ISSN 2712-8008
Том / Volume 25 №2 2025
Домашняя страница журнала: http://vestnik.pstu.ru/geo/
НЕДРОПОЛЬЗОВАНИЕ

УДК 622+519.688: 550.3: 553.9
Статья / Article
© ПНИПУ / PNRPU, 2025

Автоматизация межскважинной корреляции с использованием алгоритма динамической трансформации временной шкалы: опыт применения

И.С. Тунев, В.А. Барях

ООО «ЛУКОЙЛ-Инжиниринг» (Российская Федерация, 614015, г. Пермь, ул. Пермская, 3а)

Automation of interwell correlation using dynamic time scale transformation algorithm: application experience

Ivan S. Tunev, Vadim A. Bariakh

LUKOIL-Engineering LLC (3a Permskaya st., Perm, 614015, Russian Federation)

Получена / Received: 24.10.2024. Принята / Accepted: 05.12.2024. Опубликована / Published: 01.04.2025

Ключевые слова:
межскважинная корреляция, автоматическая корреляция скважин, автокорреляция, ГИС, каротаж, геофизика, алгоритм динамической трансформации временной шкалы, алгоритм, программирование, кластеризация, анализ данных, статистика, обработка данных.

Представлен подход к автоматизации межскважинной корреляции, основанный на алгоритме динамической трансформации временной шкалы Dynamic Time Warping. Данный метод является эффективным инструментом для анализа и сравнения временных рядов (в данном случае каротажных кривых). Описаны ключевые этапы реализации данного подхода, начиная с предварительной подготовки данных: выбор наиболее информативных кривых геофизических исследований скважин, обработка пропусков, сглаживание и стандартизация данных. Затем применяются методы кластеризации (например, k-means с использованием PCA) для определения эталонных скважин, которые охватывают все возможные варианты поведения кривых геофизических исследований скважин из имеющихся данных. На следующем этапе используется алгоритм Dynamic Time Warping для вычисления сходства между эталонными и целевыми скважинами, что позволяет найти оптимальное соответствие между ними. Наконец, выполняется прогнозирование положения пластопересечений в целевых скважинах на основе экспертных отметок в эталонных скважинах.

Опыт применения показал высокую согласованность результатов алгоритма с экспертными оценками в 85 % случаев, подтверждая его эффективность. Отмечены расхождения, которые в ряде случаев были вызваны субъективными факторами при экспертной интерпретации, в то время как алгоритм Dynamic Time Warping продемонстрировал эффективность в правильном распознавании заданных ему шаблонов поведения кривых геофизических исследований скважин.

Предложенный подход не только повышает точность и объективность интерпретации данных, но и служит инструментом для выявления и коррекции субъективных ошибок, связанных с человеческим фактором. Это особенно важно при работе с большими объемами данных, где традиционные методы анализа становятся крайне трудоемкими и уязвимыми для ошибок.

Таким образом, использование Dynamic Time Warping открывает новые возможности для автоматизации процессов анализа и корреляции данных в геофизических исследованиях, способствуя повышению эффективности работы специалистов в данной области.

Keywords:
interwell correlation, automatic well correlation, autocorrelation, geophysical well logging, logging, geophysics, dynamic time scale transformation algorithm, algorithm, programming, clustering, data analysis, statistics, data processing.

The article presents an approach to automation of interwell correlation based on the Dynamic Time Warping algorithm for dynamic time scale transformation. This method is an effective tool for analyzing and comparing time series (in this case, well logging curves). The key stages of implementing this approach are described, starting with preliminary data preparation: selecting the most informative well logging curves, processing gaps, smoothing and standardizing the data. Then, clustering methods (for example, k-means using PCA) are applied to determine reference wells that cover all possible behavior options for well logging curves from the available data. At the next stage, the Dynamic Time Warping algorithm is used to calculate the similarity between reference and target wells, which allows finding the optimal match between them. Finally, the position of formation intersections in target wells is predicted based on expert marks in reference wells.

Experience shown high consistency of the algorithm results with expert estimates in 85% of cases, confirming its effectiveness. Discrepancies were noted, which in some cases were caused by subjective factors in expert interpretation, while the Dynamic Time Warping algorithm demonstrated effectiveness in correctly recognizing the assigned behavior patterns of well logging curves.

The proposed approach based on Dynamic Time Warping not only improves the accuracy and objectivity of data interpretation, but also serves as a tool for identifying and correcting subjective errors associated with the human factor. This is especially important when working with large volumes of data, where traditional analysis methods become extremely labor-intensive and vulnerable to errors.

Thus, the use of Dynamic Time Warping opens up new opportunities for automating the processes of data analysis and correlation in geophysical research, contributing to the increased efficiency of specialists in this field.

Тунев Иван Сергеевич (ORCID: 0009-0003-5075-242X) – геолог 1-й категории отдела мониторинга геологического строения (тел.: +007 (952) 659 16 39, e-mail: Ivan.Tunev@lukoil.com). Контактное лицо для переписки.
Барях Вадим Александрович (ORCID: 0009-0006-6453-5313) – начальник отдела мониторинга геологического строения (тел.: +007 (902) 794 48 72, e-mail: Vadim.Baryakh@lukoil.com).

Ivan S. Tunev (ORCID: 0009-0003-5075-242X) – 1st Category Geologist at Department of Geological Structure Monitoring (tel.: +007 (952) 659 16 39, e-mail: Ivan.Tunev@lukoil.com). The contact person for correspondence.
Vadim A. Baryakh (Author ID in Scopus: 57205619691, ORCID: 0009-0006-6453-5313) – Head of Department of Geological Structure Monitoring (tel.: +007 (902) 794 48 72, e-mail: Vadim.Baryakh@lukoil.com).

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:
Тунев, И.С. Автоматизация межскважинной корреляции с использованием алгоритма динамической трансформации временной шкалы: опыт применения / И.С. Тунев, В.А. Барях // Недропользование. – 2025. – Т.25, №2. – С. 66–74. DOI: 10.15593/2712-8008/2025.2.2

Please cite this article in English as:
Tunev I.S., Bariakh V.A. Automation of interwell correlation using dynamic time scale transformation algorithm: application experience. *Perm Journal of Petroleum and Mining Engineering*, 2025, vol.25, no.2, pp. 66-74. DOI: 10.15593/2712-8008/2025.2.2

Введение

Объем данных стремительно растет с развитием современных технологий, ставя новые вызовы перед исследователями и специалистами в различных отраслях, в том числе и в нефтяной промышленности. Традиционные методы анализа данных, особенно при работе с большими объемами информации, такой как данные геофизических исследований скважин (ГИС), часто требуют значительных временных затрат. Проведение корреляции между большим количеством скважин вручную становится крайне трудоемкой и сложной задачей.

Эксперты-геологи должны сравнивать десятки и сотни кривых друг с другом, что отнимает много времени и может приводить к субъективным выводам, что обуславливает необходимость внедрения автоматизированных аналитических процессов для повышения эффективности и точности принимаемых решений.

Представлен подход к автоматизации межскважинной корреляции, основанный на алгоритме динамической трансформации временной шкалы Dynamic Time Warping (DTW) и реализованный с использованием языка программирования Python [1–14]. В рамках работы выделены следующие основные этапы:

1) проведение подготовки и предобработки скважинных данных, а именно выбор наиболее репрезентативных кривых ГИС, обработка пропусков, сглаживание и стандартизация;

2) использование методов кластеризации (K-means, с применением PCA) для определения минимальной выборки эталонных скважин, которая будет содержать все возможные варианты и особенности кривых ГИС из имеющихся данных;

3) реализация алгоритма DTW для вычисления сходства между эталонными и целевыми скважинами при помощи матрицы расстояний. В процессе выполнения алгоритма каждая эталонная скважина сравнивается с каждой целевой, и впоследствии определяется оптимальное соответствие на основе минимального расстояния DTW;

4) определение отметок пластопересечений в целевых скважинах на основе экспертных (заранее известных) отметок в эталонных скважинах на основе найденного оптимального расстояния DTW между скважинами.

Подготовка исходной информации

В процессе анализа данных скважин, особенно в контексте геофизических исследований, первостепенным этапом является подготовка и предобработка данных. Этот этап включает в себя несколько ключевых шагов, которые обеспечивают качество и надежность последующего анализа.

Сначала выбираются наиболее информативные кривые ГИС, по которым однозначно можно отслеживать изменения свойств пластов. Кривые, которые не несут существенной информации или являются избыточными, исключаются из дальнейшего анализа, что позволяет сфокусироваться на наиболее значимых данных. Для апробации алгоритма были выбраны среднедевонские терригенные отложения (D_2), для отслеживания свойств которых будет достаточно использования нейтронного гамма-каротажа (NGK) и гамма-каротажа (GK) [15–18].

