

**ПОИСК МЕСТОРОЖДЕНИЙ-АНАЛОГОВ  
НА ОСНОВЕ КЛАСТЕРИЗАЦИИ БАЙЕСОВСКИХ СЕТЕЙ**

А. К. БЕЗБОРОДОВ\*, И. Ю. ДЕЕВА

*Университет ИТМО, Санкт-Петербург, Россия*  
\* *akbezborodov@itmo.ru*

**Аннотация.** Предложен алгоритм, разработанный для поиска нефтегазовых месторождений-аналогов, основанный на кластеризации байесовских сетей, построенных на параметрах месторождений. С помощью байесовских сетей возможно эффективно представить нефтегазовые месторождения в виде многомерного распределения с учетом сложных взаимосвязей между параметрами. Для каждого из месторождений в производственной базе данных строилась байесовская сеть на выборке из ближайших его соседей, полученных с использованием метрики косинусного расстояния. Кластеризация построенных на выборках месторождений сетей производится путем сравнения метрики расстояния Хэмминга между вытянутыми в одномерный вектор матрицами смежности. Произведен сравнительный анализ разработанного подхода и других методов поиска аналогов на основе методов машинного обучения. Приведены результаты оценки работы алгоритма, подтверждающие, что моделирование и поиск аналогов с помощью байесовских сетей является более комплексным решением задачи. Точность восстановления пропущенных значений для большинства параметров с помощью разработанного алгоритма оказалась выше, чем в существующих классических алгоритмах кластеризации.

**Ключевые слова:** нефтегазовые месторождения, месторождения-аналоги, поиск параметров, кластеризация, геологические параметры, байесовские сети

**Ссылка для цитирования:** Безбородов А. К., Деева И. Ю. Поиск месторождений-аналогов на основе кластеризации байесовских сетей // Изв. вузов. Приборостроение. 2022. Т. 65, № 1. С. 64—72. DOI: 10.17586/0021-3454-2022-65-1-64-72.

**SEARCH FOR ANALOGUE DEPOSITS BASED ON BAYESIAN NETWORKS CLUSTERING**

A. K. Bezborodov\*, I. Yu. Deeva

*ITMO University, St. Petersburg, Russia*  
\* *akbezborodov@itmo.ru*

**Abstract.** An algorithm is developed to search for oil and gas deposits-analogues, based on clustering of Bayesian networks, built on parameters of the known deposits. Using Bayesian networks, it is possible to effectively represent oil and gas fields in the form of multivariate distributions, accounting for the complex relationships between the parameters. For each of the deposits in the database, a Bayesian network was built on a sample of its nearest neighbors obtained using the cosine distance metric. Clustering of the Bayesian networks built on the samples is performed by comparing the Hamming distance metric between adjacency matrices stretched into a one-dimensional vector. The developed approach is compared to other analogue search methods based on machine learning. Results of evaluation of the algorithm performance are presented, confirming that modeling and searching for analogues using Bayesian networks is a more comprehensive solution to the problem. The accuracy of restoring missing values for most parameters using the developed algorithm turned out to be higher than in existing classical clustering algorithms.

**Keywords:** oil and gas deposits, analogue deposits, parameter search, clustering, geological parameters, Bayesian networks

**For citation:** Bezborodov A. K., Deeva I. Yu. Search for analogue deposits based on Bayesian networks clustering. *Journal of Instrument Engineering*. 2022. Vol. 65, N 1. P. 64—72 (in Russian). DOI: 10.17586/0021-3454-2022-65-1-64-72.

**Введение.** В нефтегазовой отрасли достаточно типичными являются сценарии, в которых информация о некоторых параметрах исследуемого месторождения отсутствует или ограничена. Отсутствие данных о ключевых параметрах залежей может отрицательно сказаться

на производственном процессе добычи и разведки полезных ресурсов, что впоследствии может привести к принятию экономически необоснованных решений [1]. В связи с большой стоимостью и продолжительностью разведочных бурений возникает необходимость теоретической оценки недостающих параметров нефтегазовых залежей. Одним из важнейших способов, позволяющим произвести моделирование параметров нефтегазовых месторождений, является оценка на основе месторождений-аналогов [2, 3].

