

Обзор открытых наборов скважинных данных

Д.О. Макиенко¹, И.В. Сафонов¹

¹ Национальный Исследовательский Ядерный Университет МИФИ, Каширское шоссе, 31, Москва, 115409, Российская Федерация

Аннотация

В последнее время возрастает количество исследований, посвященных применению методов машинного обучения в геофизике. Примерами исследовательских тематик являются предсказания свойств горных пород и их типизация по количественным признакам. Для построения предсказывающих моделей необходимы размеченные наборы данных. Целью работы является обзор открытых размеченных наборов скважинных данных и исследований, для которых они использовались. Проводится анализ наборов данных, использовавшихся в конкурсах по применению машинного обучения для решения геофизических задач, а также ресурсов, предоставляющих скважинные данные в открытом доступе. В работе рассматриваются наборы данных, содержащие геофизические измерения, изображения горных пород, результаты лабораторных исследований, а также результаты разметки литотипов.

Ключевые слова

Геофизические исследования скважин, изображения горных пород, открытые наборы данных, машинное обучение.

Overview of Open Well Datasets

D.O. Makienko¹, I.V. Safonov¹

¹ National Research Nuclear University MEPhI, Kashirskoye highway, 31, Moscow, 115409, Russian Federation

Abstract

Recently, the number of studies devoted to the use of machine learning methods in geophysics has been increasing. Examples of such studies include the prediction of rock properties and separation of rock types according to quantitative characteristics. Annotated datasets are required to build machine learning based models. The purpose of this paper is to review open labeled well datasets and some research used these datasets. Datasets from competitions in machine learning for geophysical problems are analyzed, as well as other publicly available sources of open well data. The paper considers datasets containing well logging, rock images, laboratory research results, as well as labeled zonation by lithotypes.

Keywords

Well logging, rock images, open datasets, machine learning.

1. Введение

В последнее десятилетие прогресс во многих областях обусловлен широким применением методов, основанных на машинном обучении (МО). Успешное решение задачи с помощью методов МО, как правило, связано с наличием большого представительного набора размеченных данных. Однако часто исследователи сталкиваются с ситуацией, когда разнообразных размеченных данных недостаточно для создания моделей, обладающих высокой обобщающей способностью. Обогащение (аугментация) и генерация искусственных данных в ряде случаев

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: makienko-dasha@mail.ru (Д.О. Макиенко); ilia.safonov@gmail.com (И.В. Сафонов)

ORCID: 0000-0001-7341-6128 (Д.О. Макиенко); 0000-0002-8270-943X (И.В. Сафонов)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

позволяет существенно повысить качество решения задач [1]. Использование открытых наборов данных даёт возможность проверить применимость и оценить обобщающую способность имеющейся модели МО, иногда позволяет улучшить модель.

При разработке месторождений в нефтегазовой индустрии выполняют геофизические исследования скважин (ГИС) [2,3]. При проведении ГИС в скважину опускают датчики, которые с некоторым шагом по глубине или времени измеряют свойства пород (электрические, радиоактивные, акустические и другие). Результаты измерений представляются в виде массивов данных с привязкой к глубине скважины. При бурении из скважины извлекается горная порода (керна), которая фотографируется. Из керна вырезаются образцы для исследования в лаборатории различных характеристик породы, например пористости и проницаемости.

На Рисунке 1 показан типичный пример визуализации скважинных данных. В левом столбце расположена шкала глубин. Во втором столбце находятся фотографии керна. В других столбцах показаны, так называемые, кривые ГИС для различных типов датчиков. При работе со скважинными данными часто можно столкнуться с отсутствием (пропуском) информации на некоторых глубинах. На рисунке 1 для ряда интервалов глубин отсутствуют фотографии керна, но нет пропусков в кривых ГИС. Однако на практике такие пропуски встречаются нередко.

