

Выявление и анализ перспективной задачи исследования по теме «Цифровая трансформация нефтегазового сектора»

Б.Н. Чигарев

Институт проблем нефти и газа РАН, г. Москва, Россия

E-mail: bchigarev@ipng.ru

Аннотация. Данная работа посвящена выявлению и анализу одной из актуальных задач научного исследования темы «Цифровая трансформация нефтегазового сектора». Актуальная область и методы исследования были определены по материалам Общества инженеров-нефтяников, для раскрытия научного аспекта выявленной актуальной задачи использовались публикации, индексируемые в библиографической и реферативной базе данных рецензируемой научной литературы Scopus. Такой подход, по мнению автора статьи, снижает предвзятость при выборе актуальных научных задач, поскольку отражает как интересы нефтегазового сектора, так и экспертов научного сообщества. В данной работе актуальная проблематика исследования определялась как по категориям классификации платформы search.spe.org, так и по авторским ключевым словам статей, проиндексированных в Scopus. Для более детального раскрытия научных аспектов таких исследований был проведен краткий обзор наиболее цитируемых статей, относящихся к рассматриваемой теме. Результаты проведенного анализа показывают, что в рамках широкой темы цифровой трансформации нефтегазового сектора для индустрии (Общества инженеров-нефтяников) и научного сообщества актуальной является задача, которую можно отразить термином «well drilling» (бурение скважин) и методом ее анализа «random forest» (случайного леса).

Ключевые слова: цифровая трансформация, нефтегазовый сектор, актуальные задачи исследования, библиометрический анализ, Scopus, VOSviewer.

Для цитирования: Чигарев Б.Н. Выявление и анализ перспективной задачи исследования по теме «Цифровая трансформация нефтегазового сектора» // Актуальные проблемы нефти и газа. 2023. Вып. 1(40). С. 67–88. <https://doi.org/10.29222/ipng.2078-5712.2023-40.art6>

Определение цели данной публикации

В 2019 году Минэнерго России был сформирован Совет по цифровой трансформации отраслей топливно-энергетического комплекса, обеспечивающий выработку единой позиции и ключевых решений по данному вопросу¹.

За прошедшие годы накоплен значительный опыт научных исследований

в области цифровой трансформации нефтегазового сектора, поэтому представляет интерес проведение анализа по выявлению актуальных научных задач в данной области исследований.

В рамках настоящей работы поставлена цель выявить и проанализировать одну из актуальных научных задач исследований, опираясь на международный опыт, изложенный как в материалах Общества инженеров-нефтяников (SPE), так и публикаций, проиндексированных в реферативной базе Scopus.

¹ <https://minenergo.gov.ru/node/14559> – ведомственный проект «Цифровая энергетика».

При этом материалы публикаций SPE использовались для выявления актуальной для нефтегазового сектора задачи, а материалы Scopus – для анализа публикаций, раскрывающих научную сторону исследований по выявленной актуальной задаче.

Введение

Чтобы оценить доминирующие задачи научных исследований по теме Digital and Oil&Gas², но не сужать тему до вынесенной в заголовке этой статьи, задаем очень широкий запрос к Scopus, ограничившись публикациями 2021 года. По запросу «TITLE-ABS-KEY (digital AND oil&gas) AND (LIMIT-TO (PUBYEAR, 2021))» получаем 524 результата, что вполне достаточно для небольшого обзора для раздела Введение. Публикации за 2021 год, с одной стороны, – достаточно новые, с другой – было время их процитировать. Из данного количества публикаций 126 относятся к публикациям китайских авторов, 83 – российских и 61 – авторов из США. Цитируемость публикации отражает проявленный к ней интерес со стороны экспертного сообщества, поэтому ограничиваемся данными по публикациям, которые процитированы минимум 10 раз. Таких публикаций оказалось 35.

Приведем список десяти чаще всего встречаемых авторских ключевых слов, взятых из данных публикаций: machine learning, artificial intelligence, drilling and completion, logging interpretation, oil and gas industry, reservoir engineering, seismic exploration, surface facility engineering, deep

learning, rock physics (см. Прил. 1). Эти данные нужны для сравнения с результатами, полученными по схеме:

- 1) запрос к публикациям SPE, отражающий отраслевой интерес;
- 2) выявление задач, интересных для отрасли;
- 3) анализ актуальных исследовательских задач, выявленных по данным научных публикаций, проиндексированных в Scopus.

Разнообразие подходов во введении и основной части статьи позволяет снизить предвзятость в оценке актуальности научных задач, относящихся к цифровой трансформации нефтегазового сектора. Если результаты получатся сопоставимыми, это можно трактовать как их устойчивость, так как выборки для анализа формировались существенно различными способами.

Для обзора опубликованных статей рассмотрим те из них, в которых встречаются перечисленные выше доминирующие авторские слова.

Из 35 публикаций статья [1] является наиболее цитируемой для ключевого слова «machine learning» (43 цитирования). Работа посвящена ускорению разработки новых металлоорганических структур (MOFs) на основе методов высокопроизводительного вычислительного скрининга (HTCS). Как показали авторы статьи, применение машинного обучения (ML) к HTCS MOFs эффективно не только для выявления скрытых связей между компонентами структуры, но и для технологий их применения, в частности, для хранения и разделения газа. Металлоорганические структуры были названы Международным союзом теоретической и прикладной химии (IUPAC) в числе десяти лучших новых технологий в химии [2].

² <https://smartgopro.com/digitalgasoil/> – отраслевая конференция и выставка по цифровым технологиям для топ-менеджеров нефтегазового сектора.

Авторы статьи [3] анализируют, как применение искусственного интеллекта может изменить нефтегазовую отрасль (ключевые слова: artificial intelligence, oil and gas industry, 39 цитирований). Основное внимание уделяется сегменту разведки и добычи, как наиболее капиталоемкой части нефтегазовой отрасли, и сегменту с огромным количеством неопределенностей, с которыми приходится сталкиваться. На основе анализа применения искусственного интеллекта (ИИ) и обзора существующих внедрений ИИ авторы анализируют последние тенденции в развитии инструментов на основе ИИ и их влияние на снижение рисков в отрасли. Для российских авторов более характерно изучение проблем именно нефтегазового сектора, а не актуальных задач чистой энергетики, примером которой может являться предыдущая публикация.

«Drilling and completion» является классической темой для нефтегазовой отрасли; наиболее цитируемая работа, содержащая данное ключевое слово, – статья [4] (22 цитирования, также включает ключевые слова: logging interpretation, reservoir engineering, seismic exploration, surface facility engineering). В данной статье описывается применение искусственного интеллекта в нефтегазовой разведке и разработке месторождений, а также обсуждаются области применения и направления развития ИИ в будущем. Машинное обучение было применено в идентификации литологии, реконструкции каротажных кривых, оценке параметров коллекторов. Глубокое машинное обучение и технология оптимизации применимы в разработке пластов, оптимизации процесса заводнения и прогнозирования добычи нефти и газа,

особенно в режиме реального времени. Данная обзорная публикация хорошо отражает области применения ИИ в нефтегазовом секторе, но не раскрывает, какие методы ИИ чаще всего реализуются при решении перечисленных в статье задач.