Следующим шагом является обработка пропусков в данных. Отсутствующие значения могут значительно повлиять на результаты анализа, поэтому их обработка является крайне важной. Пропуски могут быть заполнены средними значениями, медианами, интерполированы на основе соседних данных или удалены из анализа [19–21].

Для уменьшения влияния шума кривые ГИС необходимо сгладить. Одним из распространенных методов сглаживания является метод скользящего среднего [22, 23]. Формула для вычисления скользящего среднего выглядит следующим образом:

$$MA_t = \frac{1}{n}\sum_{i=0}^{n-1} y_{t-i}, \quad (1)$$

где MA_t – значение скользящего среднего в момент времени t; y_{t-i} – значение исходного ряда в момент времени $t - i$; n – ширина окна сглаживания.

Этот метод помогает снизить случайные колебания и выявить общие тенденции, улучшая интерпретацию данных и повышая точность анализа.

Выбор ширины окна сглаживания (n) является важным шагом. Слишком маленькое окно может не устранить достаточно шума, в то время как слишком большое окно может привести к чрезмерному сглаживанию и потере важных деталей. Выбор ширины окна сглаживания для кривых ГИС требует компромисса между устранением шума и сохранением полезной информации.

На месторождении, используемом для апробации, геологический разрез характеризуется переслаиванием мелких пропластков, которые могут иметь важное значение для корреляции. Учитывая шаг дискретизации данных в 0,1 м, для сглаживания было выбрано значение $n = 10$, что соответствует окну сглаживания в 1 м. Это позволяет уменьшить влияние шума и высокочастотных колебаний на данные, сохраняя при этом достаточную детальность для выявления мелких особенностей разреза.

Для облегчения сравнения между различными кривыми ГИС данные необходимо стандартизировать. Для стандартизации используется метод z-оценки (z-score normalization), который преобразует данные таким образом, что они имеют среднее значение 0 и стандартное отклонение 1 [24, 25]. Формула для z-оценки выглядит следующим образом:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (2)$$

где z_i – стандартизированное значение для точки i; x_i – исходное значение для точки i; μ – среднее значение исходных данных; σ – стандартное отклонение исходных данных.

После стандартизации все кривые ГИС будут иметь одинаковые статистические характеристики, что облегчает их сравнение и анализ.

В результате выполненных шагов предобработки данных ГИС были получены сглаженные и стандартизированные данные, что обеспечивает улучшение интерпретации и повышение точности анализа, а также облегчает сравнение между различными кривыми ГИС (рис. 1).

Определение минимальной выборки эталонных скважин

В задачах, связанных с построением моделей прогнозирования характеристик пластов по результатам исследований скважин, важно создать эталонную выборку. Для получения наиболее качественной и

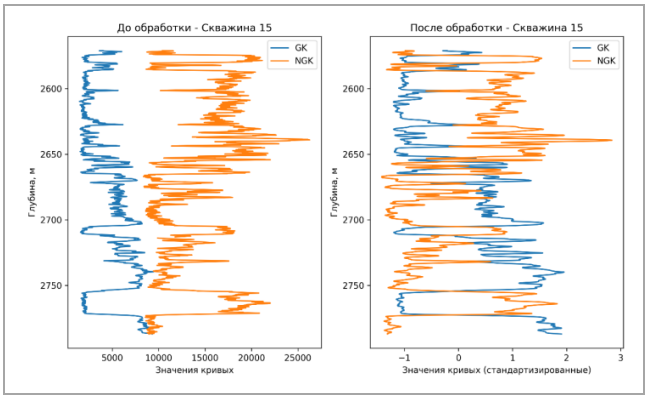


Рис. 1. Кривые ГИС до и после обработки

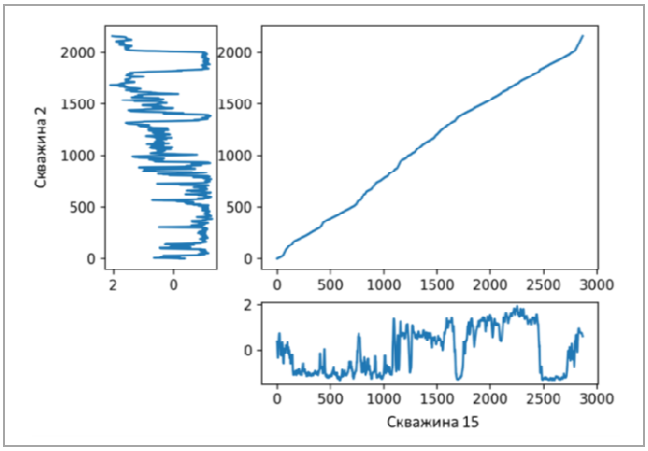


Рис. 2. Матрица расстояний между парой скважин на примере гамма-каротажа

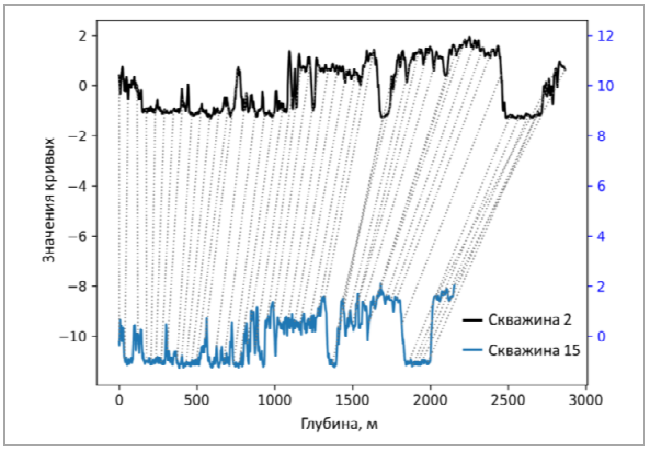


Рис. 3. График соответствия

репрезентативной эталонной выборки целесообразно использовать методы кластеризации. Кластеризация позволяет разбить всю совокупность имеющихся скважин на однородные группы (кластеры) по набору характеризующих их признаков. На основе этого формируется минимальный набор эталонных скважин, охватывающий все выделенные кластеры.

В данном исследовании в качестве ключевого параметра кластеризации предлагается использовать общее расстояние DTW [26–37] между двумя временными рядами, которые в нашем случае представлены кривыми ГИС. Общее расстояние DTW служит мерой сходства между временными рядами.

Основная идея DTW заключается в построении матрицы расстояний между всеми возможными парами точек двух временных рядов. Затем находится оптимальный путь через эту матрицу, который минимизирует суммарное расстояние между соответствующими точками (рис. 2).

В общем случае матрица расстояний DTW вычисляется следующим образом:

$$\mathrm{DTW}(X,Y) = \min_{\gamma}\left(\sum_{(i,j)\in\gamma} d(x_i, y_j)\right), \qquad (3)$$

где $\mathrm{DTW}(X,Y)$ – общее расстояние между двумя временными рядами X, Y; i, j – индексы которые указывают на позицию элемента во временных рядах X и Y соответственно; γ – множество всех возможных путей, представляющих собой последовательность пар индексов (i,j); $d(x_i, y_j)$ – функция расстояния между элементами x_i, временного ряда X и элементами y_j, временного ряда Y; $\min_{\gamma}$ – операция нахождения минимального значения среди всех возможных путей γ.

Более наглядное представление сопоставления двух временных рядов представлено на графике соответствия кривых (рис. 3), на котором показан оптимальный путь DTW в виде линии, соединяющей соответствующие точки на кривых ГИС.

При использовании в анализе нескольких временных рядов для каждой скважины, в нашем случае нейтронный гамма-каротаж и гамма-каротаж, матрица расстояний представляет собой сумму расстояний каждой пары кривых ГИС.

