

Научная статья

УДК 004.89

DOI 10.25205/1818-7900-2024-22-2-44-56

Разработка метода для оптимизации процесса получения значений числовых характеристик компонентов нефтяного сырья из графических данных хроматограмм

**Петр Андреевич Пылов¹, Роман Вячеславович Майтак¹
Даниил Евгеньевич Копылов²**

¹Кузбасский государственный технический университет им. Т. Ф. Горбачева
Кемерово, Россия

²Иркутский государственный университет
Иркутск, Россия

pylovpa@kuzstu.ru; <https://orcid.org/0009-0009-4353-6925>
maytak.roman@mail.ru; <https://orcid.org/0009-0009-4353-6925>
it-daniil@yandex.ru; <https://orcid.org/0009-0005-0554-8934>

Аннотация

Разработан вариант алгоритма (и метода на его основе) для выполнения процедуры автоматизированного восстановления числовых значений графически представленной функции сигнала хроматографа, исследующей компонентный состав проб нефтяного сырья. Проблема, на решение которой направлен разработанный метод, состоит в слабой приспособленности хроматографов к нефтяной промышленности: нефть является природным сырьем, которое не является химически чистым, поэтому в рамках хроматографического исследования происходит фиксация не всех числовых характеристик компонентов, входящих в состав исследуемого образца. Значения незафиксированных координат углеводородных соединений регистрируются специалистами нефтехимических лабораторий вручную, что существенно усложняет процесс анализа. Разработанный метод в качестве входных данных принимает полученные в лаборатории изображения хроматограмм нефти, представленные в оригинальной черно-белой цветовой гамме. Выходными данными метода является массив числовых значений координат, восстановленных с шагом в один пиксель. Величина погрешности при восстановлении значений методом значительно меньше порога, установленного нефтехимической лабораторией. Кроме автоматизации обозначенной задачи, массив полученных значений координат был векторизован с целью применения вектора в качестве входных данных в модели трансформера (вместо изображений хроматограмм) для решения задачи прогнозирования перераспределения углеводородных компонентов тяжелой нефти под воздействием катализаторов. Результатом замены типа представления входной информации стало многократное снижение времени, требуемого для получения прогноза, а также времени обучения, при этом снизилась величина усредненной ошибки предсказания.

Ключевые слова

анализ данных, предварительная обработка графических данных, обработка хроматограмм тяжелой нефти, восстановление числовых координат углеводородных компонентов

Для цитирования

Пылов П. А., Майтак Р. В., Копылов Д. Е. Разработка метода для оптимизации процесса получения значений числовых характеристик компонентов нефтяного сырья из графических данных хроматограмм // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 2. С. 44–56. DOI 10.25205/1818-7900-2024-22-2-44-56

© Пылов П. А., Майтак Р. В., Копылов Д. Е., 2024

Development of a Method for Optimization of the Process of Extraction of Values of Numerical Characteristics of Oil Feedstock Components from Graphical Data of Chromatographic Analysis

Petr A. Pylov¹, Roman V. Maitak¹,
Daniil E. Kopylov²

¹Gorbachev Kuzbass State Technical University, Kemerovo, Russian Federation

²Irkutsk State University, Irkutsk, Russian Federation

pylovpa@kuzstu.ru; <https://orcid.org/0009-0009-4353-6925>

maytak.roman@mail.ru; <https://orcid.org/0009-0009-4353-6925>

it-daniil@yandex.ru; <https://orcid.org/0009-0005-0554-8934>

Abstract

A variant of the algorithm has been developed to perform the procedure of automated recovery of numerical values of graphically represented chromatograph signal function, studying the component composition of heavy oil feedstock samples. The problem, which the developed method aims to solve, consists in the poor adaptation of chromatographs to the oil industry: oil is a natural raw material, which is not chemically pure, therefore not all numerical characteristics of the components contained in the investigated sample are fixed within the chromatographic study. In the current configuration of the method, the values from the chromatogram are recorded manually. The developed method takes as input data the images of oil chromatograms obtained in the laboratory, presented in the original black and white colour scheme. The output data of the method is an array of numerical values of coordinates reconstructed with a step of one pixel. The size of the error in the reconstruction of the values by the method is much smaller than the threshold set by the petrochemical laboratory. In addition to automating the indicated task, the array of obtained coordinate values was vectorized in order to use the vector as input data in the Transformer model to solve the problem of predicting the redistribution of hydrocarbon components of heavy oil under the influence of catalysts. As a result of the change in input data representation, the time required to obtain a prediction and the training time were reduced by a multiple, while the value of the average prediction error decreased.