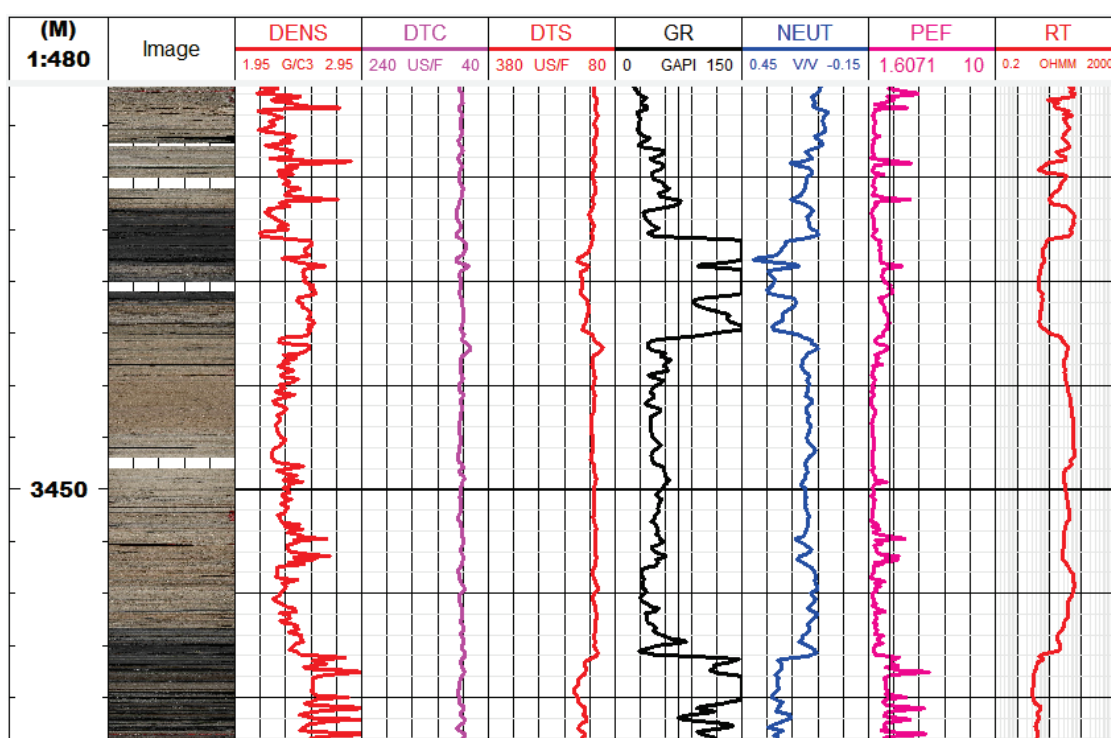


Рисунок 1 – Пример отображения скважинных данных

За прошедшие десятилетия накоплен огромный объём скважинных данных. Однако эти данные принадлежат компаниям - недропользователям и, как правило, относятся к закрытой информации. Кроме того, в ряде государств, в частности РФ [4], существуют ограничения на экспорт геологической информации, что препятствует раскрытию данных ГИС. По этим причинам даже те немногие существующие открытые наборы скважинных данных способны существенно помочь в разработке методов для их обработки и анализа.

В представленном обзоре проводится сравнительный анализ скважинных данных из пяти наборов, использовавшихся в конкурсах (соревнованиях) по применению МО для анализа геофизических данных; обсуждаются задачи, которые уже решались или могут быть решены с помощью этих наборов; демонстрируется способ визуализации данных ГИС; рассматриваются типичные проблемы качества скважинных данных; приводится описание нескольких интернет ресурсов, предоставляющих скважинные данные в открытом доступе.

2. Обзор наборов данных

В таблице 1 приведены сведения о наборах, используемых в пяти конкурсах (competitions) по анализу данных ГИС. Конкурс FORCE 2020 Machine Learning Competition в таблице и далее по тексту обозначен Force-2020. Этот конкурс проводился для двух задач: классификации классов пород – литофаций, и определения разломов по сейсмическим данным. В представленном обзоре рассматривается только первая из задач. В наборе представлен один обучающий и два тестовых набора данных. Конкурс 2016-ml-contest также посвящен предсказанию литофаций и содержит обучающий и тестовый наборы данных. Конкурсы, которые проводит SPWLA Petrophysical Data-Driven Analytics Special Interest Group, обозначены 2020-SPWLA, 2021-SPWLA и 2023-SPWLA в соответствии с годом проведения. На конкурсе 2020-SPWLA выполнялось предсказание отсутствующих кривых ГИС. В конкурсе 2021-SPWLA стояла задача предсказания характеристик породы. Конкурс 2023-SPWLA посвящен увязке кривых по глубине. Для 2020-SPWLA и 2021-SPWLA имеются обучающий и тестовый наборы данных. Для 2023-SPWLA доступен обучающий набор данных и входные данные для тестирования предсказывающих моделей. Можно ожидать, что по окончании конкурса появятся результаты разметки для тестового набора.

Таблица 1 – Наборы данных конкурсов по анализу кривых ГИС

Набор данных	Force-2020 ¹			2016-ml-contest ²		2020-SPWLA ³		2021-SPWLA ³		2023-SPWLA ³	
	Обучающий набор	Тестовый набор 1	Тестовый набор 2	Обучающий набор	Тестовый набор	Обучающий набор	Тестовый набор	Обучающий набор	Тестовый набор	Обучающий набор	Тестовый набор
Количество скважин	98	10	10	10	2	3	4	9	4	9	3
Количество строк в таблицах	1170511	136786	122397	3232, 4149	809	30143	11088	318967	11275	69304	19038
Число типов данных	29	28	29	11		9		17		5	11
Размер в Мб	267	31	29	0.2	0.05	1.9	0.7	37	1	2.4	0.9
Тип задачи	Классификация 12 классов пород			Классификация 9 классов пород		Регрессия для предсказания кривых ГИС		Регрессия для предсказания характеристик породы		Увязка глубин	
Лицензии	CC-BY 4.0			CC-BY 1.0		CC BY-NC-SA		CC BY-NC-SA		CC BY-NC-SA	

