

О МЕТОДЕ РАСЧЕТА ФИЛЬТРАЦИОННЫХ ХАРАКТЕРИСТИК СКВАЖИН УРАНОВЫХ МЕСТОРОЖДЕНИЙ НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация

Владение информацией фильтрационных характеристик вмещающих пород позволяет планировать объем добычи руды. Существующая методика расчета фильтрационных свойств скважин по добыче урана методом поземного скважинного выщелачивания базируется на системе правил, учитывающих лишь один геофизический параметр (кажущееся сопротивление-КС). Методика дает сравнительно низкую точность расчета коэффициента фильтрации. Кроме того, она не применима в случае брака при записи КС или искажения значений под действием кислоты, которая широко применяется при добыче урана. Вместе с тем, для расчета коэффициента фильтрации могут быть задействованы и другие геофизические параметры, использование которых может повысить точность расчета на 10-20% при применении методов машинного обучения.

В работе предполагается исследование применимости методов машинного обучения для прогнозирования фильтрационных свойств вмещающих пород скважин по добыче урана методом ПСВ (подземного скважинного выщелачивания) и численной оценки преимуществ подобного подхода. Будет разработан метод машинного обучения для прогнозирования фильтрационных свойств пород на основании неточно размеченных данных электрического каротажа.

Ключевые слова: Машинное обучение, геофизические исследования скважин, коэффициент фильтрации, неточные данные.

1. Введение

На Казахстан приходится 39% мировой добычи урана. Добыча ведется методом подземного скважинного выщелачивания (ПСВ). При этом методе уран добывается через сеть откачных и закачных скважин, по которой циркулирует выщелачивающий раствор. Применяемая в настоящее время методика расчета фильтрационных свойств скважин по добыче урана методом поземного скважинного выщелачивания базируется на системе правил, учитывающих лишь один геофизический параметр (кажущееся сопротивление-КС). Однако данная методика не применима в случае брака при записи КС или искажения значений под действием кислоты, которая широко применяется при добыче урана. Вместе с тем, для расчета

коэффициента фильтрации могут быть задействованы и другие геофизические параметры, использование которых может повысить точность расчета. Учет множества параметров при наличии фактических замеров может быть выполнен с помощью моделей машинного обучения (ММО). ММО делятся на две группы [1, 2, 3]:

- Обучение без учителя (*Unsupervised Learning*) (UL) [4].
- Обучение с учителем (*Supervised Learning*) (SL) [5].

UL решают задачу кластеризации, когда множество заранее не обозначенных объектов разбивается на группы путем автоматической процедуры, исходя из свойств объектов [6, 7].

SL используется тогда, когда в потенциально бесконечном множестве объектов выделяются конечные группы (классы) некоторым образом обозначенных объектов. Алгоритм использует эту первоначальную классификацию как образец и размечает не обозначенные объекты, относя их к подходящей группе (классу), исходя из свойств этих объектов. Хотя такие алгоритмы известны, выбор или разработка классификатора для конкретной практической задачи не тривиальна. Требуется проведение комплекса вычислительных экспериментов, значительная предварительная работа по подготовке данных и т.п. На протяжении многих лет внимание исследователей в области машинного обучения было сконцентрировано на нефтяной геофизике, где основной задачей является выделение коллектора. В урановой геофизике стандартный комплекс исследований содержит гораздо меньше методов, в тоже время, необходимо определять свойства множества достаточно мелких пропластков. Это делает задачу применения ММО на урановых месторождениях более сложной по сравнению с нефтяной геофизикой.

В процессе выполнения работы планируется путем проведения вычислительных экспериментов с использованием интервальных неточно размеченных данных определить алгоритм и его параметры, дающие наилучший результат предсказания коэффициента фильтрации. В настоящее время не существует систем интеллектуальной интерпретации данных геофизического исследования скважин на урановых месторождениях с целью расчета фильтрационных характеристик. Разработанный в результате проекта метод позволит определять значение коэффициента фильтрации с большей точностью. Предварительные эксперименты позволяют говорить о том, что применение ММО позволит увеличить точность расчета на 10-20%.

2. Описание вычислительных экспериментов

В качестве исходных данных использованы результаты гидрогеологических исследований скважин, выполненных на этапе разведки, как единственный источник фактических данных о фильтрационных свойствах горных пород.

