

<https://doi.org/10.23947/2587-8999-2023-6-1-70-76>

Машинное обучение в анализе влияния электромагнитного поля на скорость коррозии и солеотложения нефтепромыслового оборудования

Ш. Р. Хуснуллин¹ ✉, К. Ф. Коледина^{1,2}, С. Р. Алимбекова³, Ф. Г. Ишмуратов³

¹Уфимский государственный нефтяной технический университет, Российская Федерация, г. Уфа, ул. Космонавтов, 1

²Институт нефтехимии и катализа УФИЦ РАН, Российская Федерация, г. Уфа, пр. Октября, 141

³Уфимский университет науки и технологий, Российская Федерация, г. Уфа, ул. Карла Маркса, 12

✉ shamil.khusnullin@gmail.com

Аннотация

Введение. Образование солеотложений и коррозия нефтепромыслового оборудования на большинстве нефтяных месторождений в последние годы получила особую актуальность ввиду роста объемов добываемой нефти и увеличения ее обводненности. Отложение солей в пласте и скважинах приводит к снижению проницаемости нефтеносного пласта, дебита скважин. Целью работы является применение алгоритмов машинного обучения для моделирования воздействия электромагнитного поля на процессы солеотложений и коррозии. Предсказание результатов экспериментов позволит быстрее и точнее проводить опыты, устанавливающие влияние электромагнитных полей на процессы коррозии и отложения солей.

Материалы и методы. Для обучения моделей были использованы три группы данных, различающихся по составу изучаемого исходного модельного солевого раствора: воды Вынгапуровского и Приобского месторождений, а также водопроводная вода. Были рассмотрены следующие модели машинного обучения: линейная регрессия с регуляризацией Elastic-Net, метод k ближайших соседей, дерево решений, случайный лес и полносвязная нейросеть.

Результаты исследования. С помощью алгоритмов машинного обучения были смоделированы процессы воздействия электромагнитного поля на образования солеотложений и коррозию нефтепромыслового оборудования. Разработана программа на Python для предсказания выходных результатов экспериментов. Проведено моделирование с различными моделями и их параметрами.

Обсуждение и заключение. Из проведенных экспериментов установлено, что наилучшую точность предсказаний имеют дерево решений и случайный лес. Нейронные сети, напротив, предсказывают с наименьшей точностью. Связано это с тем, что данных в обучающих выборках слишком мало. С увеличением числа наблюдений стоит использовать нейросети различных архитектур.

Ключевые слова: солеотложение, электромагнитное воздействие, коррозия нефтепромыслового оборудования, методы множественной регрессии, машинное обучение, нейронная сеть, случайный лес.

Для цитирования. Машинное обучение в анализе влияния электромагнитного поля на скорость коррозии и солеотложения нефтепромыслового оборудования / Ш. Р. Хуснуллин, К. Ф. Коледина, С. Р. Алимбекова, Ф. Г. Ишмуратов // Computational Mathematics and Information Technologies. — 2023. — Т. 6, № 1. — С. 70–76. <https://doi.org/10.23947/2587-8999-2023-6-1-70-76>

Original article

Machine learning in the analysis of the electromagnetic field influence on the rate of oilfield equipment's corrosion and salt deposition

Sh. R. Khusnullin¹, K. F. Koledina^{1,2}, S. R. Alimbekova³, F. G. Ishmuratov³

¹Ufa State Petroleum Technical University, 1, Kosmonavtov St., Ufa, Russian Federation

²Institute of Petrochemistry and Catalysis of Russian Academy of Sciences, 141, October Ave, Ufa, Russian Federation

³Ufa University of Science and Technology, 12, Karl Marx St., Ufa, Russian Federation

✉ shamil.khusnullin@gmail.com

Abstract

Introduction. The formation of salt deposits and oilfield equipment's corrosion in most oil fields has become particularly relevant due to the increase in the volume of oil produced and the increase in its water content in recent years. The deposition of salts in the formation and wells leads to a decrease in the permeability of the oil reservoir, the flow rate of wells. The aim of the work is to use machine learning algorithms to simulate the effects of an electromagnetic field on the processes of salt deposition and corrosion. Prediction of experimental results will allow faster and more accurate experiments to establish the influence of electromagnetic fields on the processes of corrosion and salt deposition.

Materials and methods. Three groups of data were used, to train the models, differing in the composition of the studied initial model salt solution: the waters of the Vyngapurovsk's and Priobsk's deposits, as well as tap water. The following machine learning models were used: linear regression with Elastic-Net regularization, the k-nearest neighbors algorithm, the decision tree, the random forest and a fully connected neural network.