Месторождение можно считать аналогом, если оно совпадает или является схожим с рассматриваемым объектом по некоторым известным геологическим параметрам обеих залежей. Такое сходство позволяет сделать вывод, что и остальные характеристики будут аналогичны параметрам месторождения, в котором они неизвестны, что, в свою очередь, позволяет, например, произвести оценку пропусков в данных, полученных с месторождений-аналогов [3].

Следует отметить, что совокупность геологических параметров месторождения представляет собой многомерное распределение, содержащее как дискретные, так и непрерывные переменные. Определение типа зависимостей между данными имеет важное значение, так как применение классических совместных моделей распределения с использованием принципа корреляции может оказаться невозможным в случае, если зависимости не линейны, что характерно для геологических и геометрических признаков. Дополнительным преимуществом при выборе оптимального подхода является минимальное использование экспертных знаний, а в идеальном случае и полный отказ от них. В связи с этим поиск месторождений-аналогов является достаточно нетривиальной задачей, для которой на данный момент не существует общепринятого решения [4].

Нефтегазовая отрасль в России является одним из наиболее приоритетных направлений по применению современных научных методов анализа данных и технологий машинного обучения. В работе [1] наглядно показано, насколько использование предиктивной аналитики в задаче поиска аналогов месторождений позволяет оптимизировать расходы и спрогнозировать эффективность разработки новых залежей. Отечественными авторами был также разработан ряд методик по поиску аналогов на основе применения экспертных знаний.

Подходы, основанные на привлечении экспертных знаний, являются наиболее простым методом по поиску аналогов с точки зрения времени и трудозатрат. В [4] показан способ поиска аналогов с применением нескольких фильтров — параметров, важнейшим из которых является расстояние до текущей области. Метод поиска месторождений-аналогов с использованием экспертных знаний может быть улучшен путем применения метода аналогий, для которого требуется обширная производственная база данных месторождений [5, 6]. Для более точного определения неизвестных параметров исследуемой залежи предлагается предварительная обработка базы данных с разделением признаков на качественные и количественные. С использованием подобной группировки параметров появляется возможность создания списка месторождений, близких к рассматриваемому в первую очередь по выбранным экспертом параметрам. Основываясь на их сходстве, можно создать кластер потенциальных аналогов, в котором месторождения будут ранжированы в зависимости от сходства ключевых количественных признаков [6].

Подходы с приоритетным использованием экспертных знаний позволяют получить быструю оценку экспертом за короткое время, однако не гарантируют точность результатов и обычно применяются в сжатые сроки в отсутствие более надежного метода оценки неизвестных параметров. Другой недостаток таких методов — отсутствие информации о значимости или информативности пропущенных параметров. Не удается и характеризовать зависимости между параметрами, так как выставление весовых коэффициентов значимости переменных в процессе поиска аналогов экспертом является субъективной оценкой. Исходя из этого требуется рассмотреть более комплексные подходы к поиску аналогов, включающие в себя анализ многомерного распределения с учетом возможных связей между переменными.

В настоящее время активно используются подходы, основанные на использовании технологий машинного обучения. В частности, возможно применение различных алгоритмов кластеризации в целях получения групп аналогов, обладающих схожими свойствами. В геологической отрасли для моделирования различных процессов наиболее часто применяют иерархическую агломеративную кластеризацию и кластеризацию методом  $k$ -средних [7, 8].

В работах [9, 10] используется иерархическая кластеризация применительно к решению задачи по поиску аналогов месторождений. После предобработки исходных данных выполняется снижение их размерности с применением метода главных компонент. Далее выполняется обучение модели, в результате которого исходная база данных разделяется на группы месторождений со схожими свойствами.

Метод  $k$ -средних, с помощью которого возможно выделять кластер с аналогами месторождений, рассматривается в работе [11], где предлагается использовать данный вид кластеризации, предварительно снизив размерность исходной базы данных методом главных компонент. Затем для оценки качества полученных кластеров с аналогами предлагается использовать регрессионные модели.

