

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ СКВАЖИН-КАНДИДАТОВ НА ГЕОЛОГО-ТЕХНИЧЕСКИЕ МЕРОПРИЯТИЯ ПО ПРИЧИНАМ ОТКЛОНЕНИЯ

Р. Ф. ХАКИМОВ

wilmgc@gmail.com

ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

Аннотация. В статье рассмотрена методика подготовки данных, обучения и верификации нейросетевого классификатора для задачи классификации скважин-кандидатов на геолого-технические мероприятия. Описаны такие основные этапы как: нормализация данных, разбиение исходных данных на тестовую и обучающую выборки, балансировка данных. Статья затрагивает такие методы нормализации как минимаксная нормализация, метод унитарного кода, методы балансировки данных SMOTE, Random Oversampling. В качестве классификатора рассмотрена модель многослойного перцептрона. Для эксперимента использованы данные о скважинах-кандидатах на геолого-техническое мероприятие вывода скважина из бездействия. В статье так же приведена оценка полученного результата, выполненная с помощью такого метода верификации как процедура скользящего контроля.

Ключевые слова: добыча нефти; геолого-технические мероприятия скважин; нормализация данных; балансировка данных; машинное обучение; методы классификации; верификация данных; нейронные сети; oil production; geological and technical actions for wells; data normalization; data balancing; machine learning; classification methods; data verification; neural networks.

ВВЕДЕНИЕ

При эксплуатации нефтяных скважин для увеличения производительности и повышения экономической эффективности скважины применяются геолого-технические мероприятия (ГТМ) - комплекс мер геологического, технологического, технического и экономического характера. Существует несколько видов ГТМ, таких как: гидроразрыв пласта, обработка призабойной зоны, смена частоты ЭЦН, оптимизация, вывод из бездействия.

При подборе скважин-кандидатов на определенное мероприятие используются различные параметры скважины, такие как пластовое давление, забойное давление, текущие дебиты жидкости и нефти, вязкости и другие, а также статистические данные скважины.

Автоматизация первичного подбора скважин на геолого-технические мероприятия с помощью информационных систем позволяет значительно повысить эффективность и оперативность данного процесса [1].

После процесса первичного подбора скважин происходит процесс согласования ГТМ для каждой предложенной скважины - специалист либо согласовывает ГТМ либо отклоняет с указанием причины. Причины отклонения могут быть связаны с недостатком данных, аномальными данными, либо организационными причинами. Список причин отклонения может отличаться для каждого вида ГТМ.

Процесс согласования и отклонения геолого-технических мероприятий для скважин является трудоемким ввиду большого количества скважин и причин отклонения. Автоматическая

классификация скважин по причинам отклонения может существенно повысить оперативность работы специалистов.

В данной статье рассмотрен подход к разработке системы классификации скважин-кандидатов на геолого-технические мероприятия по причинам отклонения.

ПОДХОД К РЕШЕНИЮ ЗАДАЧИ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ

Для решения задачи классификации скважин-кандидатов по причинам отклонения использован метод классификации с помощью нейронной сети. Обучающая выборка построена на основе данных из системы согласования скважин-кандидатов на геолого-технические мероприятия.

Исходными параметрами алгоритма является вектор параметров для каждой i -ой скважины $X^i = \{x_1^i, x_2^i, \dots, x_n^i\}$, который включает в себя такие данные как:

- Остановочные параметры скважины (дебит жидкости, нефти, забойное давление, пластовое давление, давление насыщения, линейное давление, обводненность)
- Потенциальные параметры скважины (дебит жидкости, нефти, обводненность, забойное давление)
- Параметры экономической эффективности скважины (Суммарные затраты на ГТМ, NPV, PI)
- Предыдущая причина отклонения (если возможно)
- Признак месторождения

На основе этих данных алгоритм классифицирует скважину в один из 11 классов Y_i (ГТМ принят, низкая экономическая эффективность, аварийный фонд, некорректные данные, и т.д.).

В связи с тем, что входные данные алгоритма имеют разные типы и области значений, данные в исходном виде могут оказывать разное влияние на обучение нейронной сети [2]. Для того чтобы избавиться от этого фактора применяется нормализация данных. Среди входных параметров выделяются:

- Численные (Дебиты, давления, обводненность, экономические параметры)
- Категориальные (Признак месторождения, предыдущая причина отклонения)

Для нормализации численных параметров применяется метод минимаксной нормализации - $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$ [3]. Для нормализации категориальных параметров применяется метод унитарного кода, при котором каждому из возможных значений признака сопоставляется отдельный бинарный признак [4].

