.
27. Shi X., Wang J., Liu G., Yang L., Ge X., Jiang S. Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs // J. of Natural Gas Science and Engineering. 2016. Vol. 33. P. 687—702.
28. Cao J., Shi Y., Wang D., Zhang X. Acoustic log prediction on the basis of Kernel extreme learning machine for wells in Gjh Survey, Erdos Basin // J. of Electrical and Computer Engineering. 2017. Vol. 2017.
29. Hultquist C., Chen G., Zhao K. A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests // Remote Sensing Letters. 2014. Vol. 5, N 8. P. 723—732.
30. Williams C. K. I., Rasmussen C. E. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006.
31. Asante-Okyere S., Shen C., Yevenyo Ziggah Y., Moses Rulegeya M., Zhu, X. Investigating the predictive performance of gaussian process regression in evaluating reservoir porosity and permeability // Energies. 2018. Vol. 11, N 12. P. 3261.
32. Bukar I., Adamu M. B., Hassan U. A machine learning approach to shear sonic log prediction // SPE Nigeria Annual Intern. Conf. and Exhibition. Society of Petroleum Engineers, 2019.
33. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media, 2009.
34. Pratama H. Machine learning: using optimized KNN (K-Nearest Neighbors) to predict the facies classifications // Proc. of the 13th SEGJ Intern. Symp., Tokyo, Japan, 12—14 Nov. 2018; Society of Exploration Geophysicists and Society of Exploration Geophysicists of Japan, 2019. P. 538—541.
35. Friedl M. A., Brodley C. E. Decision tree classification of land cover from remotely sensed data // Remote Sensing of Environment. 1997. Vol. 61, N 3. P. 399—409.
36. Lee S., Jung H. S. Machine Learning Techniques Applied to Geoscience Information System and Remote Sensing. Basel: MDPI, 2019.
37. Breiman L. Random forests // Machine learning. 2001. Vol. 45, N 1. P. 5—32.
38. Friedman J. H. Greedy function approximation: a gradient boosting machine // Annals of Statistics. 2001. P. 1189—1232.



**Сведения об авторах**

- Раис Дахиевич Ахметсафин** — д-р техн. наук, доцент; Инжиниринговый центр МФТИ по полезным ископаемым, департамент цифровых технологий в индустрии; E-mail: akhmetsafinrd@mail.ru
- Римма Закиевна Ахметсафина** — канд. техн. наук, доцент; НИУ „Высшая школа экономики“, департамент программной инженерии факультета компьютерных наук; E-mail: rakhmetsafina@hse.ru

Поступила в редакцию  
19.03. 2021 г.

**Ссылка для цитирования:** Ахметсафин Р. Д., Ахметсафина Р. З. Применение методов машинного обучения для прогноза или замещения недостающих каротажных данных // Изв. вузов. Приборостроение. 2021. Т. 64, № 7. С. 532—541.

**APPLYING MACHINE LEARNING METHODS TO PREDICT OR REPLACE MISSING LOGGING DATA**

R. D. Akhmetsafin<sup>1</sup>, R. Z. Akhmetsafina<sup>2</sup>

<sup>1</sup>Moscow Institute of Physics and Technology, Engineering Center for Minerals,  
141700, Dolgoprudny, Moscow Region, 141700, Russia

<sup>2</sup>HSE University, 101000, Moscow, Russia  
E-mail: rakhmetsafina@hse.ru

Nine machine learning methods (ANN, ANFIS, ELM, FM, SVM, GPR, RF, RT, k-NN) are compared using the example of predicting acoustic logging data. With machine learning, the solution to the regression problem can be used not only for predicting geophysical fields, but also for filling in missing data. The constructed curve  $\Delta T(p)$  of the P-wave interval time can be considered as a forecasted result, if acoustic logging is expected later; if additional acoustic logging is not possible, then the synthetic curve  $\Delta T(p)$  replaces the log-derived one for further interpretation. The RF method is shown to provide the best test results.