Рассмотрим работу алгоритма на примере двух скважин – А и В. Для каждой скважины формируется матрица данных w, где каждая строка представляет собой шаг дискретизации замера (глубину), а каждый столбец – значение соответствующей кривой ГИС:

$$w_{\mathrm{A}} = \begin{bmatrix} \mathrm{NGK}_{\mathrm{A}1} & \mathrm{GK}_{\mathrm{A}1} \\ \mathrm{NGK}_{\mathrm{A}2} & \mathrm{GK}_{\mathrm{A}2} \\ \vdots & \vdots \\ \mathrm{NGK}_{\mathrm{A}n} & \mathrm{GK}_{\mathrm{A}n} \end{bmatrix} \quad w_{\mathrm{B}} = \begin{bmatrix} \mathrm{NGK}_{\mathrm{B}1} & \mathrm{GK}_{\mathrm{B}1} \\ \mathrm{NGK}_{\mathrm{B}2} & \mathrm{GK}_{\mathrm{B}2} \\ \vdots & \vdots \\ \mathrm{NGK}_{\mathrm{B}m} & \mathrm{GK}_{\mathrm{B}m} \end{bmatrix}$$

Далее создается матрица расстояний D_{AB} размером $(n+1) \times (m+1)$, где n и m – количество значений в матрицах w_A и w_B соответственно:

Каждый элемент (i,j) в матрице D_{AB} (кроме первого столбца и первой строки) вычисляется с использованием рекуррентного соотношения, которое включает в себя локальное расстояние между точками и накопленное расстояние от предыдущих шагов:

$$D_{\mathrm{AB}}(i,j) = d\left(\mathrm{NGK}_{\mathrm{A}}^{i}, \mathrm{NGK}_{\mathrm{B}}^{j}\right) + d\left(\mathrm{GK}_{\mathrm{A}}^{i}, \mathrm{GK}_{\mathrm{B}}^{j}\right) +$$

$$+ \min\left(\sum_{(p,q)\in\Gamma} w_{(p,q)} \cdot D_{\mathrm{AB}}\,(i-\mathrm{p}, j-q)\right). \qquad (4)$$

Рассмотрим подробнее составные части формулы:

1) $D_{\mathrm{AB}}\,(i,j)$ – элемент матрицы расстояний D_{AB}, который представляет собой накопленное расстояние между двумя скважинами А и В на позициях i и j соответственно;

2) $d\,(\mathrm{NGK}_A^{\mathrm{i}}, \mathrm{NGK}_B^{\mathrm{j}})$ – локальное расстояние между значениями параметра NGK в скважине А на позиции i и в скважине В на позиции j;

3) $d\,(\mathrm{GK}_{\mathrm{A}}^{i}, \mathrm{GK}_{\mathrm{B}}^{j})$ – локальное расстояние между значениями параметра GK в скважине А на позиции i и в скважине В на позиции j;

4) $\min\left(\sum_{(\mathrm{p},q)\in\Gamma} w_{(\mathrm{p},q)} \cdot D_{\mathrm{AB}}\,(i-\mathrm{p}, j-q)\right)$ – минимум из суммы взвешенных накопленных расстояний для всех допустимых переходов (p, q) из множества Γ, где Γ –

множество допустимых переходов в шаблоне шага; (p, q) – пара индексов, определяющая конкретный переход в шаблоне шага; $w_{(\mathrm{p},q)}$ – вес перехода (p, q) в шаблоне шага; $D_{\mathrm{AB}}(i-\mathrm{p}, j-q)$ – значения из предыдущих шагов в матрице расстояний D_{AB}.

Шаблон шага (step pattern) определяет, как алгоритм перемещается по матрице расстояний. Чтобы найти оптимальное выравнивание между двумя рядами, он задает допустимые переходы между элементами временных рядов и их веса. Шаблоны могут быть определены индивидуально в зависимости от конкретных задач и особенностей данных. Для решения задачи сопоставления кривых ГИС применялся шаблон Рабинера – Джуанга [38]. Хотя шаблон является стандартным, его можно модифицировать и адаптировать под конкретные задачи и особенности данных. В Python данный шаблон задается функцией rabinerJuangStepPattern (type, slope.weighting), где первый аргумент (type) – это тип шаблона шага, который определяет допустимые переходы между элементами временных рядов и их веса. Второй аргумент (slope.weighting) – это параметр, который определяет, как будут взвешиваться переходы. Классификация шаблонов шага Рабинера – Джуанга предусматривает семь основных шаблонов, а каждый шаблон имеет четыре подтипа веса. Таким образом, у нас в запасе имеется 28 способов шаблонов перехода, однако не каждая комбинация имеет смысл.

После того как матрица расстояний вычислена, находится оптимальный путь выравнивания. Этот путь представляет собой последовательность пар индексов (i, j), которая минимизирует общее накопленное расстояние между двумя временными рядами:

$$(i-\mathrm{p}, j-q) = \arg\min_{(\mathrm{p},q)\in\Gamma}\left(w_{(\mathrm{p},q)} \cdot D_{\mathrm{AB}}(i-\mathrm{p}, j-q)\right). \quad (5)$$

Таким образом, нахождение оптимального пути – это процесс обратного прохода по матрице расстояний, где на каждом шаге выбирается следующий элемент, который минимизирует сумму расстояний с учетом весов и предыдущих шагов. Этот процесс продолжается до тех пор, пока не будет достигнут начальный элемент матрицы.

Финальным шагом является нахождение общего расстояния DTW, которое представляет собой накопленное расстояние между скважинами A и B. Это значение находится в последнем элементе матрицы расстояний D_{AB}, то есть в $D_{\mathrm{AB}}(n, m)$, где n, m – длины временных рядов в скважинах A и B соответственно.

Формула для вычисления общего расстояния в DTW повторяет формулу для вычисления элементов матрицы, за исключением того, что в данном случае берется конечный элемент матрицы:

$$D_{\mathrm{AB}}(n, m) = d\left(N\mathrm{GK}_{\mathrm{A}}^{n}, \mathrm{NGK}_{\mathrm{B}}^{m}\right) + d\left(\mathrm{GK}_{\mathrm{A}}^{n}, \mathrm{GK}_{\mathrm{B}}^{m}\right) +$$

$$+\min\left(\sum_{(\mathrm{p},q)\in\Gamma} w_{(\mathrm{p},q)} \cdot D_{\mathrm{AB}}(n-\mathrm{p}, m-q)\right), \quad (6)$$

где $D_{\mathrm{AB}}(n, m)$ – общее накопленное расстояние между временными рядами скважин A и B.

Матрица расстояний при использовании нескольких временных рядов представляет собой квадратную матрицу, где каждый элемент содержит меру расстояния между соответствующими парами временных рядов (рис. 4).

Шаблон шага, используемый в данной работе, задавался функцией rabinerJuangStepPattern (4, «c») (рис. 5). Тип 4 ориентирован на выявление аномалий, подтип «c» включает весовые функции, которые могут учитывать глобальные тенденции и аномалии, что в совокупности позволяет проводить анализ аномалий с учетом глобальных паттернов и выявлять сложные взаимосвязи, что крайне важно в контексте межскважинной корреляции.

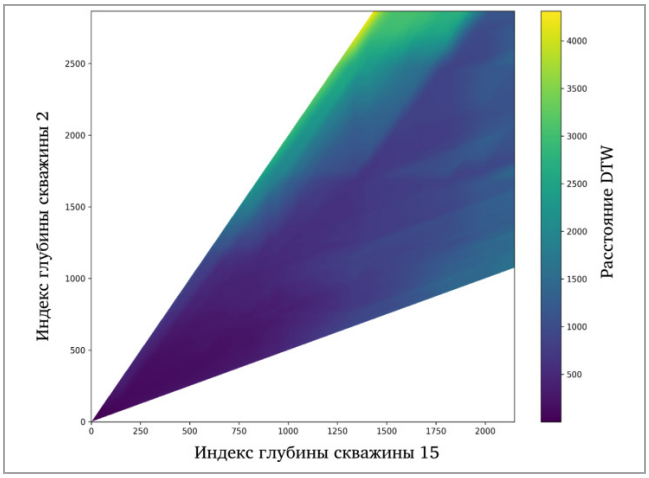


Рис. 4. Матрица расстояний между парой скважин на примере гамма и нейтронного гамма-каротажа

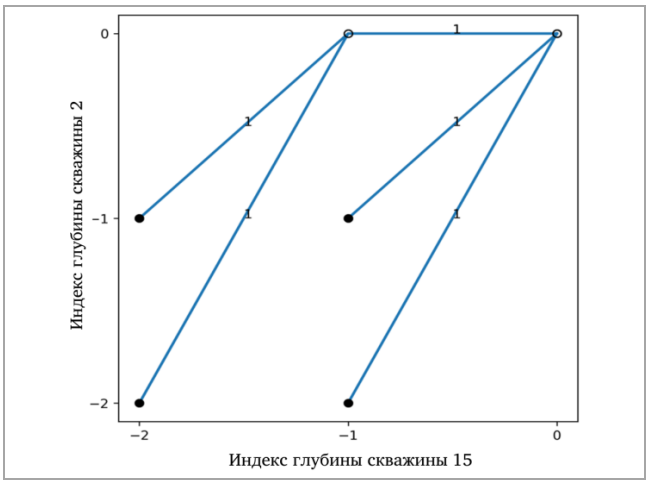


Рис. 5. Шаблон шага rabinerJuangStepPattern (4, «c»). При помощи выбора шаблона шага (цифра) и весов (буква) можно задать разные комбинации обхода, в зависимости от конфигурации сравниваемых временных рядов

После того как матрица расстояний DTW для всех пар скважин, создается квадратная матрица размером $N \times N$, где N – количество скважин. Элементы матрицы инициализируются значениями, соответствующими расстояниям DTW между парами скважин (рис. 6). Данная матрица будет использоваться в дальнейшем для кластеризации скважин на основе их сходства.

