

Исследование методов машинного обучения при анализе добычи и обслуживания нефтяных скважин

Е.А. Скачкова
Казанский национальный
исследовательский технический
университет им А. Н. Туполева -
КАИ
Казань, Россия
elen.rome557@mail.ru

А.Л. Аленин
Казанский национальный
исследовательский технический
университет им А. Н. Туполева -
КАИ
Казань, Россия
freadvice.ru@gmail.com

В.В. Мокшин
Казанский национальный
исследовательский технический
университет им А. Н. Туполева -
КАИ
Казань, Россия
vladimir.mokshin@mail.ru

Аннотация—В данной статье были рассмотрены методы определения значимых признаков, а также методы прогнозирования для оптимизации бурения нефряных 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 был выбран метод BRANN (искусственная нейронная сеть с байесовской регуляризацией), так как он показал лучшие результаты при относительно коротком времени обучения. Таким образом, нейронная сеть была обучена прогнозированию количества добываемой нефти на нефтяных скважинах.

Ключевые слова— машинное обучение, корреляционный анализ, искусственная нейронная сеть с байесовской регуляризацией.

1. ВВЕДЕНИЕ

Оптимизация добычи, наряду с управлением резервуаром, является центральной частью стратегии компании по разработке месторождений и добыче полезных ископаемых. Ключевым фактором оптимизации добычи является возможность снижения повреждения пласта при строительстве скважин и проведении регламентных работ. Уменьшение повреждения пласта может быть достигнуто, гарантируя, что эксплуатационные детали будут достигнуты до достижения продуктивной зоны до последних зарегистрированных параметров добычи.

Деятельность по оптимизации производства включает: управление профилем призабойной зоны, газоводяной конус и аппликатура, скважина с максимальным контактом с коллектором и многоствольным заканчиванием, предотвращение отложений органических и неорганических твердых частиц в призабойной зоне/заканчивании/трубопроводе, целостность скважины, предотвращение и устранение разрушения обсадной колонны и цемента, оптимизация производительности механизированной добычи на уровне месторождения и скважины, проектирование наземных сооружений и мощностей по перекачке жидкостей, устранение узких мест производственной системы [1].

Оптимизация производства позволяет повысить производительность существующего поля, и

использование охватывает несколько областей, представляющих интерес. В этом контексте вклад может исходить от применения различных технологий.

2. МЕТОДЫ РЕАЛИЗАЦИИ

Искусственные нейронные сети напоминают нейронную сложность биологического мозга, состоящую из произвольного количества узлов, каждый из которых представляет биологический нейрон. Каждый узел имеет произвольное количество входных и выходных потоков данных, оба из которых соединяют узел с другими узлами в нейронной сети. Каждый узел содержит одну функцию активации, которая вычисляет новое значение для отправки соседним нейронам. Это эквивалентно электрической передаче его биологического аналога [2]. Кроме того, узел также содержит функцию интегрирования, которая преобразует количество входных значений в одно значение. Затем это значение передается в функцию активации.

Два входа подаются в байесовский алгоритм, который представляет собой матрицу $N \times M$ из набора данных. Байесовский классификатор — это простой и эффективный алгоритм классификации, который помогает в разработке быстрых моделей машинного обучения, способных делать быстрые прогнозы. Стандартные сети с обратным распространением менее надежны, чем искусственные нейронные сети с байесовской регуляризацией (BRANN), которые могут свести к минимуму или устранить необходимость длительной перекрестной проверки.

Перед обучением гиперпараметры и веса (w) рассматриваются как случайные величины. Позже это оптимизируется с использованием байесовского подхода к регуляризации. После сбора данных функция плотности для весов может быть изменена с использованием закона Байеса:

$$P(w|D, \alpha, \beta, M) = \frac{P(D|w, \beta, M)P(w|\alpha, M)}{P(D|\alpha, \beta, M)}, \quad (1)$$

где w - вектор весов сети, M - конкретная используемая модель нейронной сети, а D - набор данных.

Поскольку набор для проверки не требуется, итеративная процедура самосогласования должна быть запущена только один раз, чтобы получить «наиболее обобщаемую» модель. Опыт показывает, что меры предосторожности, состоящие в том, чтобы выполнить

всю процедуру пять раз, достаточно, чтобы избежать любого аномального поведения. Это можно противопоставить возможным сотням или тысячам повторных вычислений с нерегулярными искусственными нейронными сетями [5].