Точная оценка проницаемости породы имеет решающее значение для разработки и управления нефтяными и газовыми пластами. Авторами публикации [5] (25 цитирований, ключевые слова: deep learning, rock physics) была разработана методика, основанная на ИИ для быстрой и точной оценки проницаемости неоднородных карбонатных пород по рентгеновскому микротомографическому изображению micro-CT (X-ray micro-CT). Авторы обнаружили, что прогнозируемая проницаемость карбонатов с помощью машинного обучения очень хорошо согласуется с проницаемостью вычислительно более трудоемкого прямого моделирования на основе «voxel-based modeling». Разработанная модель ML обеспечивает сокращение времени вычислений примерно на три порядка по сравнению с Lattice Boltzmann Method. Авторы утверждают, что предложенная схема, объединяющая различные алгоритмы обучения, визуализацию горных пород и моделирование, обладает потенциалом для быстрой и точной оценки петрофизических свойств при моделировании и определении характеристик пласта.

Приведенный выше краткий обзор показывает, что простой запрос и не менее простой анализ позволяют выявить интересные публикации, которые могут служить отправной точкой для дальнейшего сбора материалов по теме, заинтересовавшей специалиста.

Но такой подход не позволяет выявить и проанализировать, какие задачи являются приоритетными для данного класса экспертов (в нашем случае SPE) и какие методы решения данных задач чаще всего используются научным сообществом (публикации, проиндексированные в Scopus). Это и явилось мотивацией провести исследование с использованием библиометрических методов по выявлению актуальной задачи в такой широкой теме, как «Цифровая трансформация нефтегазового сектора», которая соответствовала бы как интересам индустрии (SPE), так и научного сообщества (Scopus).

Естественно, в одной публикации невозможно раскрыть все стороны озвученной выше проблемы, но вполне возможно показать подход к ее решению.

Материалы и методы

В качестве источника публикаций SPE использовался сервис search.spe.org, предоставляющий открытый доступ к системе, работающей на основе искусственного интеллекта, которая сочетает в себе машинное обучение и знания экспертов в конкретной области, для получения точной, достоверной и последовательной информации о содержании OnePetro, PetroWiki, журналов SPE и веб-сайта³.

Формирование запроса

Результаты запросов к системе search.spe.org без дополнительных ограничений в формате: название термина (число результатов в выдаче):

- digital transformation (1,038);
- digital technology (614);
- digital technologies (655);
- digital engineering (192);
- Digitalization (1,311);
- Industry 4.0 (226).

Видно, что в системе слабо работает стемминг⁴ – результаты для «digital technologies» (655) и «digital technology» (614) немного отличаются. Такое нередко встречается в системах поиска, в которых упор делается на ИИ, но недостаточное внимание уделяется предобработке исходных данных/текстов. «Digitalization» – часто употребляемый термин. Он носит более общий характер по сравнению с «digital transformation». Учитывая, что ресурс search.spe.org – сугубо специализированный и для выбора интересной и актуальной задачи научных исследований желательно иметь большую выборку библиометрических данных, он был включен в итоговый запрос.

Итоговый запрос: «“digital transformation” OR “digital technologies” OR “digital technology” OR “digital engineering” OR “Industry 4.0” OR “Digitalization”» – по нему получено 2873 записи, что больше половины суммы результатов отдельных запросов (4036). Это говорит о высокой связанности терминов в запросе и подчеркивает их отнесение к одной тематике.

Из 2873 публикаций только 186 относятся к рецензируемым статьям, то есть доминируют отраслевые материалы, а не научные публикации. Таким образом, актуальность выбираемой задачи основывается на интересах отрасли, а не научных статьях.

³ <https://onepetro.org/> – онлайн-библиотека технической литературы по нефтегазовой тематике.

⁴ Стемминг – это процесс нахождения основы слова для заданного исходного слова (<https://ru.wikipedia.org/wiki/Стемминг>).

Некоторые характеристики выборки, которые использованы для выбора актуальной задачи для детального анализа

Искусственный интеллект платформы search.spe.org позволяет отнести публикации к определенным категориям, входящим в 14 классов. Для обоснования выбора актуальной задачи исследования воспользуемся данными следующих классов: SPE Disciplines, Concept Tag, Industry и Technology. При выборе актуальной задачи исследований исходим из предположения, что актуальность можно оценить числом публикаций, относящихся к определенной категории. Для конкретизации ограничений, к каким категориям должны относиться публикации, по рассмотрению которых выбиралась актуальная задача исследований, для каждого класса оставались по 2–4 категории с наибольшим числом публикаций.

SPE Disciplines

Для SPE Disciplines платформа search.spe.org смогла классифицировать 2811 публикаций. Из них 2072 отнесены к категории Data Science & Engineering Analytics, 984 – к Management, 791 – к Reservoir Description and Dynamics, а следующие категории решено было объединить, так как они все имеют отношение к скважинам: Production and Well Operations (511), Well Completion (263) и Well Drilling (525).

Здесь применение науки о данных и инженерной аналитики к задачам эксплуатации скважин явно доминирует.

На этом этапе не рассматривается совместная фильтрация по совокупности категорий, так как просто нужны ограничения тематик для отбора научных публикаций, проиндексированных в Scopus.

Concept Tag

Для Concept Tag имеем: Upstream Oil & Gas – 2474, Artificial Intelligence – 1397, Operation – 735, Application – 675, Petroleum Engineer – 627 публикаций, относящихся к данной категории.

Вышеприведенные результаты можно интерпретировать как использование искусственного интеллекта для решения инженерных задач в сфере разведки и добычи нефти и газа.

Industry

В данном случае доминирует категория Energy (2731), ее подкатегория Oil & Gas (2685) и подподкатегория Upstream (2469).

Upstream, как самая низшая ветвь классификации, преобладает в Industry так же, как и в Concept Tag.

Technology

В данном случае доминирует категория Information Technology (2679), ее подкатегория Artificial Intelligence (1309) и две подподкатегории: Machine Learning (534) и Representation & Reasoning (439).

Чем ниже подкатегория, тем более узок класс задач, в нее входящих. Поэтому в дальнейшем используем именно их для формирования запроса по сбору научных публикаций.

Распределение категорий в четырех классах хорошо между собой согласуется.

На основании этого можно предположить, что для инженеров нефтегазовой отрасли актуальной задачей является использование машинного обучения, представления знаний и логических выводов (Machine Learning и Representation & Reasoning) в задачах эксплуатации скважин (Production and Well Operations, Well Completion и Well Drilling), т. е. выбраны наиболее конкретные подкатегории.

Проведенный анализ публикаций, проиндексированных на платформе search.spe.org, позволил сформулировать класс задач, наиболее актуальных с позиций SPE по теме «Цифровая трансформация нефтегазового сектора».

Следующий этап анализа библиометрических данных по выявлению актуальной научной исследовательской задачи, которую целесообразно изучить более детально, был проведен с использованием результатов запросов к системе Scopus.

В Scopus рассматривались публикации, проиндексированные после 2017 года.

Запросы формировались по наличию выбранных терминов в заголовках, аннотациях и ключевых словах публикации (TITLE-ABS-KEY).

Самый широкий запрос к базе Scopus, содержащий термины «well AND drilling», дал 21368 результатов. Данный запрос отражает интерес к объекту исследований.

Методы исследований, применяемые в публикациях, можно проанализировать, задав запрос: «Machine Learning OR (representation AND reasoning)», по которому Scopus выдал 322528 результатов. При этом запрос с терминами «representation AND reasoning» дает только 5688 результатов, а с «Machine Learning» – 317306. Пересечение между этими запросами небольшое – 725 публикаций, поэтому в итоговом запросе оставляем и «Machine Learning», и «representation AND reasoning».

Окончательно объединяя запросы, относящиеся к методам исследования и объекту исследования, получаем запрос: «TITLE-ABS-KEY ((well AND drilling) AND (“Machine Learning” OR (representation AND reasoning))) AND PUBYEAR > 2017»,

дающий 725 записей, из которых 688 написаны на английском языке (актуально на 30.11.2022).