Забегая вперед, стоит отметить, что использование алгоритма DTW для целей кластеризации и для поиска лучшего соответствия между кривыми ГИС имеет разные подходы:

Кластеризация:

– цель: группировка скважин на основе минимального расстояния DTW;

– метод: используется сводная матрица расстояний DTW для разделения скважин на кластеры с помощью алгоритма K-means;

– результат: каждая скважина получает метку кластера, которая определяется на основе общего сходства с другими скважинами в кластере.

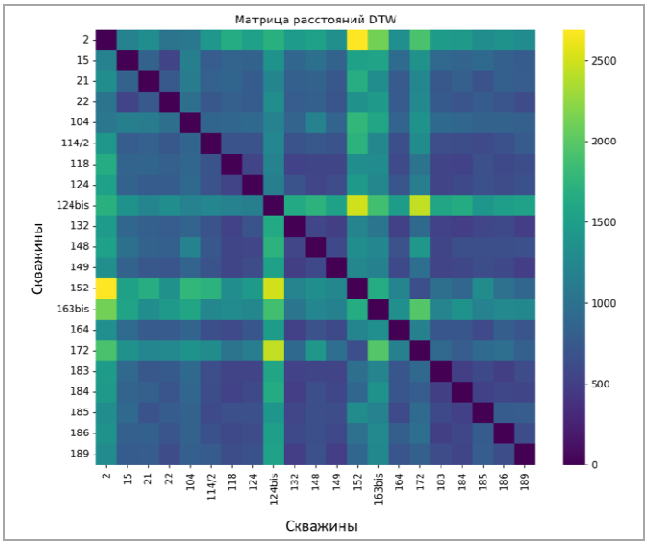
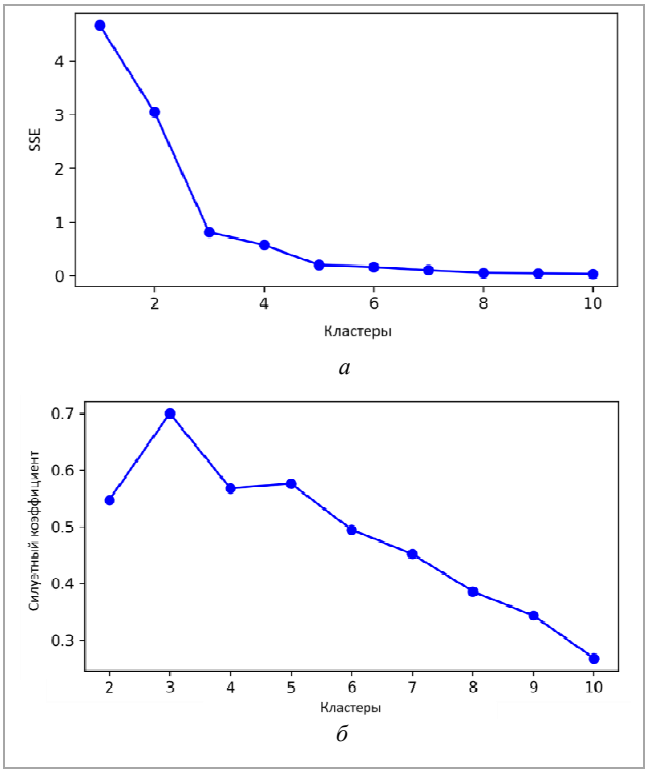


Рис. 6. Матрица расстояний DTW для всех пар скважин



а

б

Рис. 7. Графики: *а* – метод локтя; *б* – метод силуэта

Выбор лучшего соответствия между эталонной и целевой скважинами:

– цель: поиск эталонной скважины, которая наиболее похожа на целевую скважину;

– метод: сравнение каждой целевой скважины с каждой эталонной скважиной;

– результат: лучшая скважина выбирается, исходя из минимального расстояния DTW, независимо от метки кластера.

Подводя итог, DTW в кластеризации используется с целью анализа глобальных закономерностей для группировки скважин по сходству, тогда как для выбора лучшей скважины оценивается локальное сходство между двумя конкретными скважинами, независимо от их кластерной принадлежности.

Для упрощения визуализации и снижения вычислительной сложности матрица расстояний DTW преобразуется с использованием метода главных компонент (PCA). PCA – это статистический метод, который преобразует исходные переменные в новые переменные, называемые главными компонентами, которые не коррелируют между собой и упорядочены по величине их дисперсии. Это позволяет сохранить наиболее важную информацию о данных в пространстве с меньшей размерностью [39, 40].

Формально PCA выполняется путем умножения исходной матрицы данных X на матрицу весов W, которая получена из собственных векторов ковариационной матрицы X. Результатом является преобразованная матрица данных Z:

$$Z = XW. \tag{7}$$

Одним из ключевых аспектов применения метода главных компонент является определение оптимального количества главных компонент, которые задаются в процессе вычисления собственных векторов ковариационной матрицы. В нашем случае было выбрано сохранение двух главных компонент, поскольку основной задачей является визуальная интерпретация и анализ группировки скважин, что гораздо проще представить в двумерном пространстве. Кроме того, в данной работе две главные компоненты объясняют достаточно большую часть общей дисперсии данных, а именно 73 %, что делает их приемлемым компромиссом между сохранением информации и упрощением данных. Доля объясненной дисперсии для первой компоненты составляет 48 %, а для второй компоненты – 25 %.

Перед применением кластеризации необходимо определить оптимальное количество кластеров. Для решения этой задачи предлагается использовать метод локтя и метод силуэта [41–43]. Метод локтя основан на анализе суммы квадратов расстояний внутри кластеров (SSE), а метод силуэта оценивает качество кластеризации для каждого объекта. Использование обоих методов позволяет более точно определить оптимальное количество кластеров и улучшить результаты кластеризации.

Формула для вычисления SSE выглядит следующим образом:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2, \tag{8}$$

где k – количество кластеров; C_i – i-й кластер; x – точка данных в кластере C_i; μ_i – центроид кластера C_i.

Алгоритм метода локтя заключается в следующем: для каждого k от 1 до заданного максимального значения (в нашем случае задается 10 кластеров) выполняется кластеризация, вычисляется SSE для каждого k, строится график зависимости SSE от k, оптимальное количество кластеров определяется по точке, в которой кривая начинает выравниваться (рис. 7, *а*).

Коэффициент силуэта $s\,(i)$ для объекта i вычисляется по формуле:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}, \tag{9}$$

где $a(i)$ – среднее расстояние от объекта i до всех других объектов в том же кластере; $b(i)$ – среднее расстояние от объекта i до всех объектов в ближайшем соседнем кластере.

Алгоритм метода силуэта заключается в следующем: для каждого k от 2 до заданного максимального значения выполняется кластеризация,

затем вычисляется средний коэффициент силуэта для всех объектов при данном k, строится график зависимости среднего коэффициента силуэта от k, и оптимальное количество кластеров определяется по максимальному значению среднего коэффициента силуэта (рис. 7, *б*).

В результате применения методов локтя и силуэта было установлено, что оптимальным выбором являются три кластера, так как дальнейшее увеличение числа кластеров не приводит к значительному улучшению качества кластеризации. Однако на графиках наблюдается некоторая реакция на значение в пять кластеров, что может указывать на наличие более сложной структуры данных, связанной с геологическими особенностями, такими как стратиграфические несогласия, тектонические нарушения и изменения условий осадконакопления, которые могут вносить неопределенность в межскважинную корреляцию, использование в таком случае пяти кластеров поможет сформировать более репрезентативную выборку обучающих скважин, которые будут охватывать более широкий диапазон поведения кривых ГИС, и, соответственно, повысится точность проведения корреляции.

Далее проводится кластеризация с использованием метода K-means [44]. Случайным образом выбираются k центроидов, затем каждая точка данных присваивается ближайшему центроиду, что формально выражается как:

$$c_i = \arg\min_j \|x_i - \mu_j\|^2, \tag{10}$$

где c_i – индекс кластера для точки x_i; μ_j – центроид кластера j.

