

# Исследование методов машинного обучения при организации работ по добыче и обслуживанию нефтяных скважин

Е.А. Скачкова, email: elen.rome557@mail.ru<sup>1</sup>

В.В. Мокшин, email: vladimir.mokshin@mail.ru<sup>1</sup>

<sup>1</sup> КАЗАНСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. А.Н. ТУПОЛЕВА – КАИ

***Аннотация.** В данной статье были рассмотрены методы определения значимых признаков, а также методы прогнозирования для оптимизации бурения нефтяных скважин, определены их достоинства и недостатки. Для выявления наиболее значимых признаков, влияющих количество добываемой нефти, проводился отбор признаков на основе корреляционного анализа, а также сравнение и идентификация эффективных методов обучения нейронной сети. В результате сравнения был выбран метод байесовской регуляризации обратного распространения, так как он показал наименьшую ошибку распознавания при относительно коротком времени обучения. Таким образом, нейронная сеть была обучена прогнозированию количества добываемой нефти на нефтяных скважинах.*

***Ключевые слова:** машинное обучение, корреляционный анализ, байесовская регуляризация обратного распространения.*

## Введение

Оптимизация добычи - это, по сути, «управление производством», при котором вы минимизируете, максимизируете или планируете добычу нефти, газа и, возможно, воды. Например, вы можете легко максимизировать или нацелить добычу нефти и / или газа при минимизации воды, или запустить добычу нефти и газовый фактор (GOR) на заданные значения для поддержания энергии пласта. Есть множество альтернативных производственных целей. Поскольку каждая скважина, платформа и месторождение совершенно разные, предусмотрены гибкие средства контроля добычи.

Оптимизация добычи нефти жизненно важна для промышленно развитого мира. Под землей нет неограниченного запаса нефти. Это означает, что поставки нефти ограничены и должны быть максимально оптимизированы. Производители нефти должны продолжать разработку новых технологий добычи нефти из скважин. Эта труднодоступная нефть раньше осталась бы позади, когда добыча нефти была на пике.

Основные промышленно развитые страны продолжают добычу и использование сырой нефти рекордными темпами. Потребители, привыкшие к ранее низким ценам на нефть, покупали большие внедорожники и грузовики, в которых используется большое количество бензина. Поскольку экономика этих стран продолжает расти, потребности в нефтедобыче в транспортном и промышленном секторах также растут. Извлечение каждой последней капли нефти за счет оптимизации добычи помогает снизить спрос на сырую нефть.

Оптимизация добычи сырой нефти может быть достигнута несколькими способами. Буровые растворы и суспензии специального состава можно закачивать в существующие нефтяные скважины. Эта операция оптимизирует добычу нефти за счет поддержания давления в скважине, выталкивая новую нефть на поверхность. Находятся нетрадиционные источники нефти, такие как горючие сланцы и нефтеносные пески, и добывается их сырая нефть. Другие меры по оптимизации добычи включают новые методы бурения и геофизические инструменты для обнаружения и добычи неизвестных нефтяных месторождений.

### 1. Формализация задачи

Для решения задачи необходимо сгенерировать входные и выходные данные и выбрать нейронную сеть, аппроксимирующую передаточную функцию. Структура перцептрона позволяет аппроксимировать функции вида [7]:

$$y = \sum (b_i + \sum w_i \cdot x_i), \quad (1)$$

где  $x_i$  – элементы входного пространства,  $w_i$  – коэффициенты нейронной сети, полученные во время обучения,  $b_i$  – функция активации.

Многослойный перцептрон - это класс искусственных нейронных сетей прямого распространения, состоящий как минимум из трех слоев: входного, скрытого и выходного. За исключением входных нейронов, все нейроны используют нелинейную функцию активации. В качестве активационных функций нейронов используются сигмоидальные: логистический или гиперболический тангенс.

Предположим, что есть  $L$  скрытых слоев, и в каждом из них есть нейроны  $L_i$ . Вес обозначен как  $w_{ljk}$ , где  $j$  относится к  $j$ -му нейрону в  $(l-1)$ -м слое,  $k$  относится к  $k$ -му нейрону в  $l$ -м слое, а  $b$  представляет смещение. Вход обозначен как 0-й скрытый слой, а выход обозначен как  $(L+1)$ -й скрытый слой.