Однако следует отметить, что указанные подходы имеют некоторые недостатки. Оба вида кластеризации оказываются достаточно чувствительными ко входным данным большой размерности. Снижение размерности входных данных может привести к потерям исходной информации. Другой недостаток — сложность применения таких подходов в задачах, где требуется восстановить один или несколько параметров исследуемого месторождения. Месторождение, часть параметров которого пропущена, не может участвовать в процессе обучения модели, что требует дополнительной обработки при определении его кластера.

Таким образом, предлагаемый алгоритм должен эффективно работать как с непрерывными, так и дискретными параметрами требуемой размерности. Необходимо найти наиболее подходящий способ определения качества выявленных аналогов в кластерах. Следует принимать во внимание интерпретируемость итоговой модели и вовлеченность экспертных знаний.

Исходя из предъявленных требований, был сделан выбор в пользу реализации с применением технологий машинного обучения, таких как байесовские сети [12], которые позволяют при средней вычислительной сложности и незначительном применении экспертных знаний получить хорошую интерпретируемость обученной модели, способной работать с пропусками. Графическая структура байесовской сети позволяет наглядно продемонстрировать возникающие сложные взаимосвязи между исследуемыми параметрами. С помощью формирования условных распределений вероятностей в узлах сети появляется возможность не только оценивать неизвестные переменные, но и выполнять поиск аномальных значений и проводить анализ достоверности и информативности полученных данных.

Байесовские сети эффективно применяются в геологической отрасли для моделирования параметров с целью оценки возникающих неопределенностей [13—15]. В [13] с помощью байесовской сети для оценки расположения нефтегазовых месторождений выполнены прогнозы пространственного распределения ресурсов, таких как скопления углеводородов. В [14] байесовские сети применяются для анализа неопределенностей и построения процесса принятия решений при геологическом анализе месторождений. Сеть обучается с помощью нескольких известных сценариев моделирования бассейновых и нефтяных систем для различных геологических сценариев. В дальнейшем обученная сеть может использоваться для проверки сценариев принятия решений и выполнения анализа ценности информации.

Анализ и тестирование предлагаемого алгоритма производились на производственной базе данных, содержащей 442 месторождения, в которых не было пропущенных параметров. Для итоговой реализации был сделан выбор в пользу 12 параметров, наиболее точно характеризующих внутренний состав нефтегазовых месторождений.

**Байесовские сети.** Байесовская сеть является графической вероятностной моделью, представляющей собой ориентированный ациклический граф. Байесовская сеть позволяет наглядным образом представить сложные взаимосвязи между параметрами путем отображения переменных в вершинах графа и демонстрации с помощью ребер их вероятностных зависимостей по теореме Байеса. Основа данного метода — байесовский подход к определению вероятностей, позволяющий найти вероятность некоторого события при условии, что произошли другие, статистически взаимозависимые с ним события [12].

В случае если две переменные, находящиеся в вершинах графа, являются взаимосвязанными, то параметр, от вероятностного распределения которого зависит значение второй переменной, называется родителем. В свою очередь, зависимый параметр, к которому направлено ребро графа, называется потомком. При обозначении множества вершин, являющихся родителями вершины  $B_i$ , как  $\text{parents}(B_i) = PX_i$ , направленный ациклический граф является байесовской сетью, если каждой из случайных переменных ставится в соответствие одна из вершин графа, а для ребер графа выполняется следующее условие: каждая из переменных в графе условно независима от всех вершин, не являющихся ее родителями. Таким образом, можно представить полное распределение всей байесовской сети путем факторизации всех локальных произведений родителей и их потомков [12]:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i)).$$

Для моделирования многомерного распределения случайных величин с использованием байесовских сетей необходимо выполнить две задачи:

- произвести обучение структуры самой сети;
- произвести обучение условных распределений в узлах сети.

Применительно к задаче кластеризации сетей для получения месторождений-аналогов производить обучение условных распределений параметров в узлах сети не требуется. Достаточно сформировать граф, в котором при анализе выборки месторождений в качестве узлов будут выступать рассматриваемые параметры, а ребра графа, сформированные в результате обучения сети, будут соответствовать установленным взаимосвязям между параметрами.