Results. The processes of electromagnetic field influence on the formation of salt deposits and corrosion of oilfield equipment were simulated with the help of machine learning algorithms. Python program has been developed to predict the output results of experiments. Modeling with various models and their parameters is carried out.

Discussion and conclusions. It was found that the decision tree and the random forest have the best accuracy of predictions, from the experiments conducted. This is due to the fact that there is too little data in the training samples. With the increase in the number of observations, it is worth using neural networks of various architectures.

Keywords: salt deposition, electromagnetic impact, oilfield equipment's corrosion, multiple regression methods, machine learning, neural network.

For citation. Machine learning in the analysis of the electromagnetic field influence on the rate of oilfield equipment corrosion and salt deposition / Sh. R. Khusnullin, K. F. Koledina, S. R. Alimbekova, F. G. Ishmuratov // Computational Mathematics and Information Technologies. — 2023. — Vol. 6, no. 1. — P. 70–76.

<https://doi.org/10.23947/2587-8999-2023-6-1-70-76>

Введение. В последнее время особенно остро стоит проблема образования солеотложений и коррозии нефтепромышленного оборудования на большей части активно разрабатываемых нефтяных месторождений. Это связано с наращиванием объемов добычи нефти и увеличением ее обводненности. Отложение солей в пласте и скважинах приводит к снижению проницаемости нефтеносного пласта, дебита скважин, увеличению эксплуатационных расходов и отказу глубинно-насосного оборудования [1].

В НИИ ТС «Пилот» проводились эксперименты по изучению влияния электромагнитного поля, генерируемого РВК, на коррозию конструкционной низкоуглеродистой стали и на кристаллизацию карбоната кальция. Было установлено, что воздействие ЭМП уменьшает общую массу малорастворимых солей, а также обеспечивает защиту от коррозии [2, 3].

В данной работе предлагается применение алгоритмов машинного обучения для моделирования воздействия ЭМП на кристаллизацию карбоната кальция и на процесс коррозии конструкционной стали. Объектом исследования стало влияние электромагнитных полей на процессы солеотложения и коррозии. Предметом изучения и анализа является возможность применения моделей машинного обучения для моделирования процессов солеотложения и коррозии под воздействием электромагнитных полей. Цель настоящего исследования — разработка программного обеспечения для моделирования экспериментов по взаимодействию электромагнитных полей на отложения солей и коррозию. Прогноз результатов даст возможность проще и быстрее проводить подобные эксперименты. Это позволит точнее установить влияние магнитного поля на отложения карбоната кальция и на скорость коррозии нефтепромышленного оборудования.

Материалы и методы. Экспериментальные исследования электромагнитного воздействия на процессы солеотложения и коррозии были проведены для воды Вынгапуровского и Приобского месторождений, а также для водопроводной воды [4]. Для обучения алгоритмов были использованы 3 группы данных, различающихся по составу изучаемого исходного модельного солевого раствора. Анализируемые данные включают в себя информацию о составе растворов, условиях проведения исследований (скорость потока, давление, температура), па-

раметрах электромагнитных полей, воздействующих на растворы. Выходные параметры, которые необходимо прогнозировать: скорость коррозии в потоке и статичном растворе, мм/год; распределение карбоната кальция по морфологии (кальцит, арагонит, ватерит). Для каждой группы отчетов применялись различные модели, так как в каждой задаче различные входные и выходные данные [5].

Для решения поставленной задачи машинного обучения в анализе влияния электромагнитного поля на скорость коррозии и солеотложения нефтепромыслового оборудования были применены алгоритмы обучения, рассмотренные ниже.

Результаты исследования.

1. Линейная регрессия с регуляризацией Elastic Net.

Представляет собой линейную зависимость между целевой переменной и признаками:

$$\hat{y} = w_0 + w_1 * x_1 + \dots + w_D * x_D = \langle x, w \rangle + w_0 \quad [6].$$

Для нахождения оптимальных весов, минимизирующих среднеквадратичную функцию потерь:

$$MSE = \frac{1}{N} * \sum_i^N (y^i - \hat{y}^i)^2. \quad (1)$$

На практике используется метод градиентного спуска.

Для борьбы с переобучением применяется метод регуляризации Elastic Net. В функцию потерь добавляется 1-ая и 2-ая нормы весов w :

$$L(y, \hat{y}, w) = \frac{1}{N} * \sum_i^N (y^i - \hat{y}^i)^2 + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2, \quad (2)$$

где λ_1 и λ_2 — коэффициенты регуляризации.