¹ <https://github.com/bolgebrygg/Force-2020-Machine-Learning-competition>

² <https://github.com/seg/2016-ml-contest>

³ <https://github.com/pddasig>

Наборы данных, представленные в таблице 1, относятся к измерениям в скважинах, поэтому одной из характеристик наборов является количество скважин. Для всех конкурсов в тестовом и обучающем наборах используются данные разных скважин, но для одинакового месторождения. Каждый набор данных представляет собой таблицу и хранится в файле csv формата. Файлы с данными охарактеризованы количеством строк в таблицах, количеством столбцов с различными типами измерений и размером файла в Мб. Под количеством строк понимается число измерений для разных глубин скважин. Данные разных скважин либо разделены по разным файлам, либо располагаются в общей таблице. Для идентификации скважины в общей таблице может указываться её номер или название.

Для конкурса 2016-ml-contest в обучающем и тестовом наборах не было пропущенных строк, однако присутствовал файл, который включал в себя обучающий набор и дополнительно 917 строк, для которых одна кривая ГИС отсутствовала. Поэтому обучающий набор с конкурса 2016-ml-contest охарактеризован двумя значениями количества строк. В тестовом наборе 1 конкурса Force-2020 присутствует на один тип данных меньше, чем в обучающем наборе и тестовом наборе 2. В тестовом наборе конкурса 2023-SPWLA присутствуют дополнительно 6 полей для результатов предсказаний.

Рисунок 2 иллюстрирует данные четырех скважин из тестового набора 2021-SPWLA. Кривые выводились с использованием пакета matplotlib (<https://matplotlib.org>) для языка программирования Python. Для отображения пропущенных значений в скважинных данных удобно воспользоваться пакетом missingno (<https://github.com/ResidentMario/missingno>). Числа 1 и 11275 в левой части графика соответствуют номерам первой и последней строк таблицы. В самом левом столбце указан номер скважины. Серый график в правой части рисунка показывает количество значений в каждой строке таблицы. На этом графике показаны минимальное и максимальное количество данных без пропусков для строк таблицы. Белые участки в столбцах с кривыми ГИС соответствуют пропущенным значениям.

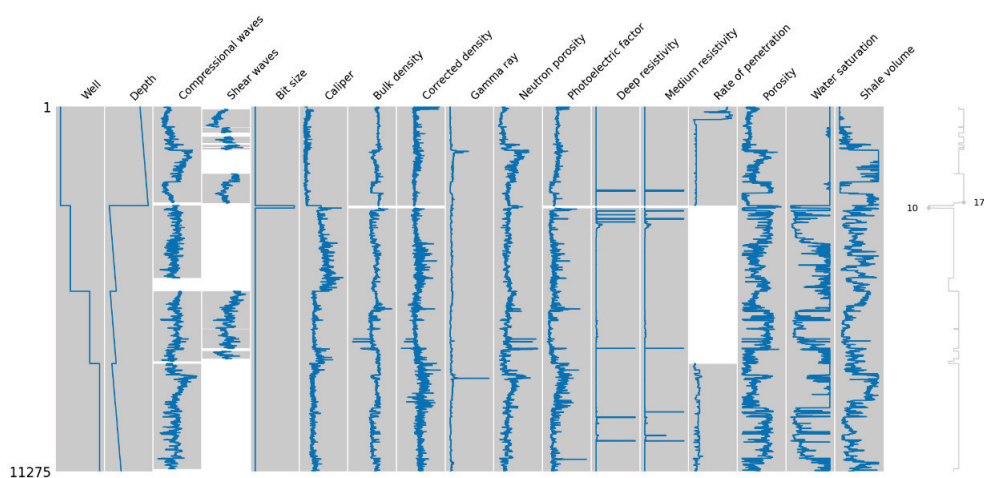


Рисунок 2 – Визуализация тестового набора 2021-SPWLA

В конкурсе Force-2020 решалась задача классификации двенадцати классов литофаций. Во многих наборах скважинных данных наблюдается дисбаланс классов, связанный с различной частотой встречаемости пород. Распределение классов в обучающем наборе Force-2020 показано на рисунке 3. Схожее распределение классов наблюдается и в тестовых наборах. В конкурсе 2016-ml-contest выполнялась классификация девяти классов пород. По набору 2020-SPWLA предсказывались акустические кривые ГИС. В соревновании 2021-SPWLA решалась задача регрессии для предсказания объема сланца, пористости и водонасыщения. Конкурс 2023-SPWLA посвящен увязке глубин между кривыми ГИС.