С помощью разработанного алгоритма выполнялось построение байесовских сетей с K2 score-функцией методом Hill-Climbing [14]. Данный алгоритм поиска оптимальной структуры сети состоит из нескольких последовательных шагов.

*Шаг 1.* Добавление ребра между каждой из пар вершин с изменением направления.

*Шаг 2.* Вычисление score-функции нового графа.

*Шаг 3.* Если полученное значение score-функции оказывается точнее, чем на шагах 1, 2, реализованных итеративно, то данная структура принимается за основную.

*Шаг 4.* Остановка алгоритма, когда значение score-функции перестает изменяться.

Выбранная score-функция имеет следующий вид:

$$\ln P(D : G) = \sum_{i=1}^n \sum_{j=1}^{q_i} \left( \ln \frac{\Gamma \left( \sum_{s=1}^{r_i} \alpha_{i,j,s} \right)}{\Gamma \left( \sum_{s=1}^{r_i} (N_{i,j,s} + \alpha_{i,j,s}) \right)} + \sum_{s=1}^{r_i} \ln \frac{\Gamma(N_{i,j,s} + \alpha_{i,j,s})}{\Gamma(\alpha_{i,j,s})} \right),$$

где  $G$  — структура графа байесовской сети;  $D$  — данные, на которых производится обучение сети;  $\Gamma$  — гамма-распределение;  $\alpha_{i,j,s}$  — параметры распределения Дирихле (для K2  $\alpha_{i,j,s} = 1$ );  $i$  — порядковый номер вершины;  $n$  — количество вершин;  $j$  — порядковый номер вершины-родителя;  $q_i$  — количество возможных значений родителей;  $r_i$  — количество

значений переменной  $i$ ;  $N_{i,j,s}$  — число строк в данных, для которых  $i$ -я вершина  $s$ -ми родителями принимает значение  $k$ .

**Кластеризация байесовских сетей.** В предложенном алгоритме для каждого из месторождений в базе данных определяются 50 его ближайших соседей с использованием метрики косинусного расстояния:

$$D_{\cos} = 1 - \frac{uv}{\|u\|_2 \|v\|_2},$$

где  $u$  и  $v$  соответствуют одномерным численным векторам параметров двух месторождений.

Дополнительные исследования были проведены при формировании ближайших соседей месторождений с помощью метрики расстояния Гауэра [16]:

$$D_{\text{Gower}} = \frac{1}{m} \sum_{f=1}^m ps_{ql}^{(f)};$$

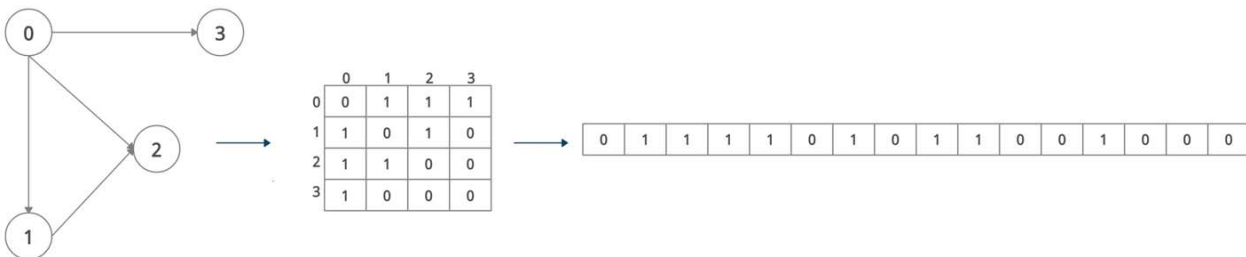
$$ps_{ql}^{(f)} = 1 - \frac{|x_{qf} - x_{lf}|}{R_f};$$

$$R_f = \max f - \min f,$$

где  $m$  — общее число признаков, а  $ps_{ql}$  соответствует частному сходству для конкретного признака  $f$  между двумя наблюдениями  $q$  и  $l$ ; для непрерывных параметров данное сходство пропорционально разности значений  $x_{qf}$  и  $x_{lf}$  в конкретном признаке  $f$ , деленной на общий диапазон признака  $R_f$ ; в случае категориальных параметров сходство равно единице, если две категории совпадают, и равно нулю в противоположном случае.