После этого центроиды пересчитываются как среднее значение всех точек в соответствующем кластере:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i, \tag{11}$$

где C_j – множество точек, назначенных кластеру j; $|C_j|$ – количество точек в этом кластере.

Шаги присваивания точек кластерам и пересчета центроидов повторяются до сходимости, то есть до тех пор, пока центроиды не перестанут изменяться (рис. 8).

По результатам кластеризации большая часть скважин сформировала два компактных кластера, что свидетельствует о схожем характере кривых ГИС в этих группах. Однако несколько скважин оказались отделены в обособленные кластеры, значительно отличаясь от основной массы данных.

Такое разделение может быть обусловлено различными факторами. Во-первых, геологические особенности, такие как тектоника или условия осадконакопления, могли привести к формированию нетипичных кривых ГИС в этих скважинах. Во-вторых, возможно, что качество данных ГИС в этих скважинах неудовлетворительное, что также могло привести к «выбросам». Выявленные уникальные группы скважин требуют более внимательного изучения и индивидуального анализа.

Для оценки качества кластеризации будем использовать индекс Дэвиса – Болдина (Davies – Bouldin Index) и индекс Калински – Харабаса (Calinski – Harabasz Index) [45].

Индекс Дэвиса – Болдина (DBI) вычисляется как среднее значение отношения внутрикластерного расстояния к межкластерному расстоянию:

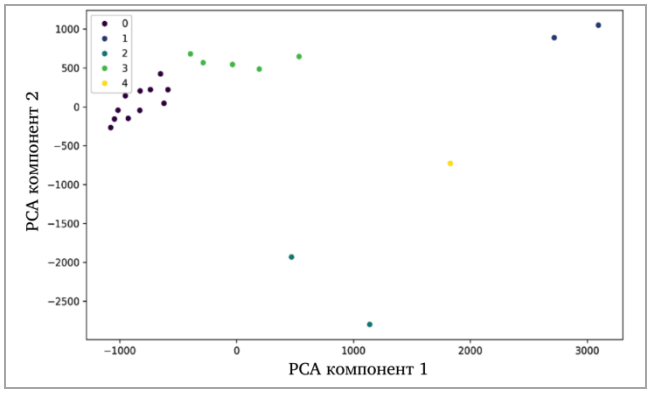


Рис. 8. Кластеризация скважин с применением PCA

$$\mathrm{DBI} = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i, c_j)}\right), \tag{12}$$

где k – количество кластеров; S_i и S_j – внутрикластерное расстояние для кластеров i и j соответственно; $d(c_i, c_j)$ – межкластерное расстояние между центроидами кластеров i и j.

Индекс Калински – Харабаса (CHI) вычисляется как отношение межкластерной дисперсии к внутрикластерной дисперсии:

$$\mathrm{CHI} = \frac{BGSS}{WGSS} \cdot \frac{n-k}{k-1}, \tag{13}$$

где $BGSS$ – межкластерная сумма квадратов; $WGSS$ – внутрикластерная сумма квадратов; k – количество кластеров.

В результате проведенной кластеризации данных с использованием метода *K*-means индекс Дэвиса – Болдина составил 0,37, что указывает на низкую внутрикластерную дисперсию и высокое межкластерное расстояние, что является признаком хорошего качества кластеризации. Индекс Калински – Харабаса равен 88,90, что также свидетельствует о хорошей разделимости кластеров.

После определения минимальной выборки скважин производится поиск лучшего соответствия между парами «эталонная скважина – целевая скважина» путем вычисления оптимального пути на матрице расстояний по алгоритму, описанному ранее, далее выбирается пара скважин с минимальным расстоянием DTW.

Определение пластопересечений в целевых скважинах производится на основе заранее известных экспертных отбивок в эталонных скважинах. Для каждой точки пластопересечения в эталонной скважине находится соответствующая точка в целевой скважине при помощи индексов из матрицы выравнивания DTW.

Сначала находится ближайшая точка в эталонной скважине к заданной глубине пластопересечения d_k. Это делается путем поиска минимальной разницы между заданной глубиной и всеми глубинами в эталонной скважине:

$$i_{\min} = \mathrm{argmin}|D_1 - d_k|, \tag{14}$$

где D_1 – массив значений глубины для эталонной скважины; $i_{\min}$ – индекс ближайшей точки в массиве глубин эталонной скважины к заданной глубине.

Затем используется матрица $\mathrm{DTW}(i, j)$. Матрица содержит информацию о том, какие индексы в эталонной скважине соответствуют индексам в целевой

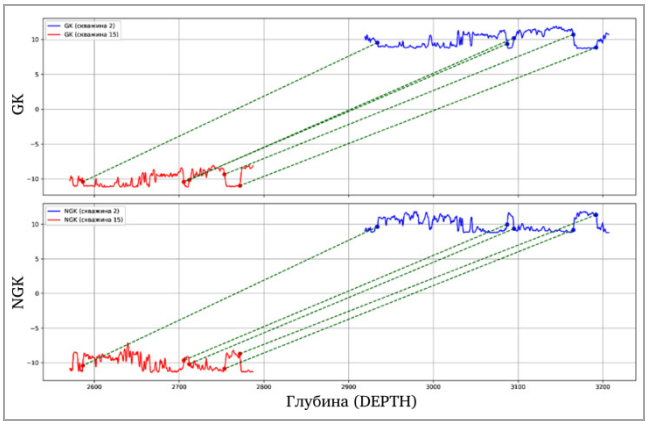


Рис. 9. Сопоставление точек пластопересечений

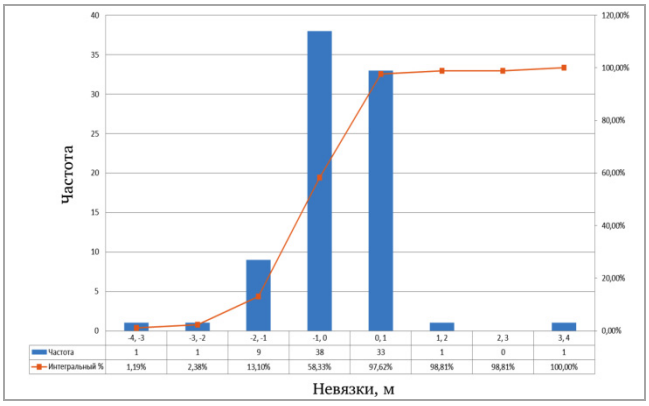


Рис. 11. Невязки между экспертными и рассчитанными пластопересечениями

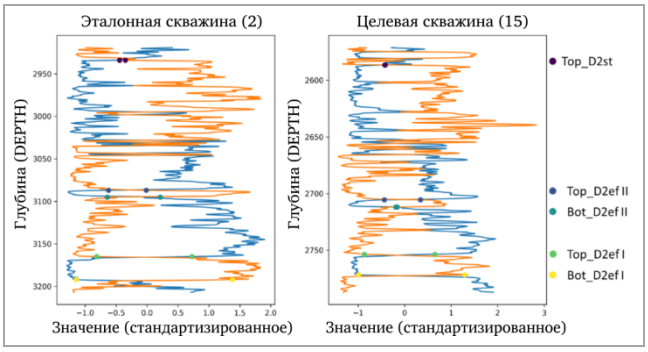


Рис. 10. Результат корреляции на примере скважин 2 и 15



Рис. 12. Схема корреляции пласта D2ef-II (скв. 189 и 21)

скважине. Для нахождения соответствующей точки в целевой скважине используется индекс $i_{\min}$:

$$j_{\min} = \mathrm{DTW}(i_{\min}, j), \qquad (15)$$

где $j_{\min}$ – индекс соответствующей точки в целевой скважине.

Таким образом, искомое пластопересечение в целевой скважине будет на глубине $D_2[j_{\min}]$, где D_2 – массив значений глубины для целевой скважины (рис. 9).

Данный алгоритм выполняется по каждой паре скважин, в результате получены рассчитанные пластопересечения по целевым скважинам (рис. 10).

Анализ результатов

Для оценки результатов был выполнен сравнительный анализ между экспертными и рассчитанными по алгоритму пластопересечениями. Общее количество пластопересечений, на основе которых была проведена апробация алгоритма, составляет 84. Распределение расхождений по диапазонам показало, что большинство результатов (84,53 %) сосредоточено в диапазоне –1…1, что указывает на высокую согласованность результатов с экспертными оценками в большинстве случаев (рис. 11).