В наборах данных часто оказывается так, что какие-либо классы присутствуют в большем количестве чем другие. Такие наборы данных могут негативно влиять на обучение нейронной сети, так как более присутствующие классы будут оказывать большее влияние на обучение чем менее присутствующие [5, 6]. Для решения проблемы несбалансированности данных в обучающей выборке используется алгоритм SMOTE (Synthetic Minority Oversample Technique), суть которого заключается в генерации искусственных экземпляров миноритарного класса [7]. Искусственные экземпляры генерируются в “соседних” областях с помощью алгоритма ближайшего соседа (KNN).

При разбиении исходных данных на обучающую и тестовую выборки в задаче многоклассовой классификации важно чтобы каждый класс был представлен равно как в обучающей, так и в тестовой выборке, иначе классификатор может иметь недостаточно данных для обучения либо для проверки. Для равномерного распределения классов среди обучающей и тестовой выборок используется алгоритм стратифицированного разделения.

Таким образом, исходное множество X делится на N подмножеств $X_k \subseteq X, k \leq N$, где N - количество классов, в соответствии с принадлежностью к классу k . После этого из каждого подмножества случайным образом выбирается $T_k = |X_k| * h$ элементов, где h - коэффициент разделения, для тестовой выборки $X^{test} = \cup_i^N T_i$, а из оставшихся элементов формируется обучающая выборка - $X^{train} = X \setminus X^{test}$.

Для решения поставленной задачи используется архитектура многослойного перцептрона, со следующими слоями (Рисунок 1):

- Входной слой, 36 элементов
- Скрытый слой, 128 элементов
- Скрытый слой, 128 элементов
- Выходной слой, 11 элементов

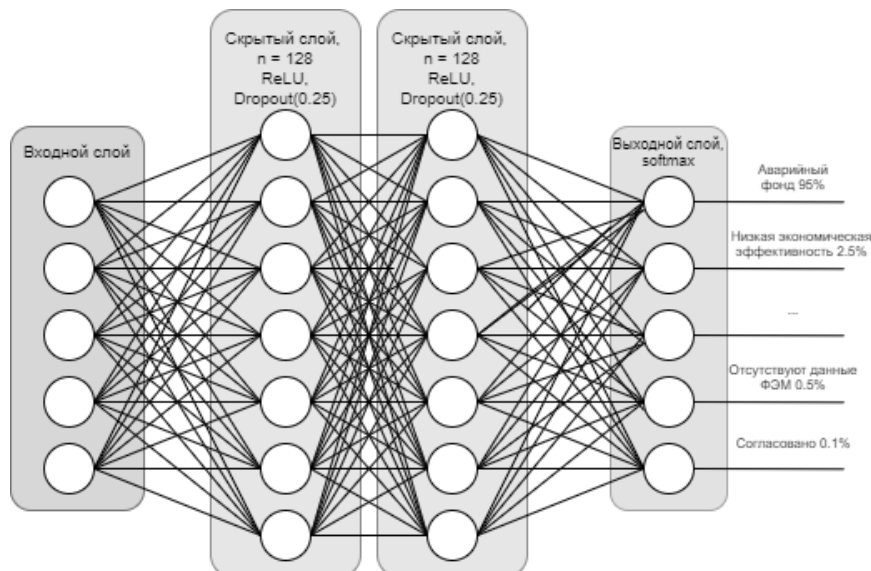


Рис. 1. Архитектура нейронной сети

В скрытых слоях используется функция активации линейного выпрямителя (ReLU) - $f(x) = \begin{cases} 0, & x < 0 \\ x, & x > 0 \end{cases}$. Для выходного слоя используется функция активации Softmax - $f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$. Областью определения функции Softmax является $(0,1)$, и ее результат можно интерпретировать как вероятность попадания в заданный класс.

Для того чтобы предотвратить переобучение нейронной сети, используется метод регуляризации Dropout, суть которого заключается в отключении определенного количества случайных нейронов слоя на каждом шаге обучения. В данном случае метод Dropout применяется к скрытым слоям нейронной сети.