Использование метрики косинусного расстояния позволило получить наилучшие результаты при исследовании качества сформированных аналогов. На полученных выборках, в состав которых входят исследуемое месторождение и его соседи, строится байесовская сеть с K2 score-функцией методом Hill-Climbing.

Таким образом, формируются  $M$  сетей в виде направленных графов, где  $M$  соответствует числу месторождений в базе данных. Первоначальные этапы обработки полученных результатов обучения для дальнейшей кластеризации данных графовых структур представлены на рисунке.



Согласно рисунку, на первом этапе для каждого месторождения имеется построенный направленный граф, характеризующий взаимосвязи между параметрами. Далее на основе структуры графа для каждой из сетей формируется матрица смежности, которая затем вытягивается в одномерный вектор.

Определение близости построенных сетей выполняется путем сравнения текущего вектора со всеми векторами, соответствующими остальным месторождениям, с помощью метрики расстояния Хэмминга:

$$D_H = \sum_{i=1}^k |x_i \oplus y_i|,$$

где для двух булевых векторов  $x$  и  $y$  для каждого из значений поэлементно вычисляется логическая операция XOR, а затем полученное число единиц суммируется для итогового результата.

С помощью данной метрики появляется возможность определить число позиций, в которых соответствующие символы двух строк отличаются друг от друга. Таким образом, на основе данного сравнения матриц смежности, соответствующих обученным графовым структурам, возможно численно охарактеризовать, насколько построенные на выборках сети близки друг к другу. Определив расстояния Хэмминга от каждого месторождения относительно всех остальных залежей, можно сформировать матрицу близости построенных сетей, строки которой характеризуют, насколько схожи текущие месторождения с остальными.

Для получения итоговых кластеров с аналогами выполняется кластеризация строк полученной матрицы методом  $k$ -средних [17]. Таким образом, удается сформировать группы, в которых месторождения являются схожими, на основе полученных байесовских сетей на выборках месторождений.

Для определения кластеров аналогов всех месторождений в базе данных рассмотренным способом требуется выполнить следующие шаги:

- 1) определение для каждого из месторождений 50 его ближайших соседей с использованием метрик расстояния;
- 2) построение байесовских сетей на полученных подвыборках;
- 3) сохранение сформированных графовых структур в виде матриц смежности, вытянутых в одномерный вектор;
- 4) определение для каждого из векторов расстояния до остальных векторов с помощью метрики расстояния Хэмминга;
- 5) кластеризация строк полученной матрицы расстояний, формирование кластеров.

**Модели для оценки качества аналогов.** Для определения точности произведенной кластеризации применение классических регрессионных моделей для восстановления параметров неприемлемо в связи с их большой чувствительностью к размеру тренировочной выборки. Для данной задачи в целях определения, насколько правильно произведена кластеризация, выполняется следующая оценка точности параметров в рамках кластеров:

- среднее значение параметра по кластеру для непрерывных параметров;
- наиболее часто встречающаяся категория по кластеру для дискретных параметров.

Применительно к данной задаче для каждого параметра каждого из месторождений в кластерах выполнялось заполнение параметра указанными способами в зависимости от типа параметра. Необходимо отметить, что данный способ не характеризует точность восстановления параметров, а является сравнительной проверкой того, насколько правильно выполнена произведенная кластеризация по сравнению с другими методами поиска аналогов.

В качестве метрики для оценки точности произведенного заполнения для непрерывных параметров используется среднеквадратическая ошибка (СКО) [18], а для дискретных параметров — средний показатель точности (Mean Accuracy Score) в рамках кластера (доля правильно заполненных категорий по отношению к общему числу месторождений в кластере).