Выборка из 688 записей использовалась для выявления актуальной темы исследований.

В качестве инструментария применялись программа для проведения библиометрических исследований VOSviewer [6], программа построения графики Scimago Graphica [7] и утилиты для работы со строками.

Результаты

Актуальные задачи исследований могут быть описаны разными терминами: авторскими ключевыми словами, индексными ключевыми словами платформы Scopus, терминами, извлеченными из текстов заголовков и аннотаций публикаций. В данной работе акцент сделан на использовании авторских ключевых слов (Author Keywords). Авторов можно рассматривать как экспертов в своей предметной области, которые стараются выбрать ключевые слова, наиболее точно, с их точки зрения, описывающие содержание статьи.

Это не умаляет важности использования других терминов для всестороннего библиометрического анализа, просто задача данной публикации ограничена выявлением и кратким анализом одной актуальной исследовательской задачи в рамках темы «Цифровая трансформация нефтегазового сектора».

Кластеризация терминов, используемых для выявления актуальной задачи, основывалась на их совместной встречаемости в поле Author Keywords. Результаты представлены на рис. 1.

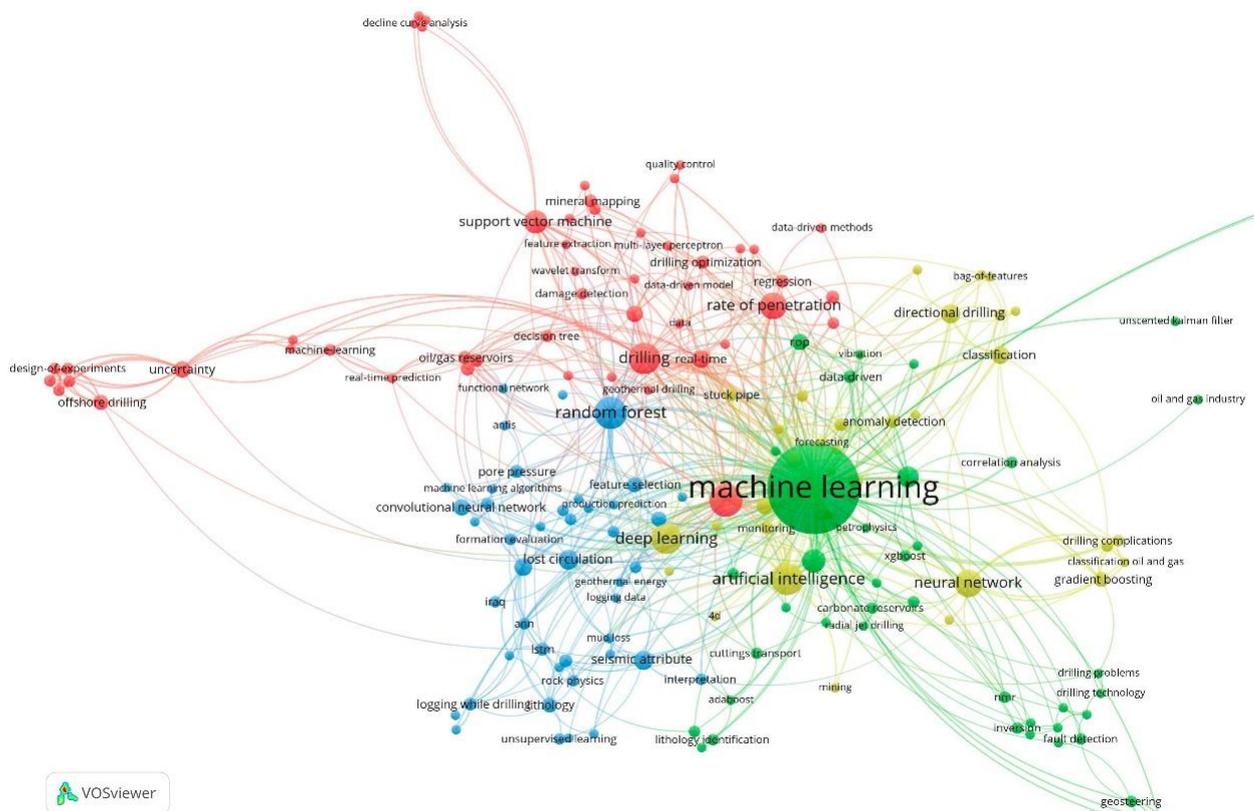


Рис. 1. Четыре кластера сети авторских ключевых слов, полученных на основе их совместной встречаемости

Всего в 688 библиометрических записях, экспортированных из Scopus, содержалось 1155 авторских ключевых слов, из которых 175 встречались два и более раза и только 29 – пять и более раз. Предварительно проводился частичный препроцессинг Author Keywords, например, термины во множественном числе заменялись на термины в единственном числе.

Авторские ключевые слова в выборке получились очень разнородными. Как видно из приведенных выше данных, только 175 из них встречаются два и более раза. Поэтому при выборе параметров по умолчанию программа VOSviewer выдавала 18 кластеров, в самом маленьком из которых содержался лишь один термин, при выборе параметра наименьшего числа

терминов в кластере в диапазоне 25–30 устойчиво получалось 4 кластера, что и использовалось в дальнейшей работе.

В табл. 1 представлены наиболее часто встречаемые термины для каждого из четырех кластеров и их характеристики.

Сопоставление названий полей в таблице и названий в экспортируемых из VOSviewer файлах:

- Название термина (label),
- № кластера (cluster),
- Общая сила связи (weight<Total link strength>),
- Встречаемость термина (weight<Occurrences>),
- Средний год публикации (score<Avg. pub. year>),
- Средняя нормализованная цитируемость (score<Avg. norm. citations>).

Таблица 1

10 наиболее часто встречаемых терминов для каждого кластера и их характеристики

Название термина	№ кластера	Общая сила связи	Встречаемость термина	Средний год публикации	Средняя нормализованная цитируемость
<i>Первый кластер, цвет красный, общее число терминов в кластере 50</i>					
artificial neural network	1	54	24	2021,1	1,5073
drilling	1	65	22	2020,1	1,2109
rate of penetration	1	37	16	2021	1,8462
support vector machine	1	30	12	2019,8	1,6723
real-time	1	31	7	2020,9	1,3735
particle swarm optimization	1	14	6	2021	2,0472
uncertainty	1	29	6	2020,3	1,279
offshore drilling	1	20	5	2020,4	0,0671
data mining	1	8	4	2021	1,3853
drilling optimization	1	12	4	2019,8	3,5014
<i>Второй кластер, цвет зеленый, общее число терминов в кластере 47</i>					
machine learning	2	333	184	2020,7	1,3863
well log	2	18	12	2020,8	1,9429
optimization	2	31	10	2020,2	2,2427
ROP	2	19	6	2019,3	3,8199
data-driven	2	16	4	2019,8	3,2969
carbonate reservoirs	2	5	3	2022	0,2988
constraints	2	6	3	2020,3	0
correlation analysis	2	4	3	2020,7	1,7659
cuttings transport	2	8	3	2021	0,565
fault detection	2	6	3	2020,3	0,798
<i>Третий кластер, цвет синий, общее число терминов в кластере 44</i>					
random forest	3	55	23	2020,8	1,3591
lost circulation	3	24	9	2020,1	2,524
seismic attribute	3	16	8	2020,3	0,9416
drilling parameters	3	20	7	2021	1,5638
convolutional neural network	3	8	6	2021	2,4845
feature selection	3	8	5	2020,6	3,1978
lithology	3	12	5	2021	1,9907
shale gas	3	7	5	2021,6	2,2762
logging while drilling	3	11	4	2020,5	0,9923
LSTM	3	12	4	2019,8	1,3448
<i>Четвертый кластер, цвет хаки, общее число терминов в кластере 33</i>					
artificial intelligence	4	64	23	2020,7	1,5927
deep learning	4	40	20	2020,95	1,6511
neural network	4	41	17	2020,4	0,2695
directional drilling	4	21	8	2020,6	2,427
classification	4	23	7	2020,3	1,9617
drilling operation	4	21	6	2021	1,9063
unsupervised machine learning	4	3	6	2021,5	2,2076
gradient boosting	4	20	5	2020,6	1,7196
logging	4	16	5	2020,8	0,2013
anomaly detection	4	16	4	2020	1,4259