Процесс обратного распространения ошибки использует частную производную каждой функции активации нейрона для определения наклона (или градиента) в направлении каждого из входящих весов. Градиент показывает, насколько круто будет уменьшаться или увеличиваться ошибка при изменении веса. Обратное распространение продолжает изменять веса до тех пор, пока частота ошибок не будет значительно уменьшена на величину, известную как скорость обучения.

Вес обновляется сзади вперед; поэтому их также называют сетью обратного распространения (ВР). Алгоритм ВР - это алгоритм на основе итеративного градиентного спуска.

На каждой итерации вектор параметров  $\beta$  заменяется новой оценкой  $\beta + \delta$ . Для определения  $\delta$  функция  $f(x_i, \beta + \delta)$  аппроксимируется ее линейризацией:

$$f(x_i, \beta + \delta) \approx f(x_i, \beta) + J_i \delta, \quad (2)$$

где

$$J_i = \frac{\partial f(x_i, \beta)}{\partial \beta} \quad (3)$$

градиент (в данном случае вектор-строка) функции  $f$  по  $\beta$ .

Сумма  $S(\beta)$  квадратов отклонений имеет минимум при нулевом градиенте по отношению к  $\beta$ . Повышение первого порядка приближения  $f(x_i, \beta + \delta)$  дает:

$$S(\beta + \delta) \approx \sum_{i=1}^m (y_i - f(x_i, \beta) - J_i \delta)^2 \quad (4)$$

или в векторной записи

$$\begin{aligned} S(\beta + \delta) &\approx \|y - f(\beta) - J\delta\|^2 = \\ &= [y - f(\beta) - J\delta]^T [y - f(\beta) - J\delta] = \\ &= [y - f(\beta)]^T [y - f(\beta)] - [y - f(\beta)]^T J\delta - \\ &\quad - (J\delta)^T [y - f(\beta)] + \delta^T J^T J\delta = \\ &= [y - f(\beta)]^T [y - f(\beta)] - \\ &\quad - 2[y - f(\beta)]^T J\delta + \delta^T J^T J\delta \end{aligned} \quad (5)$$

Взяв производную от  $s(\beta + \delta)$  по  $\delta$  и установив результат равным нулю, получим:

$$(J^T J) \delta = J^T [y - f(\beta)] \quad (6)$$

где  $J$  - матрица Якоби,  $j$ -я строка которой равна  $J_j$ , а  $f(\beta)$  и  $y$  - векторы с  $j$ -й компонентой соответственно.

Матрица Якоби, как определено выше, не является (в общем случае) квадратной  $m \times n$  матрицей, где  $n$  представляет собой прямоугольную матрицу, где  $\beta$  - количество параметров (размер вектора).

Умножение матрицы  $(J^T J)$  дает требуемую квадратную  $n \times n$  матрицу, а произведение матрицы на вектор справа дает вектор размера  $n$ . Результатом является набор  $j$ -линейных уравнений, которые можно решить для  $\delta$  [7,8].

Приведем признаки для анализа, которые используются в данной работе (табл. 1).

Таблица 1

*Признаки для анализа*

<b>Признаки</b>	<b>Описание признаков</b>
x1	Номер
x2	Дата
x3	Время
x4	Объект
x5	Товар
x6	Среда
x7	Расход жидкости м3/час
x8	Дозировка ppm
x9	Скорость потока м/с
x10	Частота Гц
x11	Время прохождения жидкости час
x12	Остаточное содержание реагента ppm
x13	Температура оС
x14	pH
x15	O2 (кислород) ppm
x16	Методика определения эффективности

Окончание табл. **Ошибка! Источник ссылки не найден.**

Признаки	Описание признаков
x17	Метод контроля коррозии
x18	Начальная средняя скорость коррозии тру
x19	Средняя скорость коррозии тру
x20	Эффективность %
y1	Начальная скорость коррозии тру
y2	Общая скорость коррозии тру
y3	Питтинг
y4	Текущая эффективность

Шумовые характеристики также могут быть среди входных характеристик. Следовательно, эффективная модель может быть построена либо с учетом методов выбора признаков, либо на основе всех признаков. Но в этом случае шум может повлиять на качество финальной модели [8].