**Результаты.** Полученные результаты предсказаний пропусков для дискретных параметрах представлены в табл. 1, а для непрерывных параметров — в табл. 2. Сравнение производилось с аналогичным заполнением параметров в рамках всей базы данных без кластеров, а также с использованием модели, представленной в [10], где для формирования кластеров применены метод главных компонент и кластеризация методом  $k$ -средних.

Таблица 1

Параметр	Точность (по всей базе данных)	Точность (СКО + кластеризация k-средних — 5 кластеров)	Точность (кластеризация байесовских сетей — 5 кластеров)	
			$D_{\cos}$	$D_{Gower}$
Литология коллектора	0,56	0,56	<b>0,58</b>	0,56
Тектонический режим	0,48	0,5	<b>0,51</b>	0,5
Тип углеводородов	0,76	<b>0,76</b>	<b>0,76</b>	0,76
Структурная принадлежность	0,28	<b>0,31</b>	0,31	<b>0,34</b>
Геосистема образования резервуара	0,27	0,28	<b>0,35</b>	0,29
Массив отложений, заполняющих бассейн	0,35	0,35	<b>0,4</b>	0,32
Тип ловушки	0,47	0,47	<b>0,48</b>	0,47

Таблица 2

Параметр	Точность (по всей базе данных)	Точность (СКО + кластеризация k-средних — 5 кластеров)	Точность (кластеризация байесовских сетей — 5 кластеров)	
			$D_{\cos}$	$D_{Gower}$
Пористость	7,64	6,71	<b>6,7</b>	7,61
Глубина кровли резервуара	1115,7	<b>512,22</b>	969,63	1112,26
Общая толщина резервуара	448,43	446,87	<b>440,54</b>	441,57
Эффективная толщина коллектора	84,17	83,61	<b>83,29</b>	82,94
Проницаемость	943,16	<b>366,21</b>	881,7	941,26

Анализ табл. 1 показывает, что для всех параметров за исключением структурной принадлежности предложенный алгоритм на основе кластеризации байесовских сетей характеризуется наибольшей точностью восстановления параметров в рамках кластеров по сравнению с другими методами. Согласно табл. 2, предложенный подход показал наилучшие результаты для всех параметров, кроме глубины кровли резервуара и проницаемости. Также следует отметить, что метрика косинусного расстояния является наиболее предпочтительной метрикой близости, которая используется на этапе формирования выборок ближайших месторождений для исследуемой залежи.

**Заключение.** Предложен алгоритм поиска аналогов месторождений на основе кластеризации байесовских сетей. Результаты предсказаний пропущенных параметров в рамках кластеров оказались точнее предсказаний, выполненных на основе всей базы данных, что говорит об эффективности проведенной кластеризации. Результаты предсказаний категориальных и большинства непрерывных параметров оказались наилучшими для кластеризации байесовских сетей предложенным методом с использованием метрики косинусного расстояния. Результаты, полученные с наименьшей точностью для параметров глубины кровли резервуара и проницаемости, могут быть связаны с достаточно сложным распределением значений

этих параметров в базе данных. Дальнейшее усовершенствование предложенного алгоритма может быть направлено на исследование других методов кластеризации графовых структур.