Keywords

data analysis, preliminary processing of graphical data, processing of heavy oil chromatograms, recovery of numerical coordinates of hydrocarbon components

For citation

Pylov P. A., Maitak R. V., Kopylov D. E., Evsikov I. A. Development of a method for optimization of the process of extraction of values of numerical characteristics of oil feedstock components from graphical data of chromatographic analysis. Vestnik NSU. Series: Information Technologies, 2024, vol. 22, no. 2, pp. 44–56 (in Russ.) DOI 10.25205/1818-7900-2024-22-2-44-56

Введение

Нефть, как энергетическое сырье, активно используется человечеством на протяжении длительного времени. За это время в предметной области нефтяного дела устоялась классификация всего природного сырья на два условных класса: легкую и тяжелую нефть [1]. Технологически добыча и переработка легкой нефти за несколько последних десятилетий была настолько хорошо налажена промышленностью, что серьезных трудностей в этой подобласти не возникает [2]. По сравнению с легкой, тяжелая нефть является гораздо более сложнокомпонентной смесью углеводородов, поэтому работать с ней значительно сложнее: затрудняется как процесс добычи, так и транспортировки (с дальнейшими ограничениями переработки, вызванными повышенной плотностью и вязкостью сырья) [3]. Для того чтобы нивелировать большинство сложностей на всех этапах жизненного цикла (от добычи сырья тяжелой нефти до превращения его в конечный энергетически ценный продукт), специалисты нефтеперерабатывающей области нашли один из способов решения проблемы в подходе, подразумевающим применение катализаторов в облагораживающей химической реакции, которая позволяет безвозвратно снизить вязкость и плотность тяжелой нефти уже на этапе ее нахождения во внутрискластовом пространстве [4].

Сложностью, которую порождает данный подход, является объемное перераспределение различных углеводородных компонентов между собой. Этот фактор в дальнейшем напрямую влияет на объемный выход конечного продукта переработки нефти и, как следствие, – на уровень коммерциализации нефтеперерабатывающего предприятия в целом.

Наиболее прецизионный способ для выбора оптимального катализатора основан на проведение хроматографического исследования [5], которое позволяет получить графическое представление компонентного состава тяжелой нефти – хроматограмму (рис. 1).

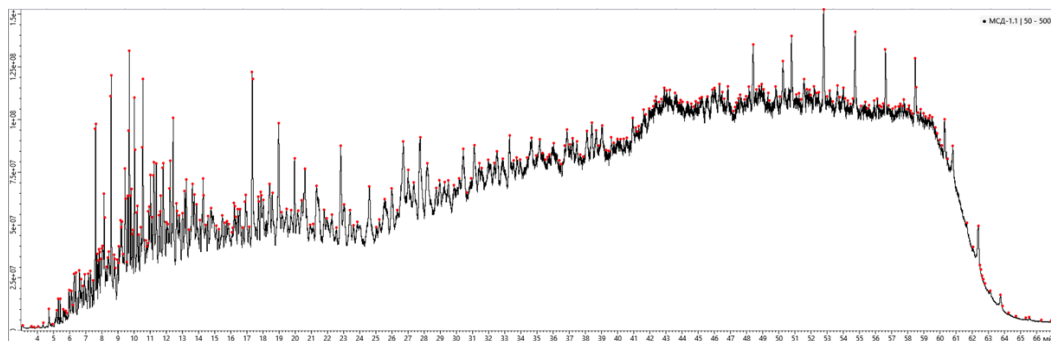


Рис. 1. Пример хроматограммы образца тяжелой нефти
Fig. 1. Example chromatogram of a heavy oil sample

Пики на хроматограмме (красные точки на рис. 1) фиксируют присутствие обособленного углеводородного компонента в исследуемом образце нефти, а значения координат пиков (время выхода по оси абсцисс и энергия выхода по оси ординат) позволяют однозначно идентифицировать название углеводородного соединения и его объемное содержание в образце. Таким образом, после проведения хроматографического исследования образца сырой тяжелой нефти в дальнейшем можно будет выполнить анализ для копии этого же образца с введенным в него катализатором, определив тем самым объемное перераспределение углеводородов на основе сопоставления результатов двух лабораторных исследований.