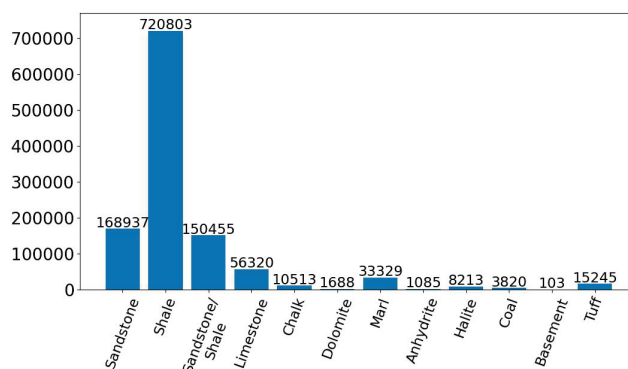


Рисунок 3 – Распределение классов литофаций в обучающем наборе Force-2020

В таблице 2 перечислены типы данных, присутствующие в конкурсных наборах. Кривые ГИС распределены по нескольким группам: гамма-излучение, электрическое сопротивление, нейтронная и объемная плотность, акустические измерения (интервальное время), фотоэлектрический фактор, собственный (самопроизвольный) потенциал. Также имеется группа с результатами интерпретации. Помимо перечисленных типов данных в наборах присутствует метаинформация, содержащая данные о бурении. В ячейках таблицы указано количество непустых строк в столбцах данных этого типа суммарно для обучающего и тестовых наборов. Для некоторых значений в скобках сделаны уточняющие тип данных комментарии. Объяснение физического смысла данных ГИС и рекомендации по их использованию для геофизической интерпретации можно найти в работах [2,3].

Таблица 2 – Типы кривых ГИС в наборах данных конкурсов

Наборы данных Типы данных	Force-2020	2016-ml-contest	2020-SPWLA	2021-SPWLA	2023-SPWLA
Гамма-излучение	1429694 70492 (spectra)	4958	40977	327123	88342
Электрическое сопротивление	695430 (shallow) 1380297 (medium) 1418609 (deep) 202657 (micro) 347997 (flushed zone)	4958	40846 (medium, deep)	299951 (medium) 299717 (deep)	88342 (deep)
Нейтронная пористость и объемная плотность	966012 (neutron porosity) 1241937 (bulk density) 1211680 (density correction)	4958 (neutron-density porosity difference) 4958 (average neutron-density porosity)	40496 (neutron porosity) 40550 (bulk density)	106813 (bulk density) 100309 (density correction) 106710 (neutron porosity)	88342 (neutron porosity) 88342 (bulk density)
Акустические измерения	1343902 (compressional) 290704 (shear)		37177 (compressional) 36366 (shear)	80508 (compressional) 47759 (shear)	
Фотоэлектрический фактор	885634	4041	40552	99729	
Собственный (самопроизвольный) потенциал	977197				
Интерпретация	1429694 (lithofacies) 1292709 (confidence)	4958 (lithofacies)		58589 (porosity) 58589 (water saturation) 56375 (shale volume)	

Существует ряд интернет-ресурсов, на которых можно найти скважинные данные в открытом доступе. В отличие от наборов конкурсов по анализу кривых ГИС, требуются определённые усилия по предобработке этих данных, в частности для сведения в таблицы с привязкой по глубине. В таблице 3 даны общие сведения о скважинных данных, доступных в открытом доступе на пяти следующих ресурсах:

- WAPIMS (Western Australian Petroleum and Geothermal Information Management System) предоставляет информацию о бурении скважин в Австралии.
- NOPIMS (National Offshore Petroleum Information Management System) содержит данные о скважинах на шельфе Австралии.
- BGS (British Geological Survey) предоставляет в открытый доступ геологические данные Великобритании.
- Volve является набором данных о месторождении в Северном море, разрабатываемом компанией Equinor.
- NLOG содержит информацию об энергетических и минеральных ресурсах Нидерландов.

Таблица 3 – Ресурсы с открытыми скважинными данными

Название	Примеры скважин	ГИС	Интерпретация ГИС	Изображения	Литологическая зонация	Стандартный анализ керна	Лицензия
WAPIMS ¹	GSWA Harvey1, DMP Harvey 2, DMP Harvey 3, DMP Harvey 4	да	-	да	да	да	Non-commercial use
NOPIMS ²	Satyr 5, Dorado 2, Barossa 4	да	-	да	-	да	CC-BY 4.0
	Dorado 3, BDC 4 06 P	да	-	да	-	-	
BGS ³	204/19-3A, 204/24a- 6, 10/01-A25, 106/20- 1	-	-	да	-	-	CC BY-NC 3.0, Open Government License
Volve ⁴	15/9-F-1, 15/9-F-1A, 15/9-F-1B	да	да	-	-	-	Equinor Open Data License
NLOG ⁵	F13-01	да	-	-	да	-	Users are permitted to copy, to download and to disclose in any way, to distribute or to simplify the information provided on this website without the prior written permission of NLOG.NL or the lawful consent of the entitled party. Users are also permitted to copy, duplicate, process or edit the information and/or layout, provided NLOG.NL is quoted as the source.
	K05-02	да	-	-	-	да	
	ANNERVEEN-06	да	-	да	-	да	