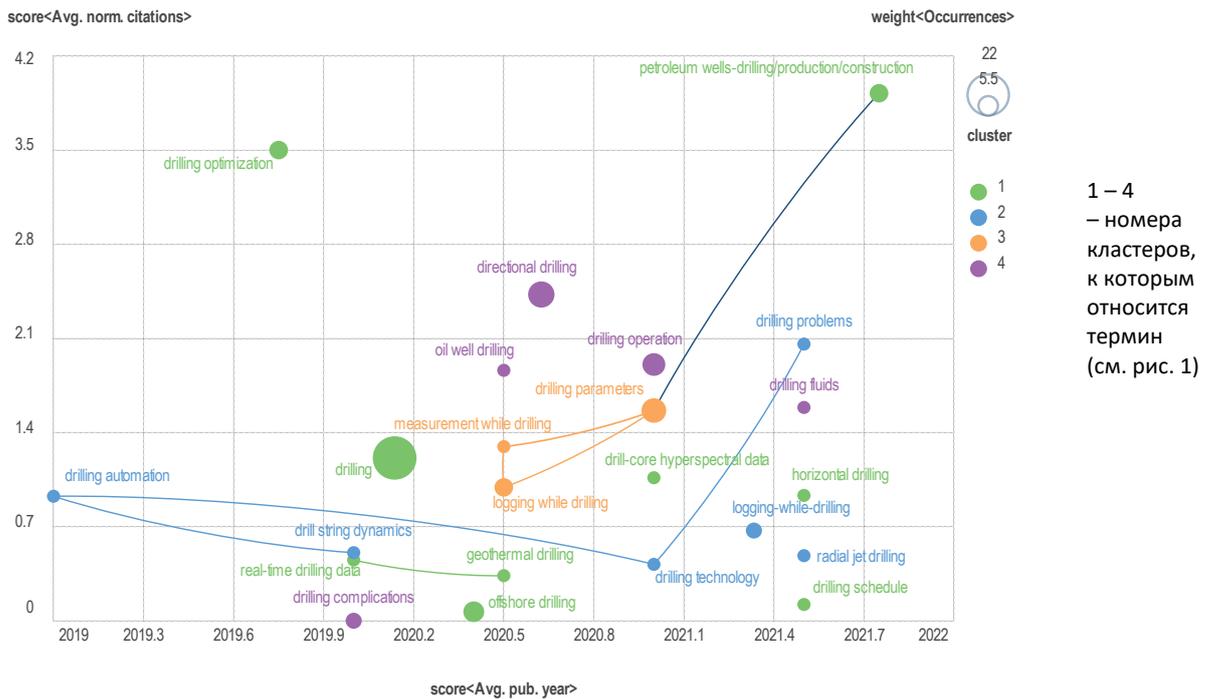


Рис. 4. Диаграмма, характеризующая цитируемость терминов, содержащих слово «drill» в координатах: средний год встречаемости термина как авторского ключевого слова (Avg. pub. year) и средняя нормализованная цитируемость публикаций, содержащих термин как авторское ключевое слово (Avg. norm. citation). Размер пузырька пропорционален частоте встречаемости термина (weight <Occurrences>)

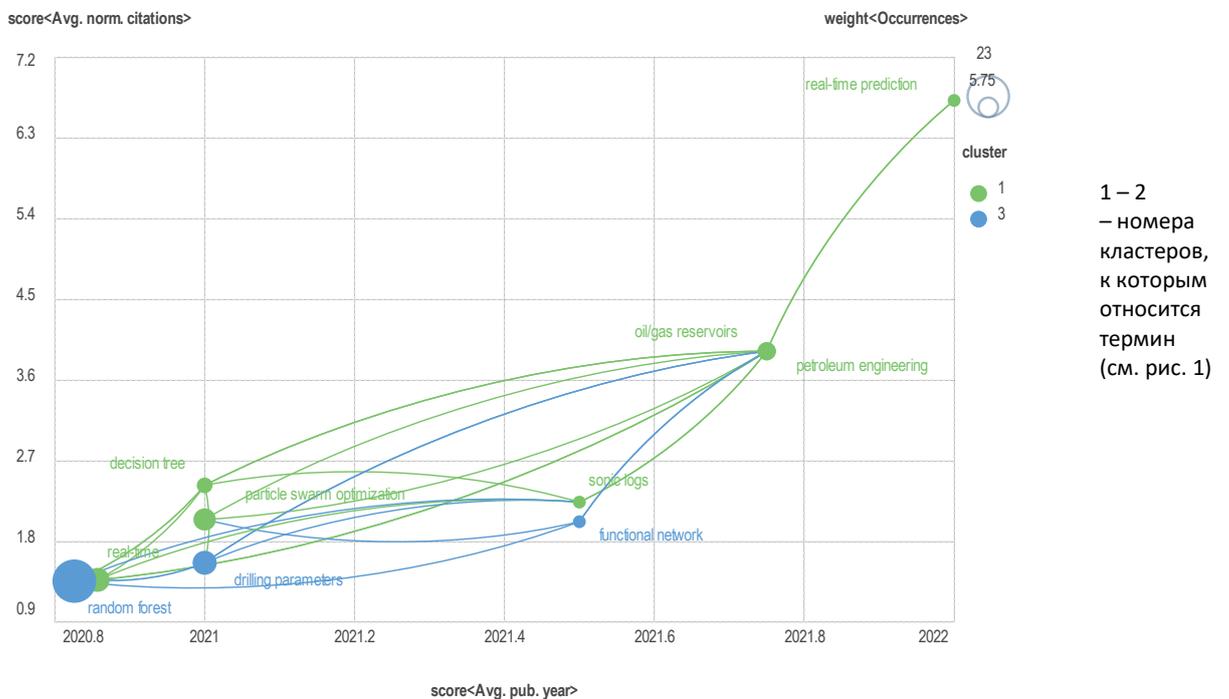


Рис. 5. Диаграмма, отражающая наиболее часто встречаемые авторские ключевые слова в публикациях, содержащих «petroleum wells-drilling» как авторское ключевое слово. Координаты на диаграмме: средний год встречаемости термина как авторского ключевого слова (Avg. pub. year) и средняя нормализованная цитируемость публикаций, содержащих термин как авторское ключевое слово (Avg. norm. citation). Размер пузырька пропорционален частоте встречаемости термина (weight <Occurrences>)

«Random forest» является наиболее часто встречаемым термином в данной выборке. «Real-time» и «real-time prediction» также являются значимыми терминами. Близкая связанность «random forest» и «drilling parameters» может указывать, в каких задачах используется данный метод классификации. Метод «random forest» является развитием метода «decision tree», это может объяснить их совместную встречаемость в одной выборке.

Термины «sonic logs», «oil/gas reservoirs» и «petroleum engineering» логически сочетаются с выбранным автором статьи для выявления актуальной задачи термином «petroleum wells-drilling».