## 2. Корреляционный анализ как метод отбора признаков

Корреляция, такая как линия наилучшего соответствия, полезна для определения линейной зависимости количественных переменных. В корреляции каждое наблюдение помогает нам определить, какая из двух переменных влияет на другую и в какой степени. Переменная, которая влияет, называется исследовательской переменной, а та, которая затрагивается, называется переменной ответа. Основываясь на степени воздействия и взаимосвязи, мы можем спланировать соответствующий анализ.

Для количественной оценки корреляции между данными в качестве критерия оценки используется коэффициент корреляции в статистике. Вычисляя коэффициент корреляции, можно судить о взаимосвязи между различными выборками и степенью корреляции.

Чтобы определить критическое значение коэффициента корреляции воспользуемся критическим значением критерия Стьюдента:

$$r_{ijkpum} = \pm \sqrt{\frac{t_{ijkpum}^2}{t_{ijkpum}^2 + n - 2}}; i = 1, n; j = 1, N + M + V + L \quad (7)$$

где  $r_{i, крит}$  – критическое значение критерия Стьюдента для рекомендуемого уровня значимости  $\alpha = 0,05$ , определяемого по статистическим таблицам [9] при  $n - 2$  степенях свободы,  $n$  – количество значений в исходной выборке.

Таким образом, корреляцию выше 0,3162 и ниже  $-0,3162$  считаем значимой.

Таким образом, сделан вывод о том, что главная проблема прогнозирования заключается в проблеме отбора значимых признаков для наиболее точного обучения.

### 3. Результаты обучения

Берем полученную выборку признаков, чтобы сформировать эффективную модель, основанную на методе регуляризации байесовского обратного распространения ошибки. Делим выборку на обучающую (80%) и тестовую (20%), проводим обучение и денормализуем данные. На рисунках представлены данные, обученные по методу байесовской регуляризации обратного распространения (рис. 1-4).