¹ https://wapims.dmp.wa.gov.au/WAPIMS_2 ² https://www.ga.gov.au/nopims_2

³ <https://webapps.bgs.ac.uk/data/offshoreWells>, ⁴ <https://www.equinor.com/energy/volve-data-sharing>,

⁵ <https://www.nlog.nl/en>

В таблице 3 представлена информация о наличии данных ГИС, геофизической интерпретации ГИС, изображений, литологической зонации и результатов стандартного анализа керна. Прочерки в ячейках таблицы соответствуют случаям, когда такие данные отсутствуют или не удалось получить к ним доступ. Набор данных Volve содержит данные ГИС для двадцати четырех скважин. Данные из набора Volve используются в конкурсах SPWLA. На сайтах WAPIMS, NOPIMS, BGS и NLOG есть информация о большом количестве скважин и виды данных для них могут отличаться. Поэтому в таблице указаны названия рассматриваемых скважин. На сайте BGS рассматриваются данные из раздела о скважинах на шельфе (Offshore Hydrocarbon Wells).

Часто данные ГИС можно найти в виде отчетов с изображениями кривых, например, в формате pdf. Так как извлечение данных из изображений кривых является отдельной нетривиальной задачей, такие форматы данных в обзоре не рассматривались. В Таблице 3 учтена информация доступная в числовом виде, которая хранится в специфичных для нефтегазовой и геологической отраслей форматах las или dliis. Существуют пакеты для языка программирования Python для работы с такими форматами, например, lasio (<https://lasio.readthedocs.io/en/latest/>), dliisio (<https://dliisio.readthedocs.io/en/latest/index.html>).

Под литологической зонацией подразумевается описание типов пород, сделанное для интервалов глубин. Литологическая зонация может храниться в виде таблиц, текста или схематического изображения, и описана с разной степенью детальности. Под стандартным анализом керна понимаются результаты лабораторных исследований над образцами,

выбуренными из керна на некоторых глубинах. При стандартном анализе керна определяются характеристики пород, например, пористость и проницаемость.

Изображения керна получают путем фотографирования пород в дневном и ультрафиолетовом свете, а также с использованием компьютерной томографии. Примеры изображений пород с различных ресурсов показаны на рисунке 4. Выбуренный керна обычно раскладывают по ящикам с указанием соответствующего ему интервала глубин. На одной фотографии, как правило, есть изображения нескольких ящиков. На фотографиях керна могут присутствовать дыры округлой формы, которые появляются в результате выпиливания образцов для дальнейшего исследования в лаборатории. В места отсутствия породы добавляются деревянные, картонные или другие вставки.