Возможной актуальной задачей может служить использование метода «random forest» в определении оптимальных параметров бурения, в том числе в условиях реального времени. Эти ключевые термины присутствуют на рис. 2–5.

Целесообразно дать некоторое пояснение к ключевому слову «functional network», которое в последнее время нередко встречается для описания анатомической и функциональной организации мозга в терминах сложных сетей. Часто целью подобных моделей является выявление различий между возможными условиями функционирования элементов такой сети. В основе подобных моделей лежит определение дискриминационных признаков и количественная оценка различий между ними. Более детально с сущностью «functional network» можно ознакомиться в работе [9]. В текущем контексте использование подходов «functional network» может быть интересно для обоснования выбора параметров бурения («drilling parameters» на рис. 5).

Но данная задача скорее относится к перспективным/поисковым исследованиям, нежели к актуальным/востребованным, выявлению которых посвящена статья.

Из собранных данных, которые, согласно основному запросу, относятся к задачам бурения «well AND drilling», выбираем публикации, содержащие «random forest» в авторских ключевых словах.

Дадим краткое описание тематики каждой из публикаций.

Увеличение скорости проходки «rate of penetration» (ROP) сокращает время, необходимое для бурения, что приводит к снижению затрат. Эмпирические формулы, разработанные для прогнозирования ROP, имеют ограниченное применение в полевых условиях. Методы сбора данных о бурении в реальном времени и вычислительные технологии с годами совершенствуются, поэтому авторы статьи [10] применили подход к прогнозированию, основанный на построении модели данных, а не на эмпирических формулах.

После оптимизации гиперпараметров метод «random forest regressor» с 300 деревьями решений дал лучшие результаты, чем «artificial neural network» (ANN), для проанализированного набора данных. Метод «random forest» строит несколько деревьев решений и объединяет их вместе, чтобы получить более точный и стабильный прогноз. Он широко используется и входит во многие аналитические пакеты. Большим преимуществом метода случайного леса является то, что его можно использовать как для классификации, так и для задач регрессии. Более детально с информацией о данном методе можно ознакомиться в публикациях [11, 12]. Первая процитирована в Scopus 62394 раза, а вторая – 2444 раза (актуально на 30.11.2022).

Сопоставим ключевые слова статьи: «rate of penetration, predictive modeling, geothermal energy, machine learning, deep learning, random forests, artificial neural network, python programming» с терминами, представленными на графике (см. рис. 5). К термину на графике «drilling parameters» можно отнести ключевое слово «rate of penetration», к «functional network» – «artificial neural network», область применения «geothermal energy» является востребованной темой чистой энергетики, наиболее близкой к задачам геологии и разработки из всех других методов получения энергии, относящихся к чистым.

Таким образом, ключевые термины статьи имеют сходство с терминами, представленными на рис. 5. И статья может рассматриваться как относящаяся к актуальной тематике.

В работе [13] авторы рассматривали проблему картирования пород при анализе гиперспектральных данных керна и их классификации по нескольким меткам. Классификация по нескольким меткам осуществлялась с использованием алгоритма random forest (RF). Для создания обучающего набора данных использовались минералогические данные сканирующей электронной микроскопии (SEM) высокого разрешения. Гиперспектральные данные охватывали видимый и ближний инфракрасный (VNIR) и коротковолновый инфракрасный (SWIR) диапазоны электромагнитного спектра. Таким образом, данная статья раскрывает еще один актуальный аспект, связанный с бурением, в котором алгоритм random forest используется для классификации данных по керну.

Цель работы [14] заключалась в разработке модели машинного обучения для прогнозирования результатов акустического каротажа по данным бурения

в режиме реального времени. В качестве алгоритмов обучения для построения моделей прогнозирования результатов акустического каротажа по данным бурения сложных разрезов, включающих известняки, песчаники, сланцы и карбонатные формации, были использованы алгоритмы Decision Tree (DR) и random forest (RF). Результаты показали, что разработанные модели обладают высокой способностью предсказывать звуковые сигналы по данным бурения. Прогнозирование акустического сигнала с помощью разработанных моделей позволяет сэкономить затраты и время на сбор данных и обеспечивает оценку параметров бурения в режиме реального времени. Как отмечалось выше, random forest фактически является усовершенствованным методом Decision Tree, поэтому и на графиках, и часто в текстах эти термины встречаются совместно. Данная статья раскрывает третий аспект использования метода random forest – построение моделей прогнозирования результатов акустического каротажа по данным бурения в режиме реального времени. В статье [15] авторы предлагают алгоритм вероятностного случайного леса (Probabilistic Random Forest) и исследуют его преимущества по сравнению с восемью существующими алгоритмами интерпретации литологии каротажа. Сравнительные эксперименты на 9 реальных задачах интерпретации литологических данных каротажа подтвердили целесообразность и преимущества предложенного метода. Авторы утверждают, что вероятностная модель способна предоставить больше информации о неоднородности и свойствах пласта. Пояснение: алгоритм Probabilistic Random Forest (PRF) рассматривает признаки и метки как функции распределения вероятностей, а не как детерминированные величины [16].

Ценность данной работы заключается не только в еще одном объекте исследования (литологической интерпретации данных каротажа), но и в модификации самого метода «Probabilistic random forest», предоставляющего, по мнению авторов статьи, больше информации о свойствах пласта.

Суммируя краткое рассмотрение содержания четырех публикаций, можно утверждать, что выбранная с использованием библиометрического метода актуальная задача по использованию метода «random forest» для решения проблем, требующих решения в реальном времени (real time и real-time prediction) и связанных с бурением скважин (drilling parameters), является адекватной. В скобках названия терминов приводятся в том виде, в котором они представлены на графиках (см. рис. 5 и 6).

Контекст процесса цифровой трансформации нефтегазового сектора может быть очень широким, но актуальные задачи, учитывающие как интересы промышленности (SPE), так и научные наработки (Scopus), чаще всего бывают весьма конкретными и приземленными. Хорошо отработанные и легко интерпретируемые методы типа «random forest» могут быть не менее актуальными, чем трудно интерпретируемые и требующие больших выборок для этапа обучения методы с использованием нейронных сетей.

«Приземленные» вопросы, связанные с процессом бурения, могут быть не менее актуальными в промышленных условиях, чем создание цифровых двойников. Вторые задачи можно скорее отнести не к актуальным/востребованным, а к перспективным/поисковым исследованиям,

которые позволяют занять нарождающиеся рынки и наметить темы исследований. Нарождающиеся темы исследований «Emerging research topics» требуют несколько иного подхода к проведению библиометрического анализа. В этом случае постановка вопроса задается не текущими проблемами индустрии, а денежными потоками и политическими решениями, при этом темы библиометрически будут определяться не абсолютными значениями, а их ростом и диссеминацией.

Библиометрический анализ не рассчитан на получение «единственно верного решения», чаще всего он позволяет снизить предвзятость в принятии решений. В данном исследовании ставился вопрос о выборе задачи, относящейся к теме «Цифровая трансформация нефтегазового сектора» и востребованной/актуальной именно с точки зрения промышленности. Более того, библиометрические данные, экспортированные из Scopus, проанализированы с использованием авторских ключевых слов. Аналогичный анализ можно провести, используя индексные ключевые слова поля Index Keywords, кластеризацию ключевых терминов, полученных из текстов заголовков и аннотаций или, например, опираясь на кластеризацию публикаций, основанную на их библиографическом сочетании (bibliographic coupling). Результаты могут существенно различаться, но удовлетворять заданным ограничениям, например, на частоту встречаемости терминов, совместную встречаемость терминов или учитывать цитируемость публикаций.