**Keywords:** machine learning, regression problem, missing data replacement, acoustic logging

**REFERENCES**

1. Caté A., Perozzi L., Gloaguen E., Blouin M. *The Leading Edge*, 2017, no. 3(36), pp. 215–219.
2. Pandey S., Saraiya R. *Marine Acquisition Workshop 2018*, 2018.
3. Mas J.F., Flores J.J. *Intern. Journal of Remote Sensing*, 2008, no. 3(29), pp. 617–663.
4. Rumelhart D.E., Hinton G.E., Williams R.J. *Nature*, 1986, no. 6088(323), pp. 533–536.
5. Akinnikawe O., Lyne S., Roberts J. *Unconventional Resources Technology Conf.*, Houston, Texas, July 23–25, 2018, pp. 1507–1522.
6. Cracknell M.J., Reading A.M. *Computers & Geosciences*, 2014, vol. 63, pp. 22–33.
7. Rodriguez-Galiano V., Sanchez-Castillo M., Chica-Olmo M., Chica-Rivas M.J.O.G.R. *Ore Geology Reviews*, 2015, vol. 71, pp. 804–818.
8. Yilmaz I., Kaynar O. *Expert Systems with Applications*, 2011, no. 5(38), pp. 5958–5966.
9. Gupta I., Devegowda D., Jayaram V., Rai C., Sondergeld C. *Interpretation*, 2019, no. 3(7), pp. 1–56.
10. Onalo D., Oloruntobi O., Adedigba S., Khan F., James L., Butt S. *Journal of Petroleum Science and Engineering*, 2019, vol. 175, pp. 1049–1062.
11. Takagi T., Sugeno M. *IEEE Transactions on Systems, Man and Cybernetics*, 1985, no. 1, pp. 116–132.
12. Jang J.S.R. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, no. 3(23), pp. 665–685.
13. Karaboga D., Kaya E. *Artificial Intelligence Review*, 2019, no. 4(52), pp. 2263–2293.
14. Cabalar A.F., Cevik A., Gokceoglu C. *Computers and Geotechnics*, 2012, vol. 40, pp. 14–33.
15. Akhmetsafin R.D., Akhmetsafina R.Z. *Geofizicheskiy vestnik*, 2016, no. 1, pp. 11–14. (in Russ.)
16. Basarir H. *Intern. Journal of Mining, Reclamation and Environment*, 2019, no. 1(33), pp. 61–74.
17. Babuška R. *Fuzzy modeling for control*, Springer Science & Business Media, 2012, vol. 12.
18. Gustafson D.E., Kessel W.C. *IEEE Conf. on Decision and Control including the 17th Symposium on Adaptive Processes*, 1978, vol. 17, pp. 761–766.
19. Grima A. M., Babuška R. *Intern. Journal of Rock Mechanics and Mining Sciences*, 1999, no. 3(36), pp. 339–349.
20. Kieu D.T., Kopic A., Kitzig M.C. *Geophysical Prospecting*, 2018, no. 9(66), pp. 1667–1683.
21. Vapnik V. *The nature of statistical learning theory*, Springer Science & Business Media, 2013.
22. Chang C.C., Lin C.J. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, no. 3(2), pp. 27.
23. Al-Anazi A.F., Gates I.D. *Computers & Geosciences*, 2012, vol. 39, pp. 64–76.
24. Bagheripour P., Gholami A., Asoodeh M., Vaezzadeh-Asadi M. *Journal of Petroleum Science and Engineering*, 2015, vol. 125, pp. 95–99.

25. Huang G.B., Zhou H., Ding X., Zhang R. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2011, no. 2(42), pp. 513–529.
26. Olatunji S.O., Selamat A., Raheem A.A.A. *2010 Fourth Asia Intern. Conf. on Mathematical/Analytical Modelling and Computer Simulation*, IEEE, 2010, pp. 29–33.
27. Shi X., Wang J., Liu G., Yang L., Ge X., Jiang S. *Journal of Natural Gas Science and Engineering*, 2016, vol. 33, pp. 687–702.
28. Cao J., Shi Y., Wang D., Zhang X. *Journal of Electrical and Computer Engineering*, 2017, vol. 2017.
29. Hultquist C., Chen G., Zhao K. *Remote Sensing Letters*, 2014, no. 8(5), pp. 723–732.
30. Rasmussen C.E., Williams C.K.I. *Gaussian Processes for Machine Learning* the MIT Press, Cambridge, MA, 2006.
31. Asante-Okyere S., Shen C., Yevenyo Ziggah Y., Moses Rulegeya M., Zhu X. *Energies*, 2018, no. 12(11), pp. 3261.
32. Bukar I., Adamu M.B., Hassan U. *SPE Nigeria Annual Intern. Conf. and Exhibition, Society of Petroleum Engineers*, 2019.
33. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media, 2009.
34. Pratama H. *The 13th SEGJ Intern. Symposium*, Tokyo, Japan, November 12–14, 2018, Society of Exploration Geophysicists and Society of Exploration Geophysicists of Japan, 2019, pp. 538–541.
35. Friedl M.A., Brodley C.E. *Remote sensing of environment*, 1997, no. 3(61), pp. 399–409.
36. Lee S., Jung H.S. *Machine Learning Techniques Applied to Geoscience Information System and Remote Sensing*, MDPI, 2019, 438 p., ISBN 978-3-03921-215-6.
37. Breiman L. *Machine Learning*, 2001, no. 1(45), pp. 5–32.
38. Friedman J.H. *Annals of Statistics*, 2001, pp. 1189–1232.

**Data on authors**

- |                              |   |
|------------------------------|---|
| <b>Rais D. Akhmetsafin</b>   | — Dr. Sci., Associate Professor; MIPT Engineering Center for Minerals, Department of Digital Technologies in Industry;<br>E-mail: akhmetsafinrd@mail.ru |
| <b>Rimma Z. Akhmetsafina</b> | — PhD, Associate Professor; HSE University, School of Software Engineering, Faculty of Computer Science;<br>E-mail: rakhmetsafina@hse.ru                |

**For citation:** Akhmetsafin R. D., Akhmetsafina R. Z. Applying machine learning methods to predict or replace missing logging data. *Journal of Instrument Engineering*. 2021. Vol. 64, N 7. P. 532–541 (in Russian).

DOI: 10.17586/0021-3454-2021-64-7-532-541