Однако фиксация соответствующих числовых координатных значений пиков с хроматограммы происходит автоматически только в том случае, когда органическое соединение уже есть в базе данных хроматографа. Поскольку базы данных являются собственностью производителей хроматографов (устройств, выполняющих хроматографический анализ) и не могут быть расширены в лабораториях [6], а сами устройства ориентированы на фармацевтическую промышленность, в которой специалисты оперируют химически чистыми соединениями, то сотрудники нефтехимических лабораторий для большинства пиков вынуждены вручную выполнять фиксацию значений через аппроксимацию числовых данных с графика хроматограммы [7].

В работе [8] задача прогнозирования перераспределения углеводородных компонентов (при вводе катализатора) была автоматизирована моделью мультимодального трансформера. Ключом к решению стал анализ всего изображения хроматограммы путем загрузки изображения в модель трансформера. Несмотря на то что такой подход с высокой точностью решает поставленную задачу, стоит заметить, что процесс является очень ресурсозатратным, так как со всего изображения (рис. 1) вычленяется для анализа график, фиксирующий сигнал хроматографа, при этом большая площадь анализируемого изображения является белой заливкой, которая не несет в себе смысловой информации.

Исходя из этого, было решено провести дополнительное исследование, посвященное оптимизации процесса обработки входных данных с целью упрощения пайплайна разработанного прикладного решения на базе интеллектуальной модели.

Исходные данные

В качестве набора исходных данных выступают лабораторные хроматограммы тяжелой нефти, полученные с помощью хроматографов «Хроматэк» серии «Кристалл» на основе образцов из Ашальчинского, Мордово-Кармальского и Пионерского месторождений высоковязкой нефти. Территориально все обозначенные месторождения находятся в Республике Татарстан. Отметим, что данные из Ашальчинского и Мордово-Кармальского месторождения аналогичны тем, которые были использованы в работе [8].

Каждая хроматограмма является файлом изображения размером 1302×558 пикселей, представленном в одноканальном режиме, имеющем два уровня градации серого цвета: черный и белый. Размер выборки данных составил 220 484 хроматограмм, из которых 210 000 принадлежат к Ашальчинскому месторождению, 2310 – Мордово-Кармальскому, 8174 – Пионерскому. На отличающихся месторождениях присутствуют свои особенности нефтяного сырья, которые будут проявляться в изменяющемся представлении графика на хроматограмме [9].

Отметим, что на всех представленных графических данных существует общая черта, которая должна быть учтена в процессе построения метода для фиксации числовых значений координат: в правом верхнем углу каждой хроматограммы присутствует отметка хроматографа, которая регистрирует порядковый номер исследования (рис. 2).

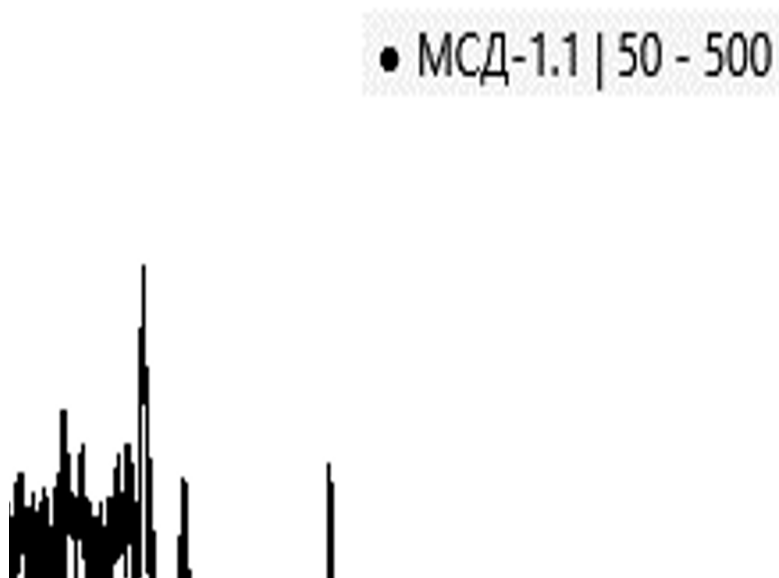


Рис. 2. Регистрационная отметка номера исследования, присваиваемая хроматографом каждой хроматограмме
Fig. 2. Identifies the run number assigned to each chromatogram by the chromatographer

Регистрационная отметка устанавливает однозначное соответствие между конкретным лабораторным исследованием пробы нефти и его хроматограммой, поэтому в рамках решения задачи фиксации числовых значений координат она не содержит в себе полезной информации.

