

**ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ
ДЛЯ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ ГИС В КОНТЕКСТЕ ЗАДАЧИ
ВЫДЕЛЕНИЯ ТЕРРИГЕННЫХ КОЛЛЕКТОРОВ**

Аннотация: В статье рассматривается применение алгоритмов машинного обучения для оперативной интерпретации данных геофизических исследований скважин в контексте задачи выделения терригенных коллекторов. Представлен анализ полученных результатов.

Ключевые слова: машинное обучение, геофизические исследования скважин, коллектор, интерпретация, machine learning.

Введение

Одной из задач геофизики является разведка месторождений полезных ископаемых. Для изучения свойств и структуры геологического разреза производится бурение разведочных скважин с последующим отбором керна (образцов горных пород). Зачастую, отбор керна из скважин затруднен или невыгоден с экономической точки зрения. В этом случае применяются другие методы получения геологической информации – методы геофизических исследований скважин (ГИС) [1]. Суть данных методов состоит в измерении показаний различных датчиков, характеризующих физические, химические и другие свойства горных пород вдоль ствола скважины при помощи специальных зондов и наземной аппаратуры. В настоящее время бурение любой скважины обязательно сопровождается комплексом ГИС[2]. Полученные показания датчиков передаются инженерам-геофизикам для интерпретации.

Геологическая интерпретация данных геофизических исследований скважин – сложная задача, имеющая важнейшее практическое значение для разведки месторождений полезных ископаемых. Зависимость геологических

параметров от полученных показаний с датчиков в подавляющем большинстве случаев носит сложный и нелинейный характер. При интерпретации данных значимость имеют не абсолютные значения полученных сигналов с зонда, а их соотношения[1]. Более того, инженеру-геофизику нужно учитывать погрешности измерений сигналов (шум, растяжение кабеля спускаемого зонда) и специфику территории. Все это требует ощутимых временных затрат инженера-геофизика. Принимая во внимание все сказанное выше, можно сделать вывод об эффективности применения алгоритмов машинного обучения[3] для задач интерпретации данных ГИС [4,5,6], т.к. данные методы обладают свойством адаптивности, обобщения, извлечения знаний и моделирования сложных нелинейных зависимостей в массивах данных.

Одним из первых этапов разведки месторождения является оперативная интерпретация данных ГИС. Под оперативной интерпретацией данных ГИС[1] понимают подготовку и выдачу (как можно быстрее) геологической службе буровых предприятий заключения о наличии в разрезах скважин нефтегазонасыщенных пластов. Она производится на всех этапах разведки и эксплуатации нефтегазовых месторождений, включая бурение первых скважин, когда отсутствуют достоверные сведения об изучаемых геологических разрезах и не установлены конкретные зависимости между геофизическими величинами и коллекторскими свойствами. Поэтому определяемые характеристики носят качественный или полуквантитативный характер [1].

Постановка задачи

Одной из задач оперативной интерпретации данных ГИС является задача выделения пластов-коллекторов. Выделяют несколько типов коллекторов: терригенные, карбонатные и др. Для каждого типа характерны свои значения данных ГИС. Как было отмечено выше, необходимо учитывать эту специфику, поэтому рассмотрим задачу выделения терригенных коллекторов с помощью алгоритмов машинного обучения.

Наиболее надежно терригенные коллекторы выделяются [1] по совокупности диаграммы ПС (Метод потенциалов самопроизвольной поляризации), кривой ГК (Гамма-каротаж).

Постановка исходной прикладной задачи звучит следующим образом:

С помощью имеющихся данных комплекса геофизических исследований необходимо выделить терригенные коллекторы вдоль глубины ствола скважины.

Из постановки следует, что задача является *классификационной*.



Рис. 1. Схема процесса интерпретации данных ГИС.

На рис.1 представлена схема процесса интерпретации ГИС. В данном случае интерпретатор (инженер-геофизик) получает данные с датчика №1, с датчика №2 и датчика №3. Далее он проводит анализ значений и соотношений показаний датчиков вдоль всего ствола скважины, используя дополнительную информацию (опыт работы, специфика территории), и определяет интервалы наличия коллекторов.