Вычислительные эксперименты включают предобработку данных, формирование тренировочного и тестового наборов данных и расчет параметров классификационной или регрессионной модели. Предобработка данных имеет в данном случае особенности, поскольку в качестве источников данных служат интервальные отсчеты, полученные в ходе электрического каротажа скважин. Формирование тренировочного и тестового набора данных также имеет особенности поскольку их необходимо формировать с учетом скважинного разбиения данных и разбивки данных на фолды.

- Для поиска наилучших параметров регрессий и коэффициентов регуляризации произведён grid-search по параметрам трёх алгоритмов с записью результатов тестирования методов cross-validation.

- Сформирован набор моделей по принципу grid-search.

- Проведено тестирование, в результате которого получены показатели качества моделей, включая среднеквадратичную ошибку и determination coefficient, после чего модели были отсортированы по этим показателям на тестовой выборке и был выбран топ лучших моделей с оптимальными настройками мета-параметров.

- Для обучения модели был сформирован датасет, содержащий данные по месторождению «Буденновское».

- Разбиение на фолды для кросс-валидации производилось двумя способами:

Обычное стратифицированное разбиение на фолды, обеспечивающее попадание в тестовую и тренировочную выборку пропорционального общей выборке количество объектов с различными целевыми (значение коэффициента фильтрации) значениями

Разбиение на фолды по скважинам — при таком подходе в тестовую выборку отделялась каждая отдельная скважина, чтобы исключить возможность переобучения модели на конкретные физические и фильтрационные свойства отдельных скважин, поскольку простое стратифицированное разбиение на фолды не может обеспечить расчёта показательных метрик качества оценки в таких случаях.

- Перебор регрессоров и их параметров происходил по принципу grid search (решётчатый поиск), при котором выбираются значимые параметры классификатора и происходит их перебор в заданном интервале и с заданным шагом, после чего происходит выбор той модели, у которой показатели качества на тестовых выборках оптимальны.

В таблице 1 приведены регрессоры, показавшие наилучшие результаты на разных наборах входных данных.

	Перепровер	Correlation train	Correlation test	MSE train	MSE test
КС	MLP Regressor - one hidden layer with 201 neurons, 600 max iterations	0,494	0,613	32,389	35,872
КС+Код породы	MLP Regressor - one hidden layer with 201 neurons, 850 max iterations	0,502	0,488	30,968	38,384
КС+ПС	MLP Regressor - one hidden layer with 201 neurons, 600 max iterations	0,596	0,535	25,785	30,472
КС+ПС+ Код породы	MLP Regressor - one hidden layer with 251 neurons, 100 max iterations	0,698	0,570	21,244	30,363
КС	Расчет по принятой методике		0,261		48,32

Таблица 1 – результаты регрессоров на разных наборах входных данных.

Таким образом, примененный подход позволил оценить фильтрационные свойства пород лучше, чем существующая методика. Выявлено, что корреляция с фактическими данными выше и меньше среднеквадратичная ошибка. Расширение датасета за счет каротажных данных по большему числу скважин позволит повысить точность модели.

ЛИТЕРАТУРА

1. Ayodele T. O. Types of machine learning algorithms //New advances in machine learning. – 2010. – Т. 3. – С. 19-48.
2. Ibrahim H. A. H. et al. Taxonomy of machine learning algorithms to classify real time interactive applications //International Journal of Computer Networks and Wireless Communications. – 2012. – Т. 2. – №. 1. – С. 69-73.
3. Muhamedyev R. Machine learning methods: An overview //Computer Modelling & New Technologies. – 2015. – Т. 19. – №. 6. – С. 14-29.
4. Hastie T., Tibshirani R., Friedman J. Unsupervised learning //The elements of statistical learning. – Springer, New York, NY, 2009. – С. 485-585.
5. Kotsiantis S. B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques //Emerging artificial intelligence applications in computer engineering. – 2007. – Т. 160. – №. 1. – С. 3-24.
6. Jain A. K., Murty M. N., Flynn P. J. Data clustering: a review //ACM computing surveys (CSUR). – 1999. – Т. 31. – №. 3. – С. 264-323.
7. Ashour B. W., Ying W., Colin F. Review of Clustering Algorithms. Non-Standard Parameter Adaptation for Exploratory Data Analysis //Studies in Computational Intelligence. – 2009. – Т. 249. – С. 7-28.

Данная работа выполнена при поддержке грантового финансирования научных проектов КН МОН РК № AP09562260 «Разработка метода расчета фильтрационных характеристик скважин урановых месторождений на основе алгоритмов машинного обучения, работающих с неточно размеченными интервальными данными».