Для реализации заданной модели использовался язык программирования Python, а также пакет keras для реализации нейросетевой модели и пакет sklearn для обработки исходных данных и валидации результатов. API-служба реализована на языке Python с помощью фреймворка Flask.

По результатам тестирования описанной модели, средняя оценка точности по метрикам precision, accuracy и recall составила 78%, 90% и 76%. Полученная точность позволяет использовать результаты классификации для рекомендации пользователю наиболее подходящих причин отклонения. Для повышения точности классификации по этой причине отклонения необходима информация о бригадах, которой в данном исследовании не было в исходной выборке.

ЗАКЛЮЧЕНИЕ

Процесс согласования и отклонения скважин-кандидатов на геолого-технические мероприятия является трудоемким и ресурсозатратным, автоматизация некоторых аспектов данного процесса может повысить его эффективность и надежность.

В ходе исследования реализовано программное решение для классификации скважин-кандидатов на геолого-техническое мероприятие «Вывод из бездействия» по причинам отклонения с помощью нейросетевого классификатора. Результаты эксперимента показали, что нейронная сеть с заданными параметрами решает задачу классификации скважин с точностью, достаточной для отображения оператору в качестве предложения для принятия решения. Увеличение количества входных параметров о скважине, таких как: информация о предыдущих мероприятиях, информация о мероприятиях на данном пласте может существенно увеличить точность классификатора.

СПИСОК ЛИТЕРАТУРЫ

1. А.Н. Ситников, Р.Н. Асмандияров, А.А. Пустовских, А.Ю. Шеремеев, Р.З. Зулкарниев, Д.Ю. Колупаев, Н.В. Чебыкин, А.А. Кириллов. Формирование программ геолого-технических мероприятий с помощью цифровой информационной системы «Подбор ГТМ» // ПРОНЕФТЬ. Профессионально о нефти. 2017. № 2(4). С. 39-46.
2. Sola, J. & Sevilla, Joaquin. Importance of input data normalization for the application of neural networks to complex industrial problems. Nuclear Science, IEEE Transactions. 1997. vol. 44. no 18. P. 1464 - 1468.
3. S. Patro and K. K. Sahu. Normalization: A preprocessing stage // arXiv preprint. 2015. [Электронный ресурс]. URL: <https://arxiv.org/ftp/arxiv/papers/1503/1503.06462.pdf> (дата обращения: 16.05.2021)
4. Potdar, Kedar & Pardawala, Taher & Pai, Chinmay. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications. 2017. vol. 175. no. 4. P. 7-9.
5. Shuo Wang, Member, and Xin Yao. Multiclass Imbalance Problems: Analysis and Potential Solutions. IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics. 2012. vol. 42, no. 4. P. 1119-1130.
6. В.Н. Никулин, И.С. Канищев, И.В. Багаев. Методы балансировки и нормализации данных для улучшения качества классификации // Компьютерные инструменты в образовании. 2016. №3. С. 16-24
7. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research. 2002. vol. 16. Pages: 321–357.

ОБ АВТОРЕ

ХАКИМОВ Ринат Фанисович, аспирант 4-го курса кафедры ВМик УГАТУ.

METADATA

Title: Machine learning methods for classification of wells-candidates for geological and technical actions by rejection reasons.

Author: R. F. Khakimov

Affiliation: Ufa State Aviation Technical University (UGATU), Russia.

Email: wilmgc@gmail.com

Language: Russian.

Source: Molodezhnyj Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), no. 2 (25), pp. 97-100, 2021. ISSN 2225-9309 (Print).

Abstract: The article discusses the methodology for data preparation, training and verification of a neural network classifier for the problem of classifying candidate wells for geological and technical actions. The following main stages are described: data normalization, splitting the initial data into test and training samples, data balancing. The article touches upon such normalization methods as minimax normalization, unitary code method, SMOTE data balancing methods, Random Oversampling. A multilayer perceptron model is considered as a classifier. For the experiment, we used data on candidate wells for geological and technical actions to bring the well out of operation. The article also provides an assessment of the result obtained using such a verification method as a sliding control procedure.

Key words: oil production; geological and technical actions for wells; data normalization; data balancing; machine learning; classification methods; data verification; neural networks.

About author:

ХАКИМОВ, Rinat Fanisovich, postgraduate student 4-year, Ufa state aviation technical University.