Опишем математическую модель процесса. Данные датчиков ГИС – это функция от одной переменной (глубины ствола скважины).

$$y = f_{name}(d) \quad (1)$$

Интерпретатор – функция N-переменных, где N – число датчиков, участвующих в интерпретации, name – название кривой ГИС, d- глубина вдоль ствола.

$$z = F(f_{name1}(d), f_{name2}(d), \dots, f_{nameN}(d)) \quad (2)$$
$$z \in \{0,1\}$$

Математическая постановка задачи звучит следующим образом:

Необходимо подобрать функцию F , определенную на множестве $\{0,1\}$, которая возвращает 1, если при данном наборе значений в скважине присутствует терригенный коллектор и 0 в противном случае. Определить интервалы коллекторов вдоль ствола скважины, используя функцию F .

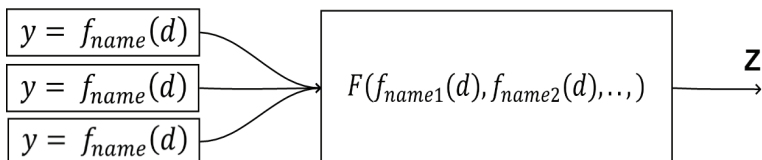


Рис. 2. Математическая модель процесса интерпретации ГИС.

В роли функции интерпретации будет выступать алгоритм (метод) машинного обучения. В работе были использованы следующие методы [3, 7, 8, 9, 10]:

1. Ансамбли моделей: adaboost, gradient boosting, random forest, bagging, extra trees, extreme gradient boosting
2. Алгоритм k-nearest neighbors
3. Деревья принятия решений: decision tree, extra tree
4. Метод опорных векторов (ядро – сигмоид)
5. Многослойная нейронная сеть. Число нейронов на внутренних слоях составило (5,3,1)

Эксперимент

В качестве входных данных использовались показания датчиков ПС, ИК (Индукционный каротаж) и ГК. Также в ходе экспериментов был добавлен специальный аргумент. Синтезированный параметр определялся как разность показаний ПС и ИК на определенной глубине.

Всего для эксперимента было использовано 75557 показаний со 108 скважин. В обучающую выборку вошло 8024 показания (~10%) с 11 скважин.

Все данные прошли нормировку [3,7,8]. Для оценки качества обобщающей способности алгоритма использовалась процедура кросс-валидации [3,7,8].

Результаты обучения представлены в таблице:

Таблица 1. Результаты обучения алгоритмов

Алгоритм	Точность		Ст. отклонение
	Тренировочная выборка	Тестовая выборка	
ExtraTrees	1	0.921387	0.003291227
Bagging	0.993851	0.919477	0.004341433
RandomForest	0.994038	0.917733	0.003540633
k-nearest neighbors	0.941795	0.912666	0.0034648
GradientBoosting	0.935251	0.911337	0.003485667
Extreme gradient boosting	0.926215	0.91005	0.003519867
DecisionTree	1	0.904277	0.002686527
ExtraTree	1	0.901744	0.006050033
AdaBoost	0.908787	0.897841	0.004745867
Метод опорных векторов	0.890154	0.891154	0.006202733
Многослойная нейронная сеть	0.888076	0.890241	0.0060496

Как видно из таблицы, все алгоритмы показали приемлемый результат. При проведении тестирования на реальных данных произошло снижение точности алгоритмов.

Таблица 2. Результаты алгоритмов на реальных данных

Алгоритм	Точность	Стд. Отклонение
Многослойная нейронная сеть	0.836004	0.121893
Extreme gradient boosting	0.808702	0.125434
ExtraTrees	0.802129	0.133246
k-nearest neighbors	0.80188	0.117094
Метод опорных векторов	0.789811	0.172834
GradientBoosting	0.773262	0.134646
AdaBoost	0.769521	0.146228
RandomForest	0.727287	0.173611
DecisionTree	0.723322	0.154546
Bagging	0.709825	0.175045
ExtraTree	0.699206	0.186389

Из полученных результатов видно, что лучший результат показала многослойная нейронная сеть.