Методы

Для решения задачи восстановления значений временного ряда (зависимости времени выхода углеводородного компонента от его энергии выхода) на основе его графического

представления была учтена важная особенность изображений, представленных исключительно в черном и белом цвете. Реализованный программный метод позволяет восстановить временной ряд благодаря последовательному решению нескольких операций над хроматограммой.

1. Ретуширование белыми пикселями прямоугольного участка регистрационной отметки хроматографа, находящегося в правом верхнем углу хроматограммы.

2. Ограничение зоны исследования метода только на область графика (рис. 3). Обоснованием этому является постоянство значений единиц измерения, находящихся на осях координат (так как характеристики всех проб нефти изменяются в одном диапазоне значений, а изменяется только форма графика).

3. Выполнение условного «разрезания» графика на вертикальные полосы шириной в 1 пиксель, начиная от начала системы координат.

4. Поиск интенсивности пикселя, соответствующей черному цвету на каждой из полученных вертикальных полос.

5. Фиксирование значения удаленности (в пикселях) обнаруженной «черной точки» относительно начала отсчета, которому соответствует верхний левый угол. Величина удаленности рассчитывается от верха графика, а затем инвертируется для получения значения по оси ординат. В том случае, если черных пикселей на одной вертикальной полосе оказывается несколько, то для «серии пикселей» фиксируется наибольшее значение (случай представлен на рис. 4) по оси ординат (по такому принципу функционирует и сам хроматограф, так как для отдельных компонентов нефти энергия выхода возрастает не мгновенно, а инерциально, поэтому из исследования графика может сложиться ошибочное мнение, что одному дискретному значению времени соответствует несколько значений энергии: истинным значением будет максимальная величина зарегистрированной энергии). Именно по этой причине величина удаленности рассчитывается от верха графика, так как такой способ позволяет существенно экономить временные затраты на фиксацию значения энергии в вертикальной полосе.

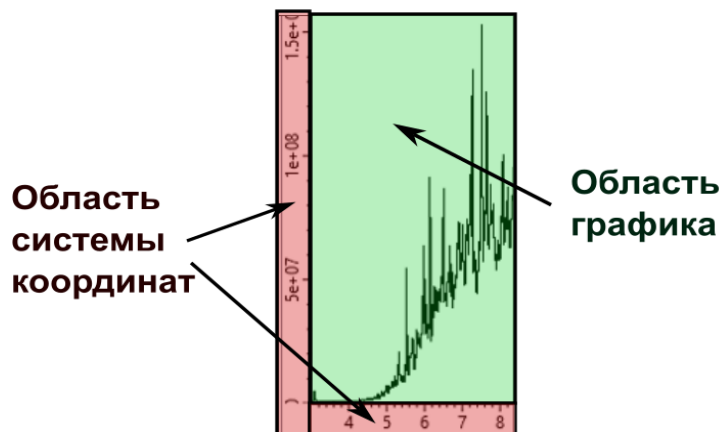


Рис. 3. Графическое представление области графика, исследуемой методом
Fig. 3. Graphical representation of the area of the graph under study by the method

Предлагаемый метод позволяет выполнить конвертацию данных хроматограмм из графической формы представления информации в числовую, которая теперь формализована в виде пар значений координат. Для расчета величины погрешности метода была использована метрика среднеквадратического отклонения, которая позволяет определить разницу между известными значениями координат углеводородных компонентов, зафиксированными хроматографом, и восстановленными при помощи метода.