Тем не менее встречаются значительные расхождения. В ходе детального анализа скважин с наличием значительных (>1 м) невязок было установлено, что не во всех случаях причиной погрешностей является работа алгоритма.

Рассмотрим пару скважин 189 и 21 (рис. 12). По скважине 21 невязка между рассчитанным и экспертным глубинами кровли пласта D2ef-II составляет 3,1 м.

В эталонной и целевой скважинах эксперт применил различные методики выделения кровли пласта D2ef-II, что привело к расхождению в результатах. В противоположность этому алгоритм DTW продемонстрировал свою эффективность, корректно выделив отбивки согласно обучающему шаблону, независимо от субъективных факторов. Этот случай подчеркивает, что алгоритм может выявлять ошибки, которые остались незамеченными экспертами. Таким образом, использование алгоритма DTW не только повышает точность и объективность интерпретации данных, но и служит инструментом для выявления и коррекции субъективных ошибок, связанных с человеческим фактором.

Заключение

Предложенный в работе подход на основе алгоритма динамической трансформации временной шкалы показал высокую эффективность и точность межскважинной корреляции, выполненной в автоматическом режиме. Такой подход позволяет значительно сократить временные затраты, повысить объективность результатов и выявить ошибки, связанные с человеческим фактором при экспертной интерпретации, что делает его перспективным для широкого практического применения в геологической и геофизической отраслях, особенно при работе с большими объемами данных.

Библиографический список

1. Keogh, E. Scaling up Dynamic Time Warping for Datamining Applications / E. Keogh, M. Pazzani // Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2002. – P. 285–289. DOI: 10.1145/347090.347153
2. Scikit-learn: Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort [et al.]// Journal of Machine Learning Research, – 2011. – No. 12. – P. 2825–2830. DOI: 10.48550/arXiv.1201.0490
3. McKinnev. W. Data structures for statistical computing in Python / W. McKinney // Proceedings of the 9th Python in Science Conference, – 2010. – P. 1–6. DOI: 10.25080/Majora-92bf1922-00a
4. Hunter, J.D. Matplotlib: A 2D graphics environment / J.D. Hunter // Computing in Science & Engineering. – 2007. – No. 9 (3). – P. 90–95.
5. Семенов, И.А. Использование возможностей библиотеки Matplotlib для визуализации данных в работе инженеров и исследователей / И.А. Семенов, М.Д. Шишкин // Современные технологии и научно-технический прогресс. – 2024.– № 11. – С. 169–170.
6. Таршхоева, Ж.Т. Язык программирования Python. Библиотеки Python / Ж.Т. Таршхоева // Молодой ученый. – 2021. – № 5 (347). – С. 20–21.
7. Галимуллин, Н.Р. Python: использование Python для автоматизации повседневных задач / Н.Р. Галимуллин // Экономика и управление: проблемы, решения. – 2024. – Т. 9, № 9 (150). – С. 69–76. DOI: 10.36871/ek.up.p.r.2024.09.09.010
8. Ефремов, Е.А. Разработка графических интерфейсов с помощью графической библиотеки Tkinter / Е.А. Ефремов // Молодежный научно-технический вестник. – 2017. – № 6. – С. 44.
9. Дрянкова, Д.А. Обработка пропущенных значений и дубликатов в данных с помощью библиотеки Pandas для языка программирования Python / Д.А. Дрянкова // Дневник науки. – 2023. – № 6 (78). DOI: 10.51691/2541-8327_2023_6_11
10. Van der Walt, S. The NumPy array: A structure for efficient numerical computation / S. Van der Walt, S.C. Colbert, G. Varoquaux // Computing in Science & Engineering. – 2011. – No. 13(2). – P. 22–30. DOI: 10.1109/MCSE.2011.37
11. Array programming with NumPy / C.R. Harris, K.J. Millman, S.J. Van der Walt [et al.] // Nature. – 2020. – No. 585 (7825). – P. 357–362. DOI: 10.1038/s41586-020-2649-2
12. Li, Ya. Exploring Indexing and Slicing in NumPy Arrays / Ya. Li, Yu. Shao, Ch. Fu // Journal of Physics: Conference Series. – 2024. – Vol. 2832, no. 1. – P. 012019. DOI: 10.1088/1742-6596/2832/1/012019
13. Waskom, M. Seaborn: Statistical data visualization / M. Waskom // Journal of Open Source Software. – 2021. – Vol. 6, no. 60. – P. 3021. DOI: 10.21105/joss.03021
14. Machine learning in petrophysics: Advantages and limitations / Ch. Xu, L. Fu, T. Lin, W. Li, Sh. Ma // Artificial Intelligence in Geosciences. – 2022. – No. 3. DOI: 10.1016/j.aiig.2022.11.004
15. Костиков, Д.В. Сравнение методов уменьшения обучающей выборки при интерпретации данных геофизических исследований скважин с помощью нейронной сети / Д.В. Костиков, А.И. Ларюхин, В.Е. Лялин // Труды международного симпозиума «Надежность и качество». – 2007. – Т. 1. – С. 118–123.
16. Гулин, А.Б. Применение методов машинного обучения для интерпретации данных геофизических исследований скважин / А.Б. Гулин // Проблемы разработки месторождений углеводородных и рудных полезных ископаемых. – 2022. – Т. 1. – С. 41–45.
17. Ковалевский, Е.В. Уточнение геологических моделей посредством использования автоматической корреляции скважин / Е.В. Ковалевский, Г.Н. Гогоненков, М.В. Перепечкин // Недропользование XXI век. – 2007. – № 4. – С. 28–31.
18. Lineman, D.J. Well to well log correlation using knowledge-based systems and dynamic depth warping / D.J. Lineman, J.D. Mendelson, M.N. Toksos // SPWLA 28th Annual Logging Symposium “High-impact palynology in petroleum geology: applications from Venezuela (Northern South America)” (June 29 – July 2, 1987). – AAPG Bull., 1987. – Vol. 86. – P. 421–453.
19. Multiple Imputation by Chained Equations: What is it and how does it work? / M.J. Azur, E.A. Stuart, C. Frangakis, P.J. Leaf // International Journal of Methods in Psychiatric Research. – 2011. – Vol. 20, no. 1. – P. 36–49. DOI: 10.1002/mpr.329
20. Siddique, J. A Comparison of Six Methods for Missing Data Imputation / J. Siddique, T.R. Belin // Journal of Biometrics & Biostatistics. – 2008. – Vol. 3, no. 2. – P. 1–15. DOI: 10.4172/2155-6180.1000224
21. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data / R. Wu, S.D. Hamshaw, L. Yang [et al.] // IEEE Sensors Journal. – 2022. – Vol. 22, no. 11. – P. 10671–10683. DOI: 10.1109/jsen.2022.3166643
22. Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring / Y. Chen, D. Li, Y. Li, X. Ma, J. Wei // Lecture Notes of the Institute for Computer Sciences. – 2016. – No. 181. – P. 193–203. DOI: 10.1007/978-3-319-49655-9_26
23. Soltanian, M. Recognition of Visual Events using Spatio-Temporal Information of the Video Signal / M. Soltanian, Sh. Ghaemmaghami // Signal and Data Processing. – 2021. – Vol. 18, no. 1. – P. 134–119. DOI: 10.52547/jsdp.18.1.134
24. Al-Faiz, M.Z. The Effect of Z-Score Standardization (Normalization) on Binary Input Due the Speed of Learning in Back-Propagation Neural Network / M.Z. Al-Faiz, A.A. Ibrahim, S.M. Hadi // Ira. J. Inf. Commun. Technol. – 2018. – Vol. 1. – P. 42–48. DOI: 10.31987/ijict.1.3.41
25. Al Shalabi, L. Normalization as a preprocessing engine for data mining and the approach of preference matrix / L. Al Shalabi, Z. Shaaban // International Conference on Dependability of Computer Systems, 25–27 May 2006. – Szklarska Poreba, 2006. – P. 207–214. DOI: 10.1109/DEPCOS-RELCOMEX.2006.38
26. Salvador, S. Fast DTW: Toward accurate dynamic time warping in linear time and space / S. Salvador, P. Chan // Intelligent Data Analysis. – 2004. – No. 11 (5). – P. 70–80.
27. Kate, RJ. Using dynamic time warping distances as features for improved time series classification / RJ. Kate // Data Mining and Knowledge Discovery. – 2016. – Vol. 30, no. 2. – P. 283–312. DOI: 10.1007/s10618-015-0418-x
28. Sylvester, Z. Automated multi-well stratigraphic correlation and model building using relative geologic time / Z. Sylvester // Basin Research. – 2023. – Vol. 35, no. 5. – P. 1961–1984. DOI: 10.1111/bre.12787
29. Шарафутдинов, Т.Р. Апробация алгоритма автокорреляции данных геофизических исследований скважин на примере отложений ачимовской толщи Правдинского месторождения / Т.Р. Шарафутдинов, Р.А. Шайбаков // Научно-технический вестник ОАО «НК «Роснефть». – 2012. – № 1. – С. 18–22.
30. Автоматизированная увязка сейсмических и скважинных данных с использованием алгоритма Dynamic Time Warping / А.В. Буторин, А.И. Севостьянов, С.К. Студиков, А.А. Тимиргалин // Нефтяное хозяйство. – 2019. – № 12. – С. 30–32. – DOI: 10.24887/0028-2448-2019-12-30-32
31. Sakoe, H. Dynamic programming algorithm optimization for spoken word recognition / H. Sakoe, S. Chiba // IEEE Transactions on Acoustics, Speech and Signal Processing. – 1978. – Vol. 26, iss.1. – P. 43–49. DOI: 10.1109/TASSP.1978.1163055
32. Технология оценки согласованности результатов автоматической корреляции скважин по данным геофизических исследований / К.Е. Закревский, Р.К. Газизов, Е.А. Рыжиков, К.В. Фрейдин // Нефтяное хозяйство. – 2021. – № 8. – С. 22–26. DOI: 10.24887/0028-2448-2021-8-22-26
33. Разработка подходов к автоматизированной внутрипластовой корреляции по данным геофизических исследований скважин с применением машинного обучения / И.Д. Латыпов, Л.Р. Шагимарданова, А.В. Марков, Н.А. Евграфов // Экспозиция Нефть Газ. – 2024. – № 4 (105). – С. 47–51. DOI: 10.24412/2076-6785-2024-4-47-51
34. Event DTW: An improved Dynamic Time Warping algorithm for aligning biomedical signals of nonuniform sampling frequencies / Y. Jiang, Y. Qi, W.K. Wang [et al.] // Sensor. – 2020. – Vol. 20 (9). – P. 1–13. DOI: 10.3390/s20092700
35. Study on Rock Mechanics Parameter Prediction Method Based on DTW Similarity and Machine-Learning Algorithms / W. Cai, J. Ding, Zh. Li [et al.] // Petrophysics. – 2024. – Vol. 65, no. 1. – P. 128–144. DOI: 10.30632/pjv65n1-2024a7
36. The dynamic time-warping approach to comparison of magnetic-susceptibility logs and application to lower devonian calciturbidites (Prague Synform, Bohemian Massif) / J. Hladil, P. Cejchan, L. Koptikova [et al.] // Geologica Belgica. – 2010. – Vol. 13, no. 4. – P. 385–406.
37. Романенко, А.А. Выравнивание временных рядов: прогнозирование с использованием DTW / А.А. Романенко // Машинное обучение и анализ данных. – 2011. – Т. 1, № 1. – С. 77–85.
38. Rabiner, L. An introduction to hidden Markov models / L. Rabiner, B. Juang // IEEE ASSP Magazine. – 1986. – Vol. 3. – No. 1. – P. 4–16. DOI: 10.1109/MASSP.1986.1165342
39. Денисенко, А.А. Исследование методов сокращения размерности в задаче распознавания образов / А.А. Денисенко // Международный журнал прикладных и фундаментальных исследований. – 2020. – № 12. – С. 65–69. DOI: 10.17513/mjpfi.13161
40. Basna, H. Journal of Soft Computing and Data Mining A Review of Principal Component Analysis Algorithm for Dimensionality Reduction / H. Basna, A. Adnan // Applied Soft Computing. – 2021. – No. 2. – P. 20–30. DOI: 10.30880/jscdm.2021.02.01.003
41. Бова, В.В. Методы и алгоритмы кластеризации текстовых данных (обзор) / В.В. Бова, Ю.А. Кравченко, С.И. Родзин // Известия ЮФУ. Технические науки. – 2022. – № 4 (228). – С. 122–143. DOI: 10.18522/2311-3103-2022-4-122-143
42. Орлов, В.И. Применение критерия силуэта в алгоритме автоматической группировки электрорадиоизделий космического применения / В.И. Орлов, Л.А. Казаковцев, И.С. Масич // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2016. – Т. 17, № 4. – С. 883–890.
43. Umargono, E. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula / E. Umargono, J. Suseno, S.K. Gunawan // The 2nd International Seminar on Science and Technology. – 2019. DOI: 10.2991/assehr.k.201010.019
44. Hartigan, J.A. A k-means clustering algorithm / J.A. Hartigan, M.A. Wong // Journal of the Royal Statistical Society. Series C (Applied Statistics). – 1979. – Vol. 28, iss. 1. – P. 100–108. DOI: 10.2307/2346830
45. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta / I. Firman Ashari, E. Nugroho, R. Baraku, I. Yanda, R. Liwardana // Journal of Applied Informatics and Computing. – 2023. – No. 7. – P. 89–97. DOI: 10.30871/jaic.v7i1.4947