Для примера рассмотрим результаты кластеризации публикаций, основанной на их библиографическом сочетании (Прил. 2).

Наложим следующие ограничения на выбор публикаций: статья должна быть опубликована в 2021 году (достаточно новая, но уже с накопленной цитируемостью), цитируемость – не ниже 30, а число библиографических связанных статей – не менее 50.

В нашей выборке данным условиям удовлетворяют две публикации. В первой статье [17] авторы рассматривают модель добычи сланцевого газа, построенную с помощью многоцелевого метода случайного леса. Ими показано, что многоцелевой метод случайного леса может работать с коррелированными данными о добыче сланцевого газа. Вторая статья [18] интересна тем, что в ней представлены результаты библиометрического анализа и подробный обзор по использованию контролируемого машинного обучения для анализа опасных событий при бурении. Проведенный авторами анализ показывает, что искусственная нейронная сеть является наиболее популярным алгоритмом среди исследователей. Кроме того, глубокое обучение (deep learning), метод случайного леса (random forest) и метод опорных векторов (support vector machine) становятся популярными в последнее время.

Тематика этих двух статей вполне согласуется с выводами, полученными ранее при анализе совместной встречаемости авторских ключевых слов.

При этом авторы последней статьи совершенно правы, утверждая, что «искусственная нейронная сеть является наиболее популярным алгоритмом среди исследователей», но если проанализировать, какие методы машинного обучения реализуются в проектах, получим подтверждение второго тезиса: «глубокое обучение (deep learning), случайный лес

(random forest) и метод опорных векторов (support vector machine) становятся популярными в последнее время».

Использование данных из разных полей таблицы библиометрических записей может существенно различаться. Приведем краткое сравнение встречаемости терминов, описывающих конкретные методы классификации данных в полях Index Keywords и Author Keywords.

Для индексных ключевых слов: «random forest» – 56; «support vector machine» – 104; «convolutional * networks» – 33 из 12748 общего числа терминов в столбце Index Keywords.

Для авторских ключевых слов: «random forest» – 25; «support vector machine» – 17; «convolutional * networks» – 7 из 1816 общего числа терминов в столбце Author Keywords.

Полученные приоритеты могут быть разными, но для обоих полей классические методы «random forest» и «support vector machine» встречаются чаще, чем «convolutional * networks».

Следует отметить, что полученные результаты относятся к конкретной теме бурения скважин, как актуальной для Общества инженеров-нефтяников (SPE).

Заключение

Проведенное библиометрическое исследование показало целесообразность использования двухстадийного подхода к выявлению актуальных исследовательских задач, при котором их общая направленность для индустрии задач определяется на основе анализа отраслевых публикаций. Последующий анализ научных публикаций, в рамках данных задач, позволит определить актуальные направления научных исследований.

Несмотря на очень широкую тему «Цифровая трансформация нефтегазового сектора», поставленную в начале исследования, предложенный двухстадийный библиометрический анализ позволил выявить задачу, интересную как для научных исследований, так и актуальную для нефтегазовой индустрии. Такой задачей является использование классических методов машинного обучения (в первую очередь метода случайного леса) для классификации проблем прогнозирования и оптимизации процессов бурения и эксплуатации скважин.

Актуальность задачи в данной работе оценивалась по цитируемости и количеству публикаций за последние 5 лет (2018–2022 гг.). Для научных исследований важна

не только текущая актуальность, но и выявление нарождающихся тематик (Emerging Topics).

Выявление тем требует несколько иного подхода при проведении библиометрического анализа. В этом случае необходимо учитывать, в развитие каких тем идет инвестирование, решение каких задач поддерживается политическими решениями. Нарождающиеся темы будут характеризоваться ростом числа публикаций, возрастанием цитируемости не просто отдельных публикаций, а направлений исследований, появлением новых научных журналов или тематикой спецвыпусков в престижных изданиях, тематикой конференций. Это – задача отдельного исследования.

Статья написана в рамках выполнения государственного задания ИПНГ РАН (тема «Фундаментальный базис энергоэффективных, ресурсосберегающих и экологически безопасных, инновационных и цифровых технологий поиска, разведки и разработки нефтяных и газовых месторождений, исследование, добыча и освоение традиционных и нетрадиционных запасов и ресурсов нефти и газа; разработка рекомендаций по реализации продукции нефтегазового комплекса в условиях энергоперехода и политики ЕС по декарбонизации энергетики (фундаментальные, поисковые, прикладные, экономические и междисциплинарные исследования)», № 122022800270-0).

Литература

1. Altintas C., Altundal O.F., Keskin S., Yildirim R. Machine learning meets with metal organic frameworks for gas storage and separation // Journal of Chemical Information and Modeling. 2021. Vol. 61, No. 5. P. 2131–2146. <https://doi.org/10.1021/acs.jcim.1c00191>
2. Gomollón-Bel F. Ten chemical innovations that will change our world: IUPAC identifies emerging technologies in Chemistry with potential to make our planet more sustainable // Chemistry International. 2019. Vol. 41, No. 2. P. 12–17. <https://doi.org/10.1515/ci-2019-0203>
3. Koroteev D., Tekic Z. Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future // Energy and AI. 2021. Vol. 3. P. 100041. <https://doi.org/10.1016/j.egyai.2020.100041>
4. Kuang L., Liu H., Ren Y. et al. Application and development trend of artificial intelligence in petroleum exploration and development // Petroleum Exploration and Development. 2021. Vol. 48, No. 1. P. 1–14. [https://doi.org/10.1016/S1876-3804\(21\)60001-0](https://doi.org/10.1016/S1876-3804(21)60001-0)

5. *Tembely M., AlSumaiti A.M., Alameri W.S.* Machine and deep learning for estimating the permeability of complex carbonate rock from X-ray micro-computed tomography // *Energy Reports*. 2021. Vol. 7. P. 1460–1472. <https://doi.org/10.1016/j.egy.2021.02.065>
6. *van Eck N.J., Waltman L.* Software survey: VOSviewer, a computer program for bibliometric mapping // *Scientometrics*. 2010. Vol. 84, No. 2. P. 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
7. *Hassan-Montero Y., De-Moya-Anegón F., Guerrero-Bote V.P.* SCImago Graphica: a new tool for exploring and visually communicating data // *El Profesional de la información*. 2022. Vol. 31, No. 5. P. e310502. <https://doi.org/10.3145/epi.2022.sep.02>
8. *van Eck N.J., Waltman L.* VOSviewer Manual: Manual for VOSviewer version 1.6.18. Leiden: Center for Science and Technology Studies, Leiden University, 2022. 53 p. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf (Дата обращения 03.12.2022).
9. *Zanin M., Güntekin B., Aktürk T.* et al. Telling functional networks apart using ranked network features stability // *Scientific Reports*. 2022. Vol. 12, No. 1. P. 2562. <https://doi.org/10.1038/s41598-022-06497-w>
10. *Ben Aoun M.A., Madarász T.* Applying machine learning to predict the rate of penetration for geothermal drilling located in the Utah FORGE site // *Energies*. 2022. Vol. 15, No. 12. P. 4288. <https://doi.org/10.3390/en15124288>
11. *Breiman L.* Random Forests // *Machine Learning*. 2001. Vol. 45, No. 1. P. 5–32. <https://doi.org/10.1023/A:1010933404324>
12. *Belgiu M., Drăguț L.* Random forest in remote sensing: A review of applications and future directions // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016. Vol. 114. P. 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
13. *Contreras I.C., Khodadadzadeh M., Gloaguen R.* Multi-label classification for drill-core hyperspectral mineral mapping // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020. Vol. XLIII-B3-2020. P. 383–388. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-383-2020>
14. *Gamal H., Alsaihati A., Elkatatny S.* Predicting the rock sonic logs while drilling by random forest and decision tree-based algorithms // *Journal of Energy Resources Technology*. 2022. Vol. 144, No. 4. P. 043203. <https://doi.org/10.1115/1.4051670>
15. *Ao Y., Zhu L., Guo S., Yang Z.* Probabilistic logging lithology characterization with random forest probability estimation // *Computers & Geosciences*. 2020. Vol. 144. P. 104556. <https://doi.org/10.1016/j.cageo.2020.104556>
16. *Reis I., Baron D., Shahaf S.* Probabilistic Random Forest: A machine learning algorithm for noisy datasets // *The Astronomical Journal*. 2019. Vol. 157, No. 1. P. 16. <https://doi.org/10.3847/1538-3881/aaf101>
17. *Xue L., Liu Y., Xiong Y.* et al. A data-driven shale gas production forecasting method based on the multi-objective random forest regression // *Journal of Petroleum Science and Engineering*. 2021. Vol. 196. P. 107801. <https://doi.org/10.1016/j.petrol.2020.107801>
18. *Osarogiagbon A.U., Khan F., Venkatesan R., Gillard P.* Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations // *Process Safety and Environmental Protection*. 2021. Vol. 147. P. 367–384. <https://doi.org/10.1016/j.psep.2020.09.038>