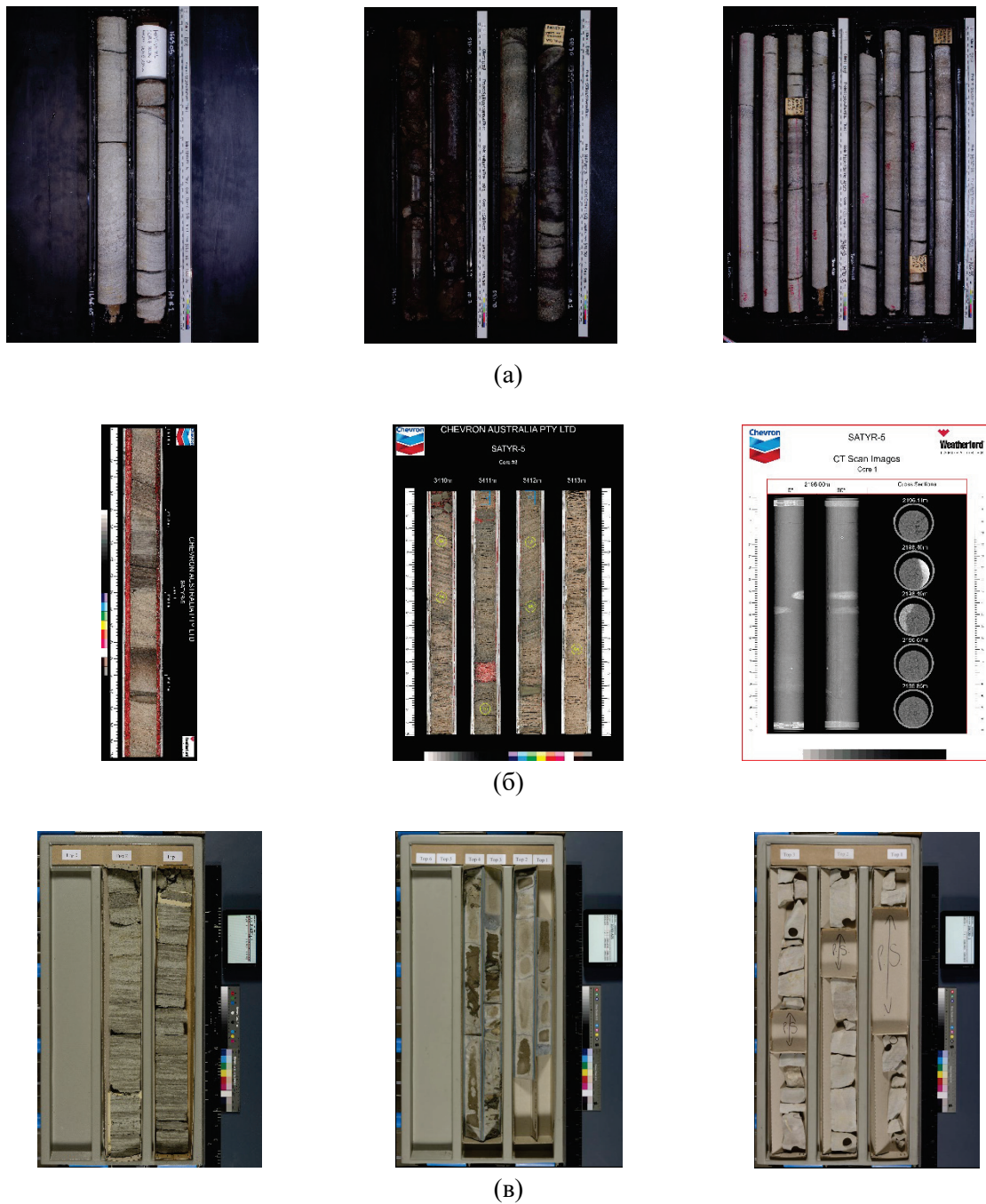


Рисунок 4 – Примеры изображений горных пород: (а) WAPIMS; (б) NOPIMS; (в) BGS

3. Обзор задач, решаемых со скважинными данными

В таблице 4 перечислены задачи, которые решались на рассмотренных наборах скважинных данных. Во втором столбце перечислены конкурсные наборы данных и интернет-ресурсы. При использовании данных интернет-ресурсов в публикациях обычно указывают примеры используемых скважин. В третьем столбце перечислены проводимые на данных исследования в виде публикаций или решений, предложенных участниками конкурсов. Конкурсные решения доступны в виде программ или кратких обзоров. Предполагается, что полный набор данных для конкурса 2023-SPWLA будет доступен после его завершения.

Таблица 4 – Задачи, решаемые со скважинными данными

Задача	Набор данных	Исследования
Классификация горных пород	Force-2020	Конкурсные решения, [5,6]
	2016-ml-contest	Конкурсные решения, [7,8,9]
	WAPIMS	[10]
	BGS	[11]
	NLOG	[12,13]
	NOPIMS	[14]
Предсказание характеристик породы	2021-SPWLA / Volve	Конкурсные решения
	WAPIMS	[15]
	Volve	[16]
Увязка глубин	2023-SPWLA	Конкурсные решения
Заполнение пропусков в кривых ГИС	Force-2020	[17]
	Volve	[17,16]
	NLOG	[18]
Предсказание кривых ГИС	2020-SPWLA / Volve	Конкурсные решения, [19,20]
	NOPIMS	[21]
Выделение на изображении неразрушенных фрагментов породы	BGS	[22]
Нахождение корреляции между геологическими разрезами разных скважин	Force-2020	[23]

Классификация горных пород проводилась как по кривым ГИС, так и по изображениям. Для наборов Force-2020, 2016-ml-contest, а также скважин, полученных из ресурсов WAPIMS и NLOG в работах [10,12], присутствовала литологическая зонация. Для скважин ресурсов BGS, NLOG и NOPIMS, которые использовались в работах [11,13,14], авторы получали литологическую зонацию в ручном или полуавтоматическом режиме.

Задача регрессии для предсказания характеристик породы решалась на трех наборах данных. Конкурсные данные 2021-SPWLA использовались для предсказания объема сланца, пористости и водонасыщения. Данные ГИС ресурса WAPIMS были использованы в работе [15] для предсказания пористости и проницаемости. В работе [16] на наборе Volve оценивалось влияние выбора кривых ГИС на качество предсказания пористости.