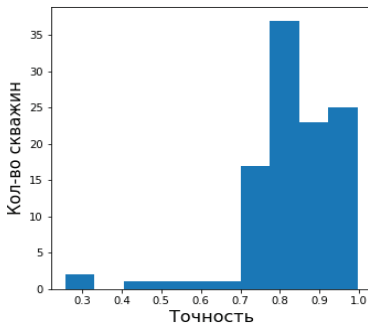


Рис. 3. Гистограмма распределения скважин по точности интерпретации с применением многослойной нейронной сети

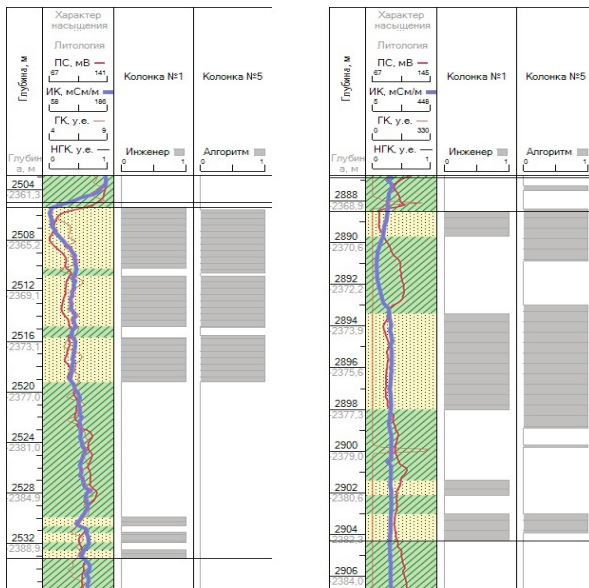


Рис. 4. Сравнение результатов интерпретации инженером и алгоритмом машинного обучения

Рассмотрим подробнее результаты работы этого алгоритма. На рис.3 показана гистограмма распределения скважин по точности интерпретации с

применением многослойной нейронной сети. Можно сделать вывод, что точность интерпретации для подавляющего количества скважин превышает 70%.

На Рис.4 представлена фрагмент планшета интерпретации данных ГИС для двух скважин. На планшете видно, что результаты интерпретации инженера-геофизика отличаются от вывода алгоритма машинного обучения. Также стоит отметить, что интервалы коллекторов, на которых кривые ГК, ПС и ИК начинают «изгибаться», определены с высоким уровнем точности. Одной из причин расхождения результатов можно назвать априорные знания геофизика, которые получены не из данных ГИС.

Заключение

Применение алгоритмов машинного обучения для решения задачи выделения коллекторов имеет смысл по ряду причин:

1. Проведение экспресс-интерпретации с помощью алгоритмов машинного обучения сразу выделит интервалы, на которые однозначно нужно обратить внимание геофизику.
2. Результаты компьютерной интерпретации могут быть использованы для последующего принятия решения экспертом.
3. Алгоритмы машинного обучения имеют возможность постоянного обучения, что повышает точность и расширяет границы применимости алгоритма (можно «обучить» алгоритм выделять не только терригенные коллекторы).
4. Описанные алгоритмы можно применять для других геофизических задач. Например, определение характера насыщения коллектора[1].

СПИСОК ЛИТЕРАТУРЫ

1. Косков В.Н. Геофизические исследования скважин: Учеб. пособие/Перм.гос.техн. ун-т. – Пермь, 2004. – 122 с.

2. Техническая инструкция по проведению геофизических исследований и работ приборами на кабеле в нефтяных и газовых скважинах: РД 153-39.0-072-01: утв. М-вом энергетики Рос. Федерации 01.07.01.
3. MachineLearning.ru - профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. //URL: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение (дата обращения: 18.02.2018)
4. Агаев Х. Б. Применение кластерного анализа для расчленения геологического разреза по данным каротажа скважины //Каротажник. – 2013. – №. 5. – С. 3-11.
5. Гафуров Д.О. Геологическая интерпретация с применением обучаемых нейронных сетей в" НейроИнформГео" данных ГИС Талаканского нефтегазоконденсатного месторождения //Известия Томского политехнического университета. Инжиниринг георесурсов. – 2006. – Т. 309. – №. 3.
6. Паклин.Н. Анализ геофизических данных //URL: <https://basegroup.ru/community/articles/geophysics> (дата обращения: 1.04.2018).
7. Marsland S. Machine learning: an algorithmic perspective. – CRC press, 2015.
8. Shalev-Shwartz S., Ben-David S. Understanding machine learning: From theory to algorithms. – Cambridge university press, 2014.
9. Chen T., Guestrin C. Xgboost: A scalable tree boosting system //Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. – ACM, 2016. – С. 785-794.
10. Natekin A., Knoll A. Gradient boosting machines, a tutorial //Frontiers in neurorobotics. – 2013. – Т. 7. – С. 21.