## СПИСОК ЛИТЕРАТУРЫ

1. *Popova O.* Analogy in the World of Geological Uncertainties, or How Reservoir Analogs May Refine Your Probabilistic Geomodel // SPE Annual Caspian Techn. Conf. and Exhibition, Astana, Kazakhstan, 31 Oct. — 1 Nov. 2018. P. 1—13. DOI: 10.2118/192551-MS.
2. *Neves L., Ferreira C., Vesely F.* Facies, petrophysical and geochemical properties of gravity-flow deposits in reservoir analogs from the Itararé Group (late Carboniferous), Paraná Basin, Brazil // Marine and Petroleum Geology. 2019. Vol. 110. P. 717—736. DOI: 10.1016/j.marpetgeo.2019.07.038.
3. *Olivo M.* et al. Characterizing scales of sedimentary heterogeneity in a fluvial—estuarine reservoir analog: An example from the Mulichinco Formation, Neuquén Basin, Argentina // Marine and Petroleum Geology. 2020. Vol. 121. DOI: 10.1016/j.marpetgeo.2020.104604.
4. *Voskresenskiy A.* et al. Variations in ranked list of reservoir analogs as an effect of search preferences // Geosciences: Converting Knowledge into Resources, Saint-Petersburg, Russia, 6 — 9 Apr. 2018. P. 1—5. (Conf. Proc.) DOI: 10.3997/2214-4609.202053163.
5. *Вологин И. С., Исламов Р. Р., Нигматуллин Ф. Н., Харисова А. В., Лознюк О. А.* Методика выбора объекта-аналога для нефтегазовой залежи по геолого-физическим характеристикам // Нефтяное хозяйство. 2019. № 12. С. 124—127.
6. *Поднебесных А. В., Хафизов А. Р.* Методика экспресс-оценки выбора объекта-аналога для залежей углеводородного сырья на основе их геологических признаков // Проблемы сбора, подготовки и транспорта нефти и нефтепродуктов. 2019. № 3. С. 9—18.
7. *Dumont M., Reninger P., Pryet A., Martelet G., Aunay B., Join J.* Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies // J. of Applied Geophysics. 2018. Vol. 157. P. 1—9. DOI: 10.1016/j.jappgeo.2018.06.020.
8. *Ren Z., Sun L., Zhai Q.* Improved k-means and spectral matching for hyperspectral mineral mapping // Intern. Journal of Applied Earth Observation and Geoinformation. 2020. Vol. 91. P. 1—12. DOI: 10.1016/j.jag.2020.102154.
9. *Rodríguez H. M.* et al. New Approach to Identify Analogue Reservoirs // SPE Economics and Management, New Orleans, Louisiana, USA, 30 Sept. — 2 Oct. 2013. P. 1—17. DOI: 10.2118/166449-MS.
10. *Da Gama e Silva R.* et al. Sensitivity analysis in a machine learning methodology for reservoir analogues // Rio Oil & Gas Expo and Conf., Rio de Janeiro, Brazil, 24 — 27 Sept. 2018. P. 1—9.
11. *Безбородов А. К., Деева И. Ю.* Алгоритм поиска месторождений-аналогов на основе методов машинного обучения // Альманах научных работ молодых ученых Университета ИТМО. СПб: НИУ ИТМО, 2021. С. 1—6.
12. *Jensen F. V.* Bayesian Networks and Decision Graphs. Springer, 2007. 448 p. DOI: 10.1007/978-0-387-68282-2.
13. *Martinelli G., Eidsvik J., Sinding-Larsen R., Rekstad S., Mukerji T.* Building Bayesian networks from basin-modelling scenarios for improved geological decision making // Petroleum Geoscience. 2013. Vol. 19, N 3. P. 289—304. DOI: 10.1144/petgeo2012-057.
14. *Scutari M.* et al. Who Learns Better Bayesian Network Structures: Accuracy and Speed of Structure Learning Algorithms // Intern. Journal of Approximate Reasoning. 2019. Vol. 115. P. 235—253. DOI: 10.1016/j.ijar.2019.10.003.
15. *Zhang L., Li D., Tang X., Cao Z., Phoon K.* Bayesian model comparison and characterization of bivariate distribution for shear strength parameters of soil // Computers and Geotechnics. 2018. Vol. 95. P. 110—118. DOI: 10.1016/j.compgeo.2017.10.003.
16. *Gower J.* A General Coefficient of Similarity and Some of Its Properties // Biometrics. 1971. Vol. 27, N 4. P. 857—871.
17. *Steinhaus H.* Sur la division des corps matériels en parties // Bull. Acad. Polon. Sci. 1956. P. 801—804.
18. *Hyndman R. J., Koehler A. B.* Another look at measures of forecast accuracy // Intern. Journal of Forecasting. 2006. Vol. 22, N 4. P. 679—688. DOI: 10.1016/j.ijforecast.2006.03.001.