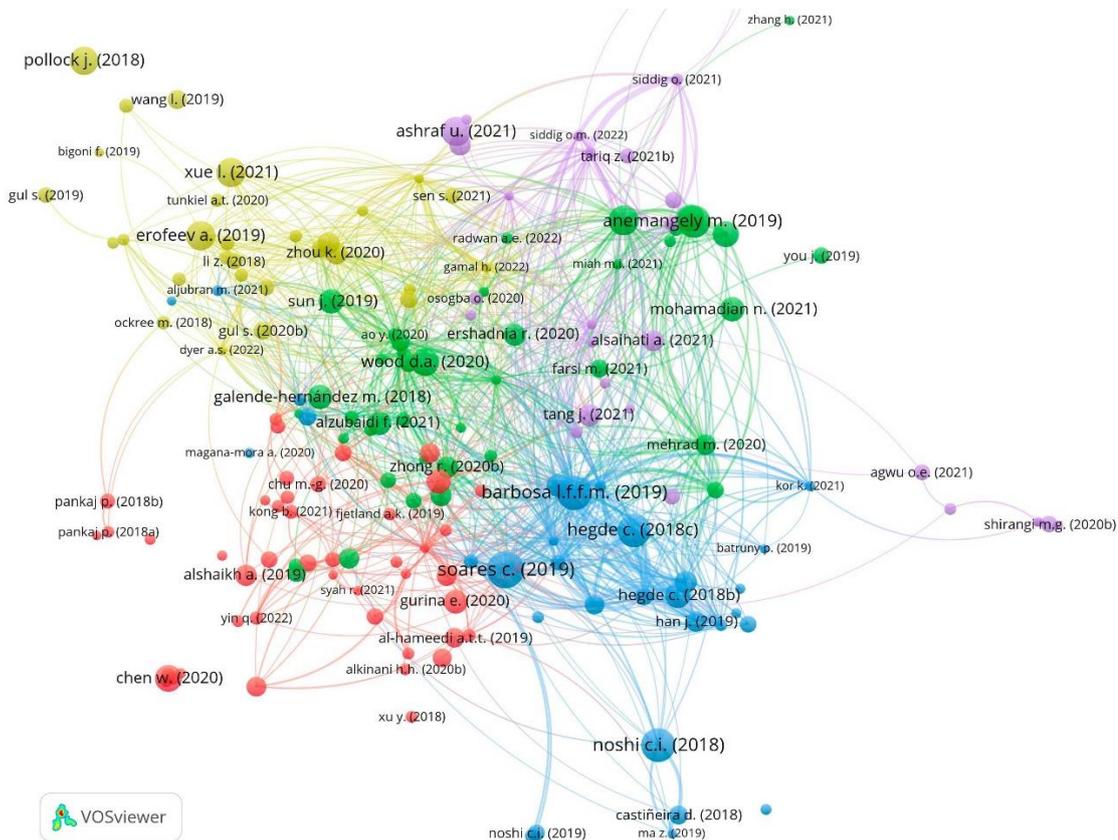
ПРИЛОЖЕНИЕ

Приложение 1

Термин в статье	Перевод
1	2
Anomaly detection	Обнаружение аномалий
Artificial intelligence	Искусственный интеллект
Artificial neural network	Искусственная нейронная сеть
Author Keywords	Авторские ключевые слова
Bibliographic coupling	Библиографическое сочетание / библиографическая связь (возникает, когда два документа ссылаются на общий третий документ; в русском чаще — «сочетание»)
Concept Tag	Тег концепции
Convolutional neural network	Сверточная нейронная сеть
Data Science & Engineering Analytics	Наука о данных и инженерная аналитика
Decision Tree	Дерево решений
Deep learning	Глубокое обучение
Digital and Oil&Gas	Цифровизация и нефтегазовая отрасль
Digital engineering	Цифровая инженерия
Digital technologies	Цифровые технологии
Digital transformation	Цифровая трансформация
Digitalization	Цифровизация
Drilling and completion	Бурение и заканчивание скважин
Emerging research topics	Зарождающиеся темы исследований
Fault detection	Обнаружение неисправностей
Gradient boosting	Градиентный бустинг
Indexed keywords	Индексируемые ключевые слова (выбираются Scopus и стандартизируются по словарям, полученным из тезаурусов, которыми владеет или лицензирует Elsevier)
Industry 4.0	Индустрия 4.0
Industry and Technology	Индустрия и технологии
Information Technology	Информационные технологии
Lattice Boltzmann Method	Метод решеток Больцмана
Logging interpretation	Интерпретация каротажа
Machine learning	Машинное обучение
Micro-CT	Микро-КТ (компьютерная томография)
Oil and gas industry	Нефтегазовая промышленность
Particle swarm optimization	Оптимизация с помощью алгоритма роя частиц

1	2
Petroleum well drilling	Бурение нефтяных скважин
Probabilistic Random Forest Production and Well Operations	Алгоритм Probabilistic Random Forest (PRF) рассматривает признаки и метки как функции распределения вероятностей, а не как детерминированные величины Эксплуатация скважин и добыча
Random forest method	Метод случайного леса
Rate of penetration	Скорость проходки
Representation & Reasoning	Репрезентация и обоснование
Reservoir Description and Dynamics	Описание и динамика резервуаров
Reservoir engineering	Разработка месторождений
Rock physics	Физика горных пород
SPE	Общество инженеров-нефтяников
SPE Disciplines	Дисциплины SPE
Scimago Graphica	Программа построения графики (https://www.graphica.app/)
Scopus	Библиографическая и реферативная база данных рецензируемой научной литературы, созданная в 2004 году академическим издательством Elsevier
Seismic exploration	Сейсморазведка
Support vector machine	Метод опорных векторов
Surface facility engineering	Проектирование наземных сооружений
Unsupervised machine learning	Неконтролируемое машинное обучение
Upstream oil & gas	Разведка и добыча нефти и газа
VOSviewer	Программа для построения и визуализации библиометрических сетей
Voxel-based modeling	Воксельное моделирование (Воксельная модель — это трехмерный растр)
X-ray	Рентгеновское излучение
search.spe.org	Система поиска и анализа информации на сайтах SPE.org, PetroWiki и OnePetro с использованием искусственного интеллекта