References

1. Keogh E., Pazzani M. Scaling up Dynamic Time Warping for Datamining Applications. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 285-289. DOI: 10.1145/347090.347153
2. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, no. 12, pp. 2825-2830. DOI: 10.48550/arXiv.1201.0490
3. McKinney W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 2010, pp. 1-6. DOI: 10.25080/Majora-92bf1922-00a
4. Hunter J.D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 2007, no. 9 (3), pp. 90-95.
5. Semenov I.A., Shishkin M.D. Ispol'zovanie vozmozhnostei biblioteki Matplotlib dlia vizualizatsii dannykh v rabote inzhenerov i issledovatelei [Using the potential of the Matplotlib library for data visualization in the work of engineers and researchers]. *Sovremennye tekhnologii i nauchno-tekhnicheskii progress*, 2024, no. 11, pp. 169-170.
6. Tarshkhoeva Zh.T. Iazyk programmirovaniia Python. Biblioteki Python [Python programming language. Python libraries]. *Molodoi uchenyi*, 2021, no. 5 (347), pp. 20-21.
7. Galimullin N.R. Python: ispol'zovanie Python dlia avtomatizatsii povsednevnykh zadach [Python: using Python to automate everyday tasks]. *Ekonomika i upravlenie: problemy, resheniia*, 2024, vol. 9, no. 9 (150), pp. 69-76. DOI: 10.36871/ek.up.p.r.2024.09.09.010
8. Efremov E.A. Razrabotka graficheskikh interfeisov s pomoshch'iu graficheskoi biblioteki Tkinter [Developing GUIs with the Tkinter GUI Library]. *Molodezhnyi nauchno-tekhnicheskii vestnik*, 2017, no. 6, 44 p.
9. Driankova D.A. Obrabotka propushchennykh znachenii i dublikatov v dannykh s pomoshch'iu biblioteki Pandas dlia iazyka programmirovaniia Python [Processing of missing values and duplicates in data using the Pandas library for the Python programming language]. *Dnevnik nauki*, 2023, no. 6 (78). DOI: 10.51691/2541-8327_2023_6_11
10. Van der Walt S., Colbert S.C., Varoquaux G. The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 2011, no. 13 (2), pp. 22-30. DOI: 10.1109/MCSE.2011.37
11. Harris C.R., Millman K.J., Van der Walt S.J. et al. Array programming with NumPy. *Nature*, 2020, no. 585 (7825), pp. 357-362. DOI: 10.1038/s41586-020-2649-2
12. Li Ya., Shao Yu., Fu Ch. Exploring Indexing and Slicing in NumPy Arrays. *Journal of Physics: Conference Series*, 2024, vol. 2832, no. 1, 012019 p. DOI: 10.1088/1742-6596/2832/1/012019
13. Waskom M. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 2021, vol. 6, no. 60, 3021 p. DOI: 10.21105/joss.03021
14. Xu Ch., Fu L., Lin T., Li W., Ma Sh. Machine learning in petrophysics: Advantages and limitations. *Artificial Intelligence in Geosciences*, 2022, no. 3. DOI: 10.1016/j.aiig.2022.11.004
15. Kostikov D.V., Lariukhin A.I., Lialin V.E. Sravnenie metodov umen'sheniia obuchaiushchei vyborki pri interpretatsii dannykh geofizicheskikh issledovanii skvazhin s pomoshch'iu neironnoi seti [Comparison of methods for reducing the training sample in the interpretation of well logging data using a neural network]. *Trudy mezhdunarodnogo simpoziuma "Nadezhnost' i kachestvo"*, 2007, vol. 1, C. 118–123.
16. Gulin A.B. Primenenie metodov mashinnogo obucheniia dlia interpretatsii dannykh geofizicheskikh issledovanii skvazhin [Application of machine learning methods to interpret well logging data]. *Problemy razrabotki mestorozhdenii uglevodorodnykh i rudnykh poleznykh iskopaemykh*, 2022, vol. 1, pp. 41-45.
17. Kovalevskii E.V., Gogonenkov G.N., Perepechkin M.V. Utochnenie geologicheskikh modelei posredstvom ispol'zovaniia avtomaticheskoi korreliatsii skvazhin [Refinement of geological models using automatic well correlation]. *Nedropol'zovanie XXI vek*, 2007, no. 4, pp. 28-31.
18. Lineman D.J., Mendelson J.D., Toksos M.N. Well to well log correlation using knowledge-based systems and dynamic depth warping. *SPWLA 28th Annual Logging Symposium "High-impact palynology in petroleum geology: applications from Venezuela (Northern South America)" (June 29 – July 2, 1987)*. AAPG Bull., 1987, vol. 86, pp. 421-453.
19. Azur M.J., Stuart E.A., Frangakis C., Leaf P.J. Multiple Imputation by Chained Equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 2011, vol. 20, no. 1, pp. 36-49. DOI: 10.1002/mpr.329
20. Siddique J., Belin T.R. A Comparison of Six Methods for Missing Data Imputation. *Journal of Biometrics & Biostatistics*, 2008, vol. 3, no. 2, pp. 1-15. DOI: 10.4172/2155-6180.1000224
21. Wu R., Hamshaw S.D., Yang L. et al. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data. *IEEE Sensors Journal*, 2022, vol. 22, no. 11, pp. 10671-10683. DOI: 10.1109/jsen.2022.3166643
22. Chen Y., Li D., Li Y., Ma X., Wei J. Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring. *Lecture Notes of the Institute for Computer Sciences*, 2016, no. 181, pp. 193-203. DOI: 10.1007/978-3-319-49655-9_26
23. Soltanian M., Ghaemmaghami Sh. Recognition of Visual Events using Spatio-Temporal Information of the Video Signal. *Signal and Data Processing*, 2021, vol. 18, no. 1, pp. 134-119. DOI: 10.52547/jsdp.18.1.134
24. Al-Faiz M.Z., Ibrahim A.A., Hadi S.M. The Effect of Z-Score Standardization (Normalization) on Binary Input Due the Speed of Learning in Back-Propagation Neural Network. *Ira. J. Inf. Commun. Technol.*, 2018, vol. 1, pp. 42-48. DOI: 10.31987/ijict.1.3.41
25. Al Shalabi L., Shaaban Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. *International Conference on Dependability of Computer Systems, 25-27 May 2006*. Szklarska Poreba, 2006, pp. 207-214. DOI: 10.1109/DEPCOS-RELCOMEX.2006.38
26. Salvador S., Chan P. Fast DTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 2004, no. 11 (5), pp. 70-80.
27. Kate RJ. Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, 2016, vol. 30, no. 2, pp. 283-312. DOI: 10.1007/s10618-015-0418-x
28. Sylvester Z. Automated multi-well stratigraphic correlation and model building using relative geologic time. *Basin Research*, 2023, vol. 35, no. 5, pp. 1961-1984. DOI: 10.1111/bre.12787
29. Sharafutdinov T.R., Shaibakov R.A. Aprobatsiia algoritma avtokorreliatsii dannykh geofizicheskikh issledovanii skvazhin na primere otlozhenii achimovskoi tolshchi Pravdinskogo mestorozhdeniia [Testing an algorithm of the well log data autocorrelation on the example of Achimov sequence of Pravdinskoye field]. *Nauchno-tekhnicheskii vestnik OAO "NK "Rosneft'"*, 2012, no. 1, pp. 18-22.
30. Butorin A.V., Sevost'ianov A.I., Stulikov S.K., Timirgalin A.A. Avtomatizirovannaia uviazka seismicheskikh i skvazhinnykh dannykh s ispol'zovaniem algoritma Dynamic Time Warping [DTW based automated seismic–well tie]. *Neftianoe khoziaistvo*, 2019, no. 12, pp. 30-32. DOI: 10.24887/0028-2448-2019-12-30-32
31. Sakoe H., Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1978, vol. 26, iss. 1, pp. 43-49. DOI: 10.1109/TASSP.1978.1163055
32. Zakrevskii K.E., Gazizov R.K., Ryzhikov E.A., Freidin K.V. Tekhnologiia otsenki soglasovannosti rezul'tatov avtomaticheskoi korreliatsii skvazhin po dannym geofizicheskikh issledovanii [Consistency evaluation technology for automatic well-log correlation using well logging data]. *Neftianoe khoziaistvo*, 2021, no. 8, pp. 22-26. DOI: 10.24887/0028-2448-2021-8-22-26
33. Latypov I.D., Shagimardanova L.R., Markov A.V., Evgrafov N.A. Razrabotka podkhodov k avtomatizirovannoi vnutriplastovoi korreliatsii po dannym geofizicheskikh issledovanii skvazhin s primeneniem mashinnogo obucheniia [Development of approaches to automated correlation from well log data using machine learning]. *Ekspozitsiia Neft' Gaz*, 2024, no. 4 (105), pp. 47-51. DOI: 10.24412/2076-6785-2024-4-47-51
34. Jiang Y., Qi Y., Wang W.K. et al. Event DTW: An improved Dynamic Time Warping algorithm for aligning biomedical signals of nonuniform sampling frequencies. *Sensor*, 2020, vol. 20 (9), pp. 1-13. DOI: 10.3390/s20092700
35. Cai W., Ding J., Li Zh. et al. Study on Rock Mechanics Parameter Prediction Method Based on DTW Similarity and Machine-Learning Algorithms. *Petrophysics*, 2024, vol. 65, no. 1, pp. 128-144. DOI: 10.30632/pjv65n1-2024a7
36. Hladil J., Cejchan P., Koptikova L. et al. The dynamic time-warping approach to comparison of magnetic-susceptibility logs and application to lower devonian calciturbidites (Prague Synform, Bohemian Massif). *Geologica Belgica*, 2010, vol. 13, no. 4, pp. 385-406.
37. Romanenko A.A. Vyravnivanie vremennykh riadov: prognozirovanie s ispol'zovaniem DTW [Time Series Alignment: Forecasting Using DTW]. *Mashinnoe obuchenie i analiz dannykh*, 2011, vol. 1, no. 1, pp. 77-85.
38. Rabiner L., Juang B. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 1986, vol. 3, no. 1, pp. 4-16. DOI: 10.1109/MASSP.1986.1165342
39. Denisenko A.A. Issledovanie metodov sokrashcheniia razmernosti v zadache raspoznavaniia obrazov [Research of dimensionality reduction methods in the pattern recognition problem]. *Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovanii*, 2020, no. 12, pp. 65-69. DOI: 10.17513/mjpfi.13161
40. Basna H., Adnan A. Journal of Soft Computing and Data Mining A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Applied Soft Computing*, 2021, no. 2, pp. 20-30. DOI: 10.30880/jscdm.2021.02.01.003
41. Bova V.V., Kravchenko Iu.A., Rodzin S.I. Metody i algoritmy klasterizatsii tekstovykh dannykh (obzor) [Methods and algorithms for text data clustering (review)]. *Izvestiia IuFU. Tekhnicheskie nauki*, 2022, no. 4 (228), pp. 122-143. DOI: 10.18522/2311-3103-2022-4-122-143
42. Orlov V.I., Kazakovtsev L.A., Masich I.S. Primenenie kriteriia silueta v algoritme avtomaticheskoi gruppirovki elektroradioizdelii kosmicheskogo primeneniia [Silhouette criterion for automatic grouping algorithm of spaceship electronic components]. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M.F. Reshetneva*, 2016, vol. 17, no. 4, pp. 883-890.
43. Umargono E., Suseno J., Gunawan S.K. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. *The 2nd International Seminar on Science and Technology*, 2019. DOI: 10.2991/assehr.k.201010.019
44. Hartigan J.A., Wong M.A. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, vol. 28, iss. 1, pp. 100-108. DOI: 10.2307/2346830
45. Firman Ashari I., Nugroho E., Baraku R., Yanda I., Liwardana R. Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*, 2023, no. 7, pp. 89-97. DOI: 10.30871/jaic.v7i1.4947

Финансирование. Исследование выполнено в рамках гранта Российского научного фонда № 22-17-00111.
Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Вклад авторов равноценен.