2. Метод k ближайших соседей.

Суть метода заключается в следующем: предсказание \hat{y} для объекта вычисляется как среднее значение целевой переменной y среди k его ближайших соседей [7]:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^k y^i. \quad (3)$$

Для нахождения расстояния между объектами используются различные метрики (функции расстояния):

– евклидово расстояние: $\rho(a, b) = \sqrt{\sum_i (a_i - b_i)^2}$; (4)

– манхэттенская метрика: $\rho(a, b) = \sum_i |a_i - b_i|$; (5)

– косинусное расстояние: $\rho(a, b) = 1 - \frac{a * b}{|a||b|}$. (6)

3. Решающие деревья.

Решающее дерево предсказывает значение целевой переменной с помощью применения последовательности простых решающих правил (которые называются предикатами). В каждом узле этого дерева находится предикат. Если он верен для текущего примера из выборки, осуществляется переход в правого потомка, если нет — в левого. Часто предикаты — это просто взятие порога по значению какого-то признака. В листьях записаны предсказания (например, значения целевой переменной y).

4. Случайный лес.

Это ансамбль моделей (композиция нескольких алгоритмов), где в качестве базового алгоритма используются деревья решений. В основе этого метода лежит бэггинг (англ. bagging, метаалгоритм, предназначенный для улучшения стабильности решения) [8]. Суть метода в следующем: из исходной выборки получают подвыборку той же размерности методом случайного выбора объектов. На каждой выборке обучается свое дерево решений, причем используются не все признаки объектов, а их случайное количество (метод случайных подпространств). Чтобы получить одно предсказание, усредняются предсказания всех моделей:

$$\hat{y}(x) = \frac{1}{k} (b_1(x) + \dots + b_k(x)). \quad (7)$$

5. Нейронная сеть.

Во всех трех задачах применялись сети следующей архитектуры: входной слой из D нейронов (количество входных параметров), скрытый слой и выходной слой, состоящий из m нейронов, по размерности целевого вектора $y = (y_1, \dots, y_m)$ [9].

Для обеспечения нелинейности преобразований применялась функция активации:

$$\text{ReLU}(x) = \max(x, 0). \quad (8)$$

6. Оценка качества моделей.

Для того чтобы определить, какая модель лучше всего аппроксимирует зависимость между признаками и зависимыми переменными, в задачах регрессии используются следующие критерии (метрики):

– среднеквадратичная ошибка (MSE): $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$; (9)

– коэффициент детерминации: $R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$; (10)

– средняя абсолютная ошибка: $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$. (11)

В таблицах 1–3 приведены результаты применения моделей для трех групп экспериментов.

Таблица 1

Значения метрик для Вынгапуровского месторождения

Модель/ Метрика	Линейная регрессия	Метод k ближайших соседей	Дерево решений	Случайный лес	Нейронная сеть
MSE	0,01453	0,0133	0,0064	0,0065	20,6408
MAE	0,0839	0,0757	0,0467	0,0478	3,2109
R^2	-5,2844	-8,2677	0,3169	0,2266	-1,0221

Таблица 2

Значения метрик для Приобского месторождения

Модель/ Метрика	Линейная регрессия	Метод k ближайших соседей	Дерево решений	Случайный лес	Нейронная сеть
MSE	0,3073	0,2623	0,1435	0,1724	13,9461
MAE	0,2489	0,1972	0,1314	0,1724	2,4746
R^2	-18,8724	-2,1167	0,5364	0,1113	-0,9355

Таблица 3

Значения метрик для водопроводной воды

Модель/ Метрика	Линейная регрессия	Метод k ближайших соседей	Дерево решений	Случайный лес	Нейронная сеть
MSE	0,3232	0,2525	0,1799	0,1845	1,1029
MAE	0,4729	0,4022	0,3251	0,3362	0,8847
R^2	-6,9638	-1,7114	0,0378	-0,0283	-4,3788

На основе описанных алгоритмов разработана программа Predict — моделирование воздействия ЭМП на процессы солеотложения и коррозии на нефтепромысловом оборудовании [10] (рис. 1).