Функция правдоподобия весов сети w описывает, насколько вероятно появление данного набора данных при заданном наборе весов:

$$P(w|D, \alpha, \beta, M) = \frac{1}{Z_w(\alpha)Z_D(\beta)} \exp(-(\beta E_D + \alpha E_w)). \quad (2)$$

Метод максимального правдоподобия выбирает веса для максимизации функции правдоподобия, которая в этом гауссовском случае совпадает с минимизацией квадрата ошибки E_D .

Для количественной оценки корреляции между данными в качестве критерия оценки используется коэффициент корреляции в статистике. Вычисляя коэффициент корреляции, можно судить о взаимосвязи между разными выборками и степенью корреляции. Вычисляем коэффициент корреляции Стьюдента и имеем: корреляцию выше 0,3162 и ниже -0,3162 считаем значимой. Таким образом, признаки, имеющие такие значения считаем значимыми [8].

Входными параметрами являются: Номер, Дата, Время, Объект, Товар, Среда, Расход жидкости м3/час, Дозировка ppm, Скорость потока м/с, Частота Гц, Время прохождения жидкости час, Остаточное содержание реагента ppm, Температура оС, pH, O2 (кислород) ppm, Методика определения эффективности, Метод контроля коррозии, Начальная средняя скорость коррозии mpy, Средняя скорость коррозии mpy, Эффективность %.

Выходными параметрами являются: Начальная скорость коррозии mpy, Общая скорость коррозии mpy, Питтинг, Текущая эффективность.

На рисунке (рис. 1) представлены данные, обученные по методу искусственной нейронной сети с байесовской регуляризацией.

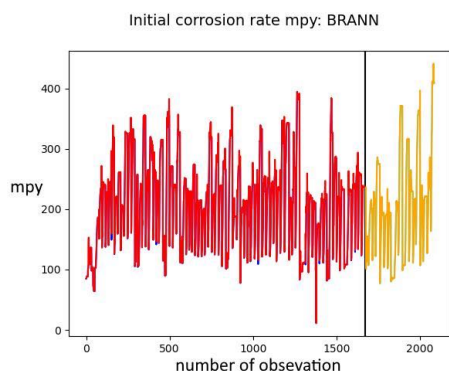


Рис. 1. Начальная скорость коррозии

Берем полученную выборку признаков, чтобы сформировать эффективную модель, основанную на методе искусственной нейронной сети с байесовской регуляризацией. Делим выборку на обучающую (80%) и тестовую (20%), проводим обучение и денормируем данные. Имеем следующие значения точности для каждого из выходных параметров: для параметра Начальная скорость коррозии mpy – 0.95, для параметра Общая скорость коррозии mpy – 0.95, для параметра Питтинг – 0.94, для параметра Текущая эффективность – 0.93. Таким образом, можно говорить о высокой точности выбранного метода.

3. ЗАКЛЮЧЕНИЕ

В 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 искусственной нейронной сети с байесовской регуляризацией. Проведен анализ методов выделения значимых признаков с использованием метода корреляции. Результаты обучения были получены по каждому методу.

ЛИТЕРАТУРА

- [1] Ticknor, J.L. A Bayesian regularized artificial neural network for stock market forecasting / J.L. Ticknor // *Expert Syst. Appl.* – 2013. – Vol. 14. – P. 5501-5506.
- [2] Xu, L. On convergence properties of the EM algorithm for Gaussian mixtures / L. Xu, M.I. Jordan // *Neural Comput.* – 1996. – Vol. 8. – P. 129-151.
- [3] 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. – Vol. 1. – P. 1226-1238.
- [4] 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. – Vol. 71. – P. 174-186.
- [5] Yakupov, D.T. Modified spectral clustering method for graphs decomposition. / D.T. Yakupov, V.V. Mokshin // *CEUR Workshop Proceedings.* – 2020. – Vol. 1. – P. 85-90.
- [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. – Vol. 21. – P. 1380-1395.
- [7] Pan, H. Annealed gradient descent for deep learning / H. Pan // *Neurocomputing.* – 2020. – Vol. 380. – P. 201-211.
- [8] Zhang, T. Adaptive forward-backward greedy algorithm for learning sparse representations / T. Zhang // *IEEE Transactions on Information Theory.* – 2011. – Vol. 57. – P. 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. – Vol. 91. – P. 376-381.
- [10] 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. – Vol. 1. – P. 180-183.