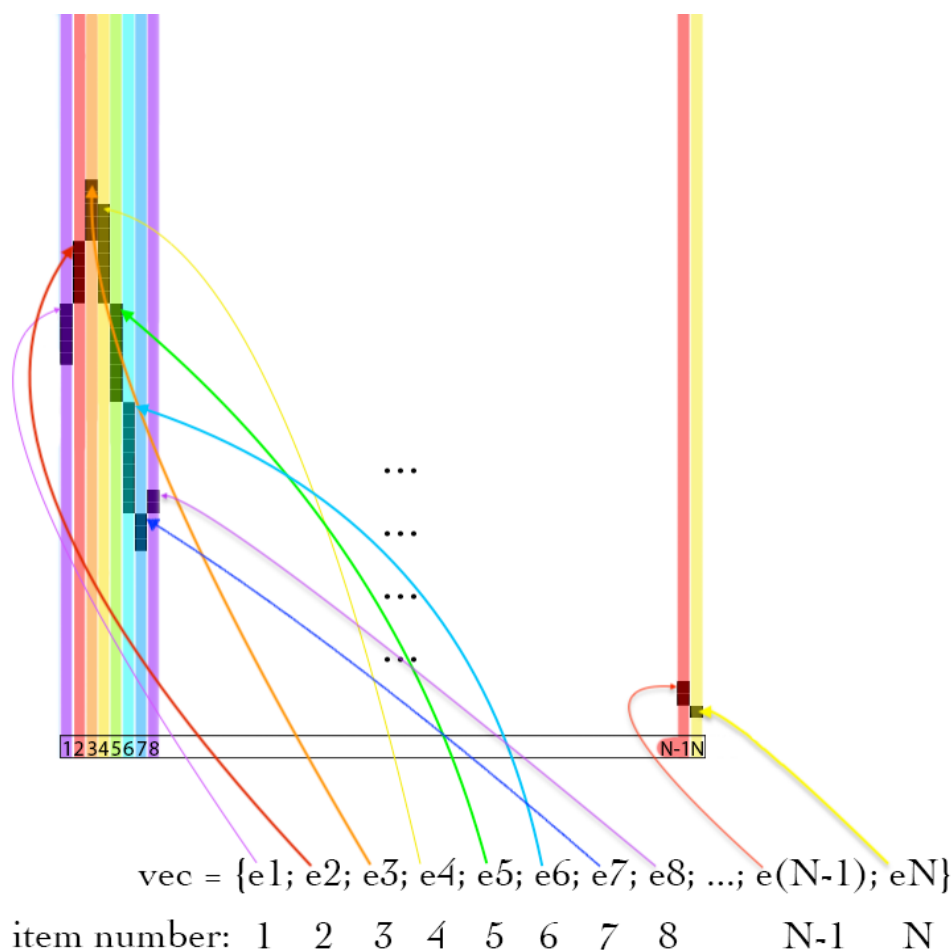


Рис. 4. Графическая схема составления вектора из пар координат восстановленных точек
 Fig. 4. Graphical scheme of vector construction from coordinate pairs of reconstructed points

Отметим, что восстановление числовых значений координат углеводородных компонентов на основе графического представления функции сигнала хроматографа является вторичной задачей. Основная задача, для решения которой был реализован метод, – упрощение пайплайна уже существующей интеллектуальной модели [8], поэтому для оценки эффективности от интеграции метода в конфигурацию модели были выделены следующие критерии:

- период времени, требуемый для обучения интеллектуальной модели;
- период времени, необходимый для предсказания выходных параметров углеводородных компонентов [8] (фактически является общим временем, требуемым для получения прогноза от модели);
- величина ошибки интеллектуальной модели при выполнении предсказаний в рамках решения прикладной задачи.

Поскольку исходные данные теперь представляются в формате числовых координат, возникла необходимость модификации пайплайна базовой модели [8].

1. Модель трансформера теперь не должна быть мультимодальной, так как отсутствует необходимость обработки разнородных данных (изображение и табличная информация); вся необходимая информация представляется в формате чисел.

2. Входные данные, представленные в виде пар координат (значения по двум осям для каждой восстановленной точки), можно преобразовать в вектор значений (рис. 4), в кото-

ром номер элемента будет характеризовать время выхода углеводородного компонента, а само значение – энергию. Так как вектор является обучающими данными для модели трансформера, было решено выполнить нормировку его значений, потому что в таком случае функция активации softmax позволит сформировать более качественную обобщающую способность модели глубокого обучения.

3. В модели трансформера остается только один энкодер, который используется в качестве входа и на который подаются векторизованные представления данных. Для повышения обобщающей способности предварительно все значения вектора (восстановленные координаты по оси ординат) нормализуются.

Результаты

В процессе применения метода для решения задачи конвертации данных из изображений в массив числовых значений координат были получены визуально идентичные графики: на рис. 5 представлено изображение оригинальной хроматограммы, а на рис. 6 – график, построенный на основе восстановленных числовых координат. Отметим, что график, изображенный на основе восстановленных значений, получен путем соединения соседних точек прямыми линиями.