Initial corrosion rate mpy: BRANN

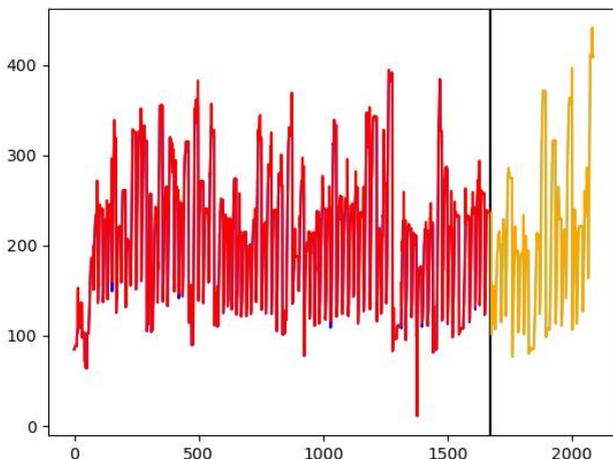


Рис. 1. Начальная скорость коррозии

General corrosion rate mpy: BRANN

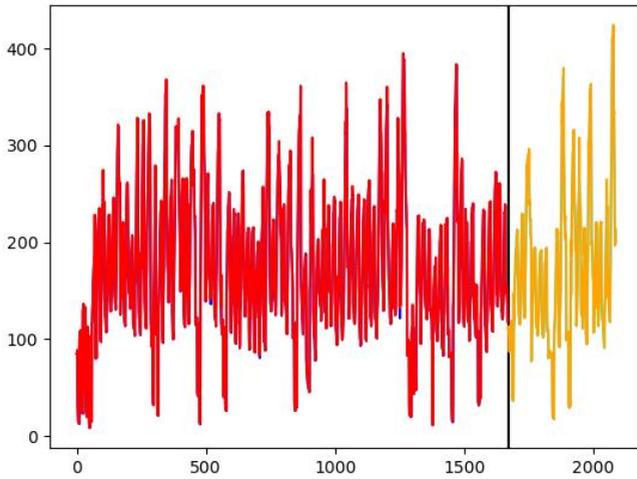


Рис. 2. Общая скорость коррозии

Pitting: BRANN

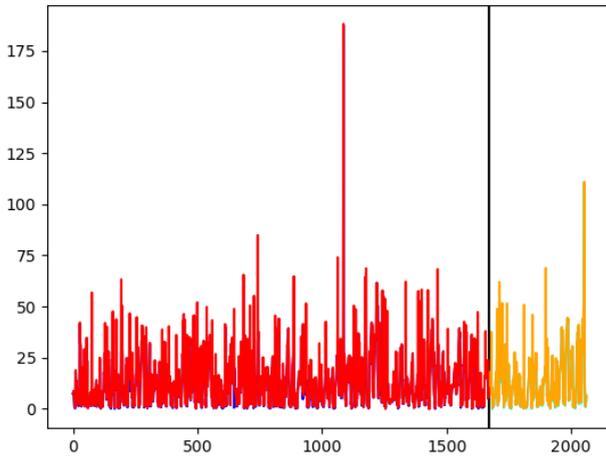


Рис. 3. ПИТТИНГ

Current efficiency: BRANN

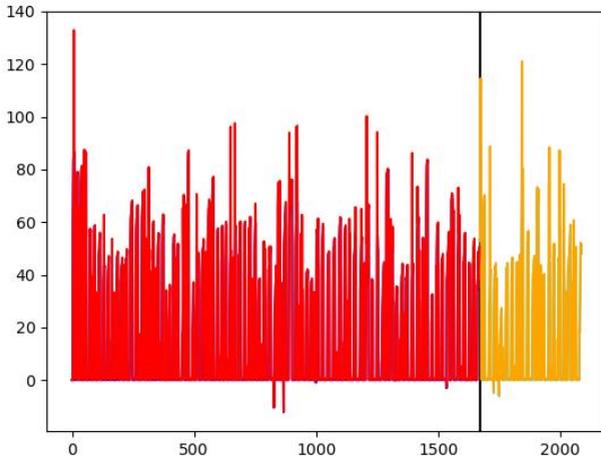


Рис. 4. Текущая эффективность

Ошибка обучения была рассчитана для разных методов обучения: метод 1 – алгоритм Левенберга-Маркварда, метод 2 – байесовская регуляризация обратного распространения, метод 3 – градиентный спуск, метод 4 – линейная регрессия, 5 – корреляционный анализ (табл. 2), выводим результаты погрешностей (рис. 5).

Таблица 2

Сравнение ошибок обучения для разных методов

Методы	Примеры значений			
	Тестовая выборка у1	Тестовая выборка у2	Тестовая выборка у3	Тестовая выборка у4
Метод 1	0.84	0.81	0.87	0.84
Метод 2	0.95	0.95	0.94	0.93
Метод 3	0.74	0.78	0.69	0.75
Метод 4	0.71	0.76	0.82	0.83
Метод 5	0.87	0.79	0.81	0.78

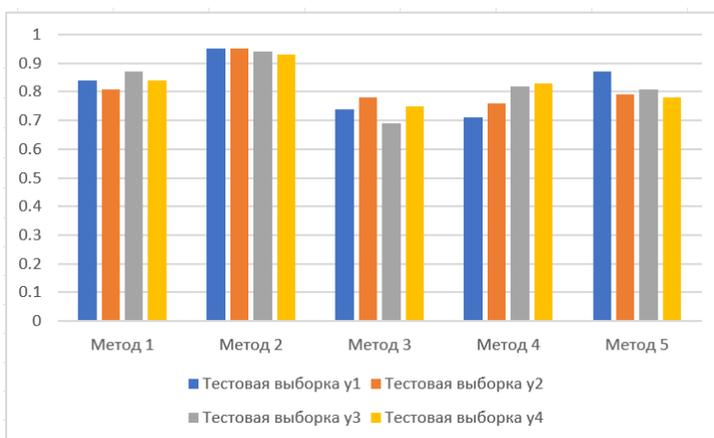


Рис. 5. Графическое сравнение ошибок обучения для разных методов

Как видно из таблицы и графика, наиболее точные результаты были получены с нейронной сетью, основанной на методе обучения байесовской регуляризации.