- Андрей Константинович Безбородов** — студент; Университет ИТМО, факультет цифровых трансформаций; E-mail: akbezborodov@itmo.ru
- Ирина Юрьевна Деева** — аспирант; Университет ИТМО, факультет цифровых трансформаций; E-mail: ideeva@itmo.ru

Поступила в редакцию 19.07.2021; одобрена после рецензирования 22.09.2021; принята к публикации 02.12.2021.

#### REFERENCES

1. Popova O. *SPE Annual Caspian Technical Conference and Exhibition*, Astana, Kazakhstan, 31 Oct.–1 Nov. 2018, pp. 1–13, DOI: 10.2118/192551-MS.
2. Neves L., Ferreira C., Vesely F. *Marine and Petroleum Geology*, 2019, vol. 110, pp. 717–736, DOI: 10.1016/j.marpetgeo.2019.07.038.
3. Olivo M. et al. *Marine and Petroleum Geology*, 2020, vol. 121, DOI: 10.1016/j.marpetgeo.2020.104604.
4. Voskresenskiy A. et al. *Geosciences: Converting Knowledge into Resources*, Saint-Petersburg, Russia, 6–9 Apr. 2018, pp. 1–5, DOI: 10.3997/2214-4609.202053163.
5. Vologin I.S., Islamov R.R., Nigmatullin F.N., Kharisova A.V., Loznyuk O.A. *Oil Industry*, 2019, no. 12, pp. 124–127. (in Russ.)
6. Podnebesnykh A.V., Khafizov A.R. *Problemy sbora, podgotovki i transporta nefti i nefteproduktov*, 2019, no. 3, pp. 9–18. (in Russ.)
7. Dumont M., Reninger P., Pryet A., Martelet G., Aunay B., Join J. *Journal of Applied Geophysics*, 2018, vol. 157, pp. 1–9, DOI: 10.1016/j.jappgeo.2018.06.020.
8. Ren Z., Sun L., Zhai Q. *International Journal of Applied Earth Observation and Geoinformation*, 2020, vol. 91, pp. 1–12, DOI: 10.1016/j.jag.2020.102154.
9. Rodríguez M.H. et al. *SPE Economics and Management*, New Orleans, Louisiana, USA, 30 Sep.–2 Oct. 2013, pp. 1–17, DOI: 10.2118/166449-MS.
10. Da Gama e Silva R. et al. *Rio Oil & Gas Expo and Conference*, Rio de Janeiro, Brazil, 24–27 Sep. 2018, pp. 1–9.
11. Bezborodov A.K., Deeva I.Yu. *Almanac of scientific works of young scientists of ITMO University*, 2021, pp. 1–6. (in Russ.)
12. Jensen Finn V. *Bayesian Networks and Decision Graphs*, Springer, 2001, 448 p., DOI: 10.1007/978-0-387-68282-2.
13. Martinelli G., Eidsvik J., Sinding-Larsen R., Rekstad S., Mukerji T. *Petroleum Geoscience*, 2013, no. 3(19), pp. 289–304, DOI: 10.1144/petgeo2012-057.
14. Scutari M. et al. *International Journal of Approximate Reasoning*, 2019, vol. 115, pp. 235–253, DOI: 10.1016/j.ijar.2019.10.003.
15. Zhang L., Li D., Tang X., Cao Z., Phoon K. *Computers and Geotechnics*, 2018, vol. 95, pp. 110–118, DOI: 10.1016/j.compgeo.2017.10.003.
16. Gower J. *Biometrics*, 1971, no. 4(27), pp. 857–871.
17. Steinhaus H. *Bull. Acad. Polon. Sci.*, 1956, pp. 801–804.
18. Hyndman R.J., Koehler A.B. *International Journal of Forecasting*, 2006, no. 4(22), pp. 679–688, DOI: 10.1016/j.ijforecast.2006.03.001.

#### Data on authors

- Andrey K. Bezborodov** — Student; ITMO University, Faculty of Digital Transformation; E-mail: akbezborodov@itmo.ru
- Irina Yu. Deeva** — Post-Graduate Student; ITMO University, Faculty of Digital Transformation; E-mail: ideeva@itmo.ru

Received 19.07.2021; approved after reviewing 22.09.2021; accepted for publication 02.12.2021.