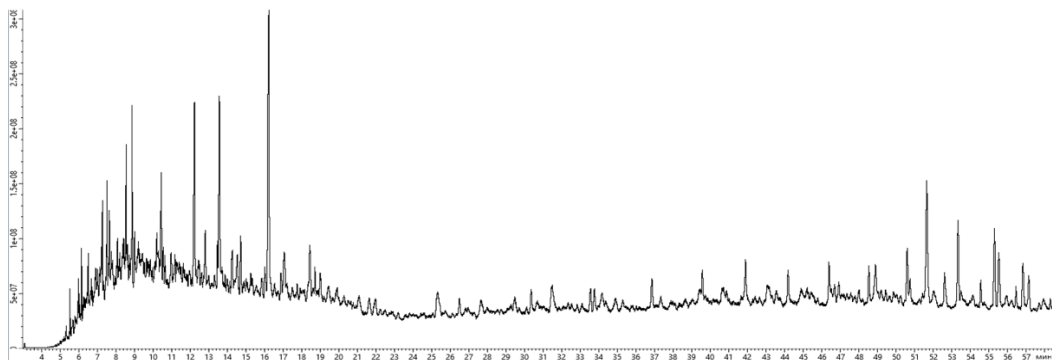


Рис. 5. Оригинальное изображение хроматограммы для пробы нефти
(с заретушированной отметкой хроматографа)

Fig. 5. Original image of the chromatogram for the oil sample
(with the mark of the chromatograph masked out)

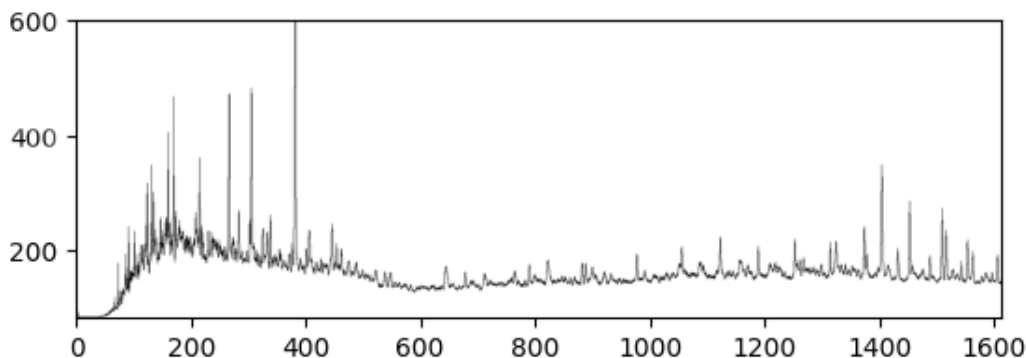


Рис. 6. График, построенный на основе восстановленных значений координат

Fig. 6. Graph based on the reconstructed coordinate values

В табл. 1 представлены числовые значения координат локальных максимумов точек хроматограммы, соответствующей рис. 5. Во втором столбце находятся значения, зарегистрированные хроматографом, в третьем – восстановленные с помощью метода (на рис. 6 значения по осям координат представлены в пикселях; для единства представления в табл. 1 шкала значений приведена к оригинальным единицам измерения).

Таблица 1

Сравнение полноты зарегистрированных значений координат точек
на хроматограмме (представлены только локальные максимумы)
на основе хроматографа и разработанного метода

Table 1

Comparison of the completeness of the recorded values of point coordinates
on the chromatogram (only local maxima are displayed)
based on the chromatograph and the developed method