**Фрагмент данных, представленных в формате:
 первый автор (год публикации)(цитируемость)(число библиографически связанных
 публикаций в рассматриваемой выборке)**



hazbeh o. (2021)(15)(130); olukoga t.a. (2021)(11)(125); tunkiel a.t. (2021)(5)(114); alali a.m. (2021)(13)(110);
 siddig o. (2021)(8)(109); gamal h. (2021b)(5)(92); kor k. (2021)(6)(81); **osarogiagbon a.u. (2021)(34)(79)**;
 tewari s. (2021)(8)(72); tariq z. (2021a)(5)(72); sabah m. (2021)(25)(69); sun j. (2021)(5)(68);
 miah m.i. (2021)(6)(59); gamal h. (2021a)(5)(59); **xue l. (2021)(47)(57)**; mahmoud a.a. (2021)(13)(56);
 aljubran m. (2021)(7)(56); tang j. (2021)(27)(51); farsi m. (2021)(19)(48); feng r. (2021)(10)(48);
 alzubaidi f. (2021)(31)(47); peaple m.d. (2021)(9)(47); jiang c. (2021)(9)(43); arnø m.l. (2021)(6)(43);
 alsaihati a. (2021)(26)(41); yin q. (2021)(13)(37); kong b. (2021)(12)(32); mohamadian n. (2021)(34)(30);
 islamov s. (2021)(17)(28); shan l. (2021)(14)(28).

Примечание – жирным шрифтом выделены проанализированные публикации.

Identification and analysis of the actual research issue of “Digital transformation of the oil and gas sector”

B.N. Chigarev

Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia

E-mail: bchigarev@ipng.ru

Abstract. This paper is devoted to identifying and analyzing one of the actual research issues of the topic “Digital transformation of the oil and gas sector”. The actual area and methods of research were determined by the materials of the Society of Petroleum Engineers, and for the disclosure of the scientific aspect of the identified actual issue, publications indexed in the bibliographic and abstract database of peer-reviewed scientific literature Scopus were used. This approach, according to the author of the article, reduces the bias in the selection of relevant scientific issues, as it reflects the interests of both the oil and gas sector and scientific community experts. In this paper, the research issues were determined both by the classification categories of the search.spes.org platform, and by the authors’ keywords of the articles indexed in Scopus. A brief review of the most cited articles revealing the topic of such works was carried out for a more detailed disclosure of the research objectives. The results of the analysis show that within the broad topic of digital transformation of the oil and gas sector, a task relevant to both industry (Society of Petroleum Engineers) and academic institutions, can be disclosed using the term “well drilling” and the random forest method.

Keywords: digital transformation, oil and gas sector, actual research issues, bibliometric analysis, Scopus, VOSviewer.

Citation: Chigarev B.N. Identification and analysis of the actual research issue of “Digital transformation of the oil and gas sector” // Actual Problems of Oil and Gas. 2023. Iss. 1(40). P. 67–88. <https://doi.org/10.29222/ipng.2078-5712.2023-40.art6> (In Russ.).

References

1. Altintas C., Altundal O.F., Keskin S., Yildirim R. Machine learning meets with metal organic frameworks for gas storage and separation // Journal of Chemical Information and Modeling. 2021. Vol. 61, No. 5. P. 2131–2146. <https://doi.org/10.1021/acs.jcim.1c00191>
2. Gomollón-Bel F. Ten chemical innovations that will change our world: IUPAC identifies emerging technologies in Chemistry with potential to make our planet more sustainable // Chemistry International. 2019. Vol. 41, No. 2. P. 12–17. <https://doi.org/10.1515/ci-2019-0203>
3. Koroteev D., Tekic Z. Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future // Energy and AI. 2021. Vol. 3. P. 100041. <https://doi.org/10.1016/j.egyai.2020.100041>
4. Kuang L., Liu H., Ren Y. et al. Application and development trend of artificial intelligence in petroleum exploration and development // Petroleum Exploration and Development. 2021. Vol. 48, No. 1. P. 1–14. [https://doi.org/10.1016/S1876-3804\(21\)60001-0](https://doi.org/10.1016/S1876-3804(21)60001-0)
5. Tembely M., AlSumaiti A.M., Alameri W.S. Machine and deep learning for estimating the permeability of complex carbonate rock from X-ray micro-computed tomography // Energy Reports. 2021. Vol. 7. P. 1460–1472. <https://doi.org/10.1016/j.egy.2021.02.065>

6. *van Eck N.J., Waltman L.* Software survey: VOSviewer, a computer program for bibliometric mapping // *Scientometrics*. 2010. Vol. 84, No. 2. P. 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
7. *Hassan-Montero Y., De-Moya-Anegón F., Guerrero-Bote V.P.* SCImago Graphica: a new tool for exploring and visually communicating data // *El Profesional de la información*. 2022. Vol. 31, No. 5. P. e310502. <https://doi.org/10.3145/epi.2022.sep.02>
8. *van Eck N.J., Waltman L.* VOSviewer Manual: Manual for VOSviewer version 1.6.18. Leiden: Center for Science and Technology Studies, Leiden University, 2022. 53 p. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf (Accessed on 03.12.2022).
9. *Zanin M., Güntekin B., Aktürk T.* et al. Telling functional networks apart using ranked network features stability // *Scientific Reports*. 2022. Vol. 12, No. 1. P. 2562. <https://doi.org/10.1038/s41598-022-06497-w>
10. *Ben Aoun M.A., Madarász T.* Applying machine learning to predict the rate of penetration for geothermal drilling located in the Utah FORGE site // *Energies*. 2022. Vol. 15, No. 12. P. 4288. <https://doi.org/10.3390/en15124288>
11. *Breiman L.* Random Forests // *Machine Learning*. 2001. Vol. 45, No. 1. P. 5–32. <https://doi.org/10.1023/A:1010933404324>
12. *Belgiu M., Drăguț L.* Random forest in remote sensing: A review of applications and future directions // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016. Vol. 114. P. 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
13. *Contreras I.C., Khodadadzadeh M., Gloaguen R.* Multi-label classification for drill-core hyperspectral mineral mapping // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020. Vol. XLIII-B3-2020. P. 383–388. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-383-2020>
14. *Gamal H., Alsaihati A., Elkatatny S.* Predicting the rock sonic logs while drilling by random forest and decision tree-based algorithms // *Journal of Energy Resources Technology*. 2022. Vol. 144, No. 4. P. 043203. <https://doi.org/10.1115/1.4051670>
15. *Ao Y., Zhu L., Guo S., Yang Z.* Probabilistic logging lithology characterization with random forest probability estimation // *Computers & Geosciences*. 2020. Vol. 144. P. 104556. <https://doi.org/10.1016/j.cageo.2020.104556>
16. *Reis I., Baron D., Shahaf S.* Probabilistic Random Forest: A machine learning algorithm for noisy datasets // *The Astronomical Journal*. 2019. Vol. 157, No. 1. P. 16. <https://doi.org/10.3847/1538-3881/aaf101>
17. *Xue L., Liu Y., Xiong Y.* et al. A data-driven shale gas production forecasting method based on the multi-objective random forest regression // *Journal of Petroleum Science and Engineering*. 2021. Vol. 196. P. 107801. <https://doi.org/10.1016/j.petrol.2020.107801>
18. *Osarogiagbon A.U., Khan F., Venkatesan R., Gillard P.* Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations // *Process Safety and Environmental Protection*. 2021. Vol. 147. P. 367–384. <https://doi.org/10.1016/j.psep.2020.09.038>