Прогноз Данные

Выберите данные: Месторождение_1

Входные данные:	
CaCl ₂ , г/л	0,567
MgCl ₂ · 6H ₂ O, г/л	0,9
NaCl, г/л	1,2
NaHCO ₃ , г/л	1,456
Частота тора, кГц	0
Частота шпули, кГц	220
Резонансная частота, кГц	0

Выходные данные:	
Кальцит, %	0,72
Арагонит, %	0,016
Ваггерит, %	0,127

Выберите модель: **Случайный лес**

Прогноз Сохранить данные

Значения метрик:	
Средняя абсолютная ошибка (MAE)	0,05
Среднеквадратичная ошибка (MSE)	0,007
Коэффициент детерминации (R ²)	0,103

Пояснение: чем меньше MSE и MAE, тем лучше. Чем ближе R² к 1, тем лучше модель.

Рис. 1. Интерфейс программы Predict

Пользователь может ввести различные входные данные (состав раствора, частоты электромагнитных полей, наличие магнитов и т. д.) и получить выходные данные — распределение карбоната кальция по морфологии и/или скорость коррозии. Для предсказания могут применяться, по выбору пользователя, следующие методы машинного обучения: линейная регрессия, метод k ближайших соседей, дерево решений, случайный лес и нейронная сеть. Для выбора лучшей модели программа выводит показания метрик качества. В качестве языка программирования был выбран Python версии 3.10. Для использования моделей машинного обучения и обработки данных применялись библиотеки numpy, pandas, sklearn. Для работы с excel-файлами использовалась библиотека openpyxl.

На рис. 2 приведены графики истинных и предсказанных значений выходных параметров для Приобского месторождения, полученные методом случайного леса. Рассмотрены скорость коррозии в потоке (истинное и прогнозируемое значение); распределение карбоната кальция в виде кальцита (истинное и прогнозируемое значение). Как видно из приведенного графика, алгоритм случайного леса довольно хорошо аппроксимирует выходные данные.

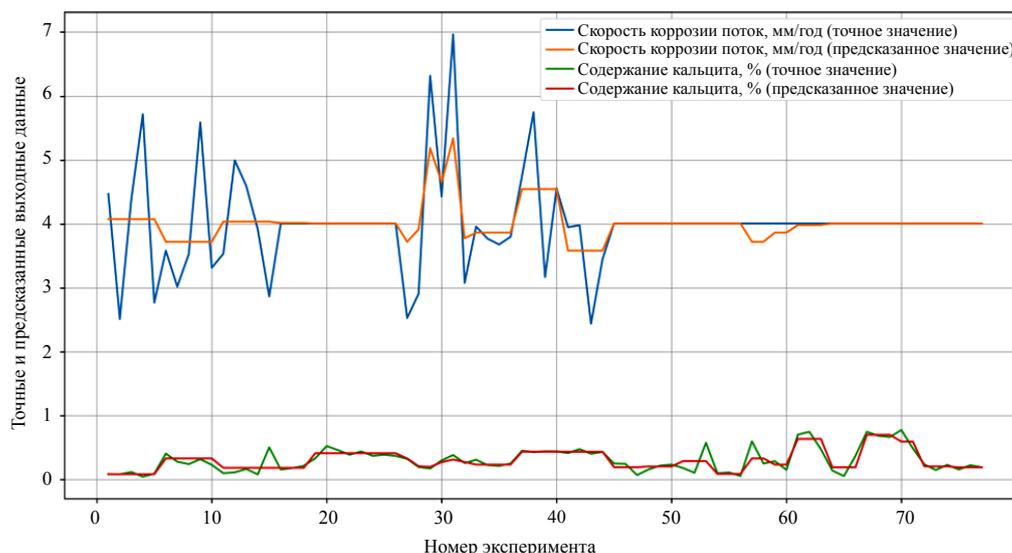


Рис. 2. График истинных и предсказанных значений методом случайного леса

Обсуждение и заключения. Как видно из результатов исследования (табл. 1–3, рис. 2), дерево решений и случайный лес имеют наименьшие значения метрик MSE и MAE и значения R^2 , наиболее близкие к 1 во всех трех задачах. Нейросети, напротив, имеют самые худшие показатели ошибок. Связано это с тем, что данных для обучения слишком мало.

Для анализа такого числа данных целесообразно использовать деревья решений и случайный лес. При увеличении количества наблюдений стоит перейти к использованию нейросетей, а также использовать иные архитектуры (с большим количеством скрытых слоев, рекуррентные сети и т. д.).