Предсказания кривых ГИС осуществлялось на наборах 2020-SPWLA и NOPIMS. Для обоих наборов предсказывались акустические кривые. Заполнение пропусков в кривых ГИС выполнялось на наборах Force-2020, Volve и NLOG. Эта задача схожа с предсказанием характеристик породы и кривых ГИС, где целевая величины непрерывна. Однако в общем случае для заполнения пропущенных значений может использоваться не только регрессионная модель, но и другие техники, например заполнение средними значениями.

К предварительной обработке данных в перечисленных работах подходят по-разному. Для данных ГИС предварительная обработка может включать:

- Выбор наиболее значимых кривых с учетом их физического понимания или с использованием специальных алгоритмов.
- Пополнение набора кривых новыми кривыми, полученными путем преобразования исходных.
- Заполнение или удаление пропущенных значений.
- Обнаружение и удаление выбросов.
- Увязку глубин между разными кривыми ГИС и изображениями пород.

Для изображений горных пород предварительная обработка включает вырезание породы из ящиков, выравнивание освещенности и коррекция цветов, удаление различного рода дефектов. В работе [22] на изображениях из набора BGS выделяют неразрушенные фрагменты породы.

Помимо вышеперечисленного на наборе Force-2020 решалась задача нахождения корреляции между геологическими разрезами для разных скважин [23].

Как правило проводятся исследования в пределах одного набора данных, содержащего информацию о скважинах из одного или нескольких близлежащих месторождений. Качество моделей оценивается на скважинах или участках скважин, которые не были задействованы в обучении. Хотя возникают сложности при построении моделей даже в рамках одного набора данных, было бы интересно попробовать построить модели, применимые к различным наборам данных.

4. Заключение

Прогресс в области МО во многом связан с количеством и качеством доступных данных. Предоставление в открытый доступ размеченных скважинных данных способствует дальнейшему развитию моделей МО, применяемых в области геофизики. Представлен обзор существующих открытых наборов скважинных данных. Рассмотрены пять наборов данных, использовавшихся в конкурсах по применению МО для решения геофизических задач, и пять интернет-ресурсов, предоставляющих различные виды скважинных данных.

На рассмотренных наборах скважинных данных решались следующие задачи:

- Классификация горных пород по данным ГИС и по изображениям.
- Предсказание характеристик породы по данным ГИС.
- Предсказание акустических кривых ГИС.
- Корректировка кривых ГИС, которая включала заполнение пропусков и увязку глубин.
- Нахождение корреляции между геологическими разрезами разных скважин по данным ГИС.
- Выделение на изображениях неразрушенных фрагментов породы.

Можно предположить, что в будущем для данных наборов будут решаться следующие задачи:

- Построение общих моделей для многих месторождений.
- Для кривых ГИС: создание специфичных для области способов фильтрации данных, например очистки кривых от выбросов; разработка способов корректировки сдвигов в кривых ГИС, возникающих вследствие проведения измерений разными приборами в разных условиях.
- Для изображений – нахождение и корректировка дефектов, определение различных видов вкраплений внутри пород.
- Корректировка дисбаланса классов горных пород при построении предсказывающих моделей.

5. Список источников

- [1] Deep neural networks for ring artifacts segmentation and corrections in fragments of CT images / A. Kornilov, I. Safonov, I. Reimers, I. Yakimchuk // 28th Conference of Open Innovation Association (FRUCT) (Moscow, 25-29 January 2021). IEEE, 2021. P. 181-193. DOI: <https://doi.org/10.23919/FRUCT50888.2021.9347587>.
- [2] Элланский М. М. Извлечение из скважинных данных информации для решения поисково-разведочных задач нефтегазовой геологии. М.: РГУ нефти и газа им. Губкина, 2000. 80 с.
- [3] Косков В. Н., Косков Б. В. Геофизические исследования скважин и интерпретация данных ГИС. Пермь: Изд-во Перм. гос. техн. ун-та, 2007. 317 с.