### Заключение

В данной работе был проведен анализ существующих методов прогнозирования добычи нефти из нефтяных скважин, предложены различные методы машинного обучения для построения модели прогнозирования показателей результата (характеристик результата) и выбран наиболее оптимальный метод регуляризации байесовского обратного распространения ошибки. Проведен анализ методов выделения значимых признаков с использованием метода корреляции. Результаты обучения были получены по каждому методу.

Кроме того, рассмотренный подход к формированию прогнозной модели может быть использован для имитационных моделей, в результате которых могут формироваться признаки [11-12], а также для оптимизации работы предприятий и различных технических систем [13-14].

### Список литературы

1. Yang X.S. Neural networks and deep learning / X.S. Yang // Introduction to Algorithms for Data Mining and Machine Learning. – 2019. – Т. 1. – С. 139–161.
2. Mokshin A.V. Adaptive genetic algorithms used to analyze behavior of complex system / A.V. Mokshin, V.V. Mokshin, L.M. Sharnin //

Communications in Nonlinear Science and Numerical Simulation. – 2019. – № 71. – С. 174–186.

3. Parallel genetic algorithm of feature selection for complex system analysis / V. V. Mokshin [и др.] // IOP Conf. Series: Journal of Physics. – 2019. – Т. 1. – С. 1092–1096.

4. Xu L. On convergence properties of the EM algorithm for Gaussian mixtures / L. Xu, M.I. Jordan // Neural Comput. – 1996. – № 8. – С. 129–151.

5. Ticknor J.L. A Bayesian regularized artificial neural network for stock market forecasting / J.L. Ticknor // Expert Syst. Appl. – 2013. – № 14. – С. 5501–5506.

6. Wali A.S. Comparative Study of Advance Smart Strain Approximation Method Using Levenberg-Marquardt and Bayesian Regularization Backpropagation Algorithm / A.S. Wali, A. Tyagi // Materials Today: Proceedings. – 2020. – № 21. – С. 1380–1395.

7. Pan H. Annealed gradient descent for deep learning / H. Pan [и др.] // Neurocomputing. – 2020. – № 380. – С. 201–211.

8. Zhang T. Adaptive forward–backward greedy algorithm for learning sparse representations / T. Zhang // IEEE Transactions on Information Theory. – 2011. – № 57. – С. 4689–4708.

9. Mokhtia M. Feature selection based on regularization of sparsity based regression models by hesitant fuzzy correlation / M. Mokhtia, M. Eftekhari, F. Saberi-Movahed // Applied Soft Computing. – 2020. – № 91. – С. 376–381.

10. Peng H. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy / H. Peng, F. Long, C. Ding // IEEE Trans Pattern Anal Machine Intell. – 2005. – Т. 1. – С. 1226–1238.

11. Yakimov I.M. The Modelling and Optimization of Machine Management System with Computer Numerical Control / I.M. Yakimov [и др.] // Lecture Notes in Electrical Engineering. – 2020. – Т. 1. – С. 724–731.

12. Mokshin V.V. Modified genetic algorithm as a new approach for solving the problem of 3d packaging / V.V. Mokshin Yakimov [и др.] // CEUR Workshop Proceedings. – 2020. – Т. 1. – С. 91–97.

13. Yakupov D.T. Modified spectral clustering method for graphs decomposition. / D.T. Yakupov, V.V. Mokshin // CEUR Workshop Proceedings. – 2020. – Т. 1. – С. 85–90.

14. Tutubalin P.I. The Evaluation of the cryptographic strength of asymmetric encryption algorithms / P.I. Tutubalin, V.V. Mokshin // Second Russia and Pacific Conference on Computer Technology and Applications. – 2017. – Т. 1. – С. 180–183.