Список литературы

1. Koledina, K. F. Solving the problem of multi-criteria optimization of the synthesis reaction of benzylalkyl esters by the method of “ideal” point and lexicographic ordering / K. F. Koledina, A. A. Alexandrova // Computational mathematics and information technologies. — 2022. — Vol. 1, no. 1. — P. 12–19. <https://doi.org/10.23947/2587-8999-2022-1-1-12-19>
2. Shaimardanova, G. F. Genetic algorithm for solving the inverse problem of chemical kinetics / G. F. Shaimardanova, K. F. Koledina // Computational mathematics and information technologies. — 2022. — Vol. 1, no. 1. — P. 41–49. <https://doi.org/10.23947/2587-8999-2022-1-1-41-49>
3. Влияние электромагнитного поля на коррозию низкоуглеродистой стали в водных минерализованных средах и на кристаллизацию карбоната кальция в присутствии ионов железа(II) / С. Р. Алимбекова, Ф. Г. Ишмуратов, В. В. Носов [и др.] // SOCAR Proceedings Special Issue. — 2021. — № 1. — С. 116–124. <http://dx.doi.org/10.5510/OGP2021SI100515>
4. Физические методы предупреждения солеотложения при нефтедобыче / С. Р. Алимбекова, Р. Н. Бахтизин, А. И. Волошин [и др.] // Нефтегазовое дело. — 2019. — № 6. — С. 31–38. <http://ngdelo.ru/files/ngdelo/2019/6/ngdelo-6-2019-p31-38.pdf>
5. Weisberg, S. Applied linear regression / S. Weisberg // New Jersey : John Wiley & Sons. — 2005. — 352 p.
6. Хуснуллин, Ш. Р. Машинное обучение в анализе влияния частоты электромагнитного поля при различных скоростях потока раствора на скорость коррозии / Ш. Р. Хуснуллин, К. Ф. Коледина, С. Р. Алимбекова // В сборнике: Технические и технологические системы. Материалы тринадцатой Международной научной конференции. Краснодар. — 2022. — С. 462–464.
7. Sun, S. An adaptive k-nearest neighbor algorithm / S. Sun, R. Huang // IEEE. — 2010. — Vol. 1. — P. 91–94.
8. Чистяков, С. П. Случайные леса: обзор // Труды Карельского научного центра Российской академии наук. — 2013. — №. 1. — С. 117–136.
9. Горбачевская, Е. Н. Классификация нейронных сетей // Вестник Волжского университета им. В. Н. Татищева. — 2012. — №. 2 (19). — С. 128–134.
10. Программа моделирования воздействия электромагнитного поля на процессы солеотложения и коррозии на нефтепромысловом оборудовании (Predict) / Хуснуллин Ш. Р., Коледина К. Ф., Алимбекова С. Р. [и др.]. // Свидетельство о регистрации программ для ЭВМ от 19 января 2023. — № 2023611354.

Поступила в редакцию 08.02.2023.

Поступила после рецензирования 01.03.2023.

Принята к публикации 02.03.2023.

Об авторах:

Хуснуллин Шамиль Рамилевич, студент кафедры информационных технологий и прикладной математики Уфимского государственного нефтяного технологического университета (РФ, 450064, г. Уфа, ул. Космонавтов, 1), кафедра «Информационные технологии и прикладная математика», [ORCID](https://orcid.org/0009-0001-9111-1111), shamil.khusnullin15@gmail.com

Коледина Камила Феликсовна, доктор физико-математических наук, профессор кафедры информационных технологий и прикладной математики Уфимского государственного нефтяного технологического университета (РФ, 450064, г. Уфа, ул. Космонавтов, 1), старший научный сотрудник Института нефтехимии и катализа УФИЦ РАН (РФ, 450075, г. Уфа, пр. Октября, 141), [ORCID](https://orcid.org/0009-0001-9111-1111), koledinakamila@mail.ru

Алимбекова Софья Робертовна, с.н.с. ИЦ «Пилот», Уфимский университет науки и технологий (РФ, 450008, г. Уфа, ул. Карла Маркса, 12), [ORCID](#), ms.sofia.al@gmail.com

Ишмуратов Фарид Гумерович, зав. лабораторией МНИЛ НХ, Уфимский университет науки и технологий (РФ, 450008, г. Уфа, ул. Карла Маркса, 12), [ORCID](#), farid_ishmuratov@mail.ru

Заявленный вклад соавторов:

Хуснуллин Шамиль Рамилевич — подготовка данных, обучение алгоритмов машинного обучения. Коледина Камила Феликсовна — консультация по методам множественной регрессии. Алимбекова Софья Робертовна — предоставление исходных данных. Ишмуратов Фарид Гумерович — предоставление исходных данных.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.