- [4] Клопов А. В. Особенности экспорта геологической информации в цифровую эпоху [Электронный ресурс] // Молодой ученый. 2018. №. 41. С. 7-9. URL: <https://moluch.ru/archive/227/53043/> (дата обращения: 03.10.2023).
- [5] Feng R. Uncertainty analysis in well log classification by Bayesian long short-term memory networks // *Journal of Petroleum Science and Engineering*. 2021. V. 205. P. 108816. DOI: <https://doi.org/10.1016/j.petrol.2021.108816>.
- [6] A Comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan / T. Merembayev, D. Kurmangaliyev, B. Bekbauov, Y. Amanbek // *Energies*. 2021. V. 14. №. 7. P. 1896. DOI: <https://doi.org/10.3390/en14071896>.
- [7] Imamverdiyev Y., Sukhostat L. Lithological facies classification using deep convolutional neural network // *Journal of Petroleum Science and Engineering*. 2019. V. 174. P. 216-228. DOI: <https://doi.org/10.1016/j.petrol.2018.11.023>.
- [8] Dunham M. W., Malcolm A., Kim Welford J. Improved well-log classification using semisupervised label propagation and self-training, with comparisons to popular supervised algorithms // *Geophysics*. 2020. V. 85. №. 1. P. O1-O15. DOI: <https://doi.org/10.1190/geo2019-0238.1>.
- [9] Hall B. Facies classification using machine learning // *The Leading Edge*. 2016. V. 35. №. 10. P. 906-909. DOI: <https://doi.org/10.1190/tle35100906.1>.
- [10] Lithology prediction using well logs: A granular computing approach / T. M. Hossain [et al.] // *Int. J. Innov. Comput. Inf. Control*. 2021. V. 17. №. 1. P. 225-244. DOI: <https://doi.org/10.24507/ijicic.17.01.225>.
- [11] Martin T., Meyer R., Jobe Z. Centimeter-scale lithology and facies prediction in cored wells using machine learning // *Frontiers in Earth Science*. 2021. V. 9. P. 659611. DOI: <https://doi.org/10.3389/feart.2021.659611>.
- [12] Analysis of ensemble methods applied to lithology classification from well logs / V. R. Leite, P. M. C. Silva, M. Gattass, A. C. Silva // 13th International Congress of the Brazilian Geophysical Society & EXPOGEF (Rio de Janeiro, Brazil, 26–29 August 2013). Society of Exploration Geophysicists and Brazilian Geophysical Society, 2013. P. 949-952. DOI: <https://doi.org/10.1190/sbgf2013-196>.
- [13] Ippolito M., Ferguson J., Jenson F. Improving facies prediction by combining supervised and unsupervised learning methods // *Journal of Petroleum Science and Engineering*. 2021. V. 200. P. 108300. DOI: <https://doi.org/10.1016/j.petrol.2020.108300>.
- [14] Interpreting the subsurface lithofacies at high lithological resolution by integrating information from well-log data and rock-core digital images / J. Jeong [et al.] // *Journal of Geophysical Research: Solid Earth*. 2020. V. 125. №. 2. P. e2019JB018204. DOI: <https://doi.org/10.1029/2019JB018204>.
- [15] Petrophysical characterisation of the Neoproterozoic and Cambrian successions in the Officer Basin / L. Wang [et al.] // *The APPEA Journal*. 2022. V. 62. №. 1. P. 381-399. DOI: <https://doi.org/10.1071/AJ21076>.
- [16] McDonald A. Data Quality Considerations for Petrophysical Machine-Learning Models // *Petrophysics*. 2021. V. 62. №. 06. P. 585-613. DOI: <https://doi.org/10.30632/PJV62N6-2021a1>.
- [17] Hallam A., Mukherjee D., Chassagne R. Multivariate imputation via chained equations for elastic well log imputation and prediction // *Applied Computing and Geosciences*. 2022. V. 14. P. 100083. DOI: <https://doi.org/10.1016/j.acags.2022.100083>.
- [18] Lopes R. L., Jorge A. M. Assessment of predictive learning methods for the completion of gaps in well log data // *Journal of Petroleum Science and Engineering*. 2018. V. 162. P. 873-886. DOI: <https://doi.org/10.1016/j.petrol.2017.11.019>.
- [19] Synthetic sonic log generation with machine learning: A contest summary from five methods / Y. Yu [et al.] // *Petrophysics*. 2021. V. 62. №. 04. P. 393-406. DOI: <https://doi.org/10.30632/PJV62N4-2021a4>.
- [20] Sonic Waves Travel-time Prediction: When Machine Learning Meets Geophysics / W. K. Wong, Y. Nuwara, F. H. Juwono, F. Motalebi, // 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST) (Miri Sarawak, Malaysia, 26-28 October 2022). IEEE, 2022. P. 159-163. DOI: <https://doi.org/10.1109/GECOST55694.2022.10010361>.

- [21] Application of conditional generative model for sonic log estimation considering measurement uncertainty / J. Jeong [et al.] // *Journal of Petroleum Science and Engineering*. 2021. V. 196. P. 108028. DOI: <https://doi.org/10.1016/j.petrol.2020.108028>.
- [22] CoreScore: a machine learning approach to assess legacy core condition / M. Fellgett [et al.] // Geological Society, London, Special Publications. 2024. V. 527. №. 1. P. SP527-2021-200. DOI: <https://doi.org/10.1144/SP527-2021-200>.
- [23] Framework for automatic globally optimal well log correlation / O. Datskiv [et al.] // *Neural Information Processing Systems (NeurIPS) Workshop on AI for Earth Sciences*. 2020. P. 1-5.