Порядковый номер	Зафиксированные значения хроматографа		Восстановленные значения		Модуль разницы значений	
	Время выхода	Энергия выхода	Время выхода	Энергия выхода	Время выхода	Энергия выхода
1	2	3	4	5	6	7
1	3,072	0,0507	3,072	0,0495	0,0002	0,0012
2	4,770	0,0244	4,771	0,0254	0,0011	0,0010
3	4,878	0,0454	4,876	0,0460	0,0016	0,0006
4	4,996	0,0804	4,995	0,0810	0,0006	0,0006
5	5,318	0,2055	5,317	0,2045	0,0012	0,0010
6	5,526	0,5466	5,526	0,5480	0,0000	0,0014
7	5,691	0,2837	5,693	0,2827	0,0017	0,0010
8	5,798	0,3076	5,798	0,3084	0,0002	0,0008
9	5,949	0,4439	5,950	0,4440	0,0009	0,0001
10	5,995	0,6377	5,995	0,6382	0,0005	0,0005
11	6,028	0,5072	6,029	0,5072	0,0014	0,0000
12	6,146	0,9121	6,145	0,9120	0,0014	0,0001
13	6,264	0,7644	6,263	0,7632	0,0007	0,0012
14	6,508	0,7779	6,508	0,7789	0,0002	0,0010
15	6,669	0,8835	6,668	0,8832	0,0007	0,0003
16	6,920	0,7346	6,920	0,7349	0,0003	0,0003
17	7,020	1,2453	7,019	1,2463	0,0008	0,0010
18	7,210	1,3642	7,208	1,3634	0,0017	0,0008
19	7,275	0,9231	7,276	0,9238	0,0011	0,0007
20	7,518	1,5379	7,519	1,5382	0,0012	0,0003
21	7,644	1,3382	7,643	1,3396	0,0006	0,0014
22	8,081	1,0128	8,080	1,0136	0,0006	0,0008

Продолжение табл. 1

1	2	3	4	5	6	7
23	8,217	0,8746	8,216	0,8743	0,0008	0,0003
24	8,550	1,8678	8,549	1,8665	0,0007	0,0013
25	8,865	2,2132	8,867	2,2129	0,0016	0,0003
26	9,005	1,0749	9,004	1,0755	0,0010	0,0006
27	9,220	0,9872	9,221	0,9876	0,0012	0,0004
28	10,202	1,7535	10,202	1,7522	0,0004	0,0013
29	10,442	1,6286	10,441	1,6284	0,0007	0,0002
30	10,542	0,9479	10,541	0,9481	0,0009	0,0002
31	10,639	0,8754	10,637	0,8769	0,0015	0,0014
32	10,954	0,8985	10,956	0,8981	0,0017	0,0004
33	11,184	0,8721	11,185	0,8719	0,0008	0,0002
34	12,216	0,8236	12,216	0,8249	0,0001	0,0013
35	12,810	1,1026	12,810	1,1024	0,0005	0,0002
36	13,466	2,3214	13,466	2,3206	0,0001	0,0008
37	13,566	2,2675	13,565	2,2680	0,0009	0,0005
38	14,269	0,9342	14,267	0,9348	0,0017	0,0006
39	14,541	0,8779	14,539	0,8784	0,0015	0,0005
40	14,717	1,0573	14,718	1,0565	0,0011	0,0008
41	16,017	3,0878	16,017	3,0877	0,0004	0,0001
42	16,229	1,1961	16,229	1,1952	0,0002	0,0009
43	16,881	0,8943	16,882	0,8937	0,0011	0,0006
44	17,067	0,9077	17,067	0,9068	0,0001	0,0009
45	18,443	0,9723	18,442	0,9728	0,0009	0,0005
46	18,715	0,7817	18,715	0,7819	0,0004	0,0002
47	18,995	0,7022	18,997	0,7035	0,0017	0,0013
48	19,394	0,5717	19,394	0,5709	0,0002	0,0008
49	19,872	0,4985	19,872	0,4988	0,0002	0,0003
50	21,091	0,5179	21,089	0,5183	0,0017	0,0004
51	21,628	0,4871	21,628	0,4877	0,0004	0,0006
52	21,961	0,4722	21,961	0,4723	0,0003	0,0001
53	25,301	0,5411	25,301	0,5399	0,0005	0,0012
54	26,501	0,4688	26,502	0,4680	0,0005	0,0008
55	27,648	0,4524	27,649	0,4525	0,0015	0,0001
56	30,363	0,5069	30,363	0,5074	0,0000	0,0005
57	31,496	0,5128	31,495	0,5142	0,0013	0,0014
58	33,563	0,5781	33,562	0,5789	0,0012	0,0008
59	33,764	0,5649	33,766	0,5640	0,0016	0,0009
60	34,190	0,5039	34,190	0,5035	0,0002	0,0004
61	36,852	0,5123	36,851	0,5120	0,0009	0,0003

Окончание табл. 1

1	2	3	4	5	6	7
62	39,590	0,5266	39,592	0,5260	0,0016	0,0006
63	41,894	0,8642	41,894	0,8633	0,0004	0,0009
64	44,180	0,7923	44,179	0,7918	0,0009	0,0005
65	46,397	0,6574	46,398	0,6568	0,0005	0,0006
66	48,642	0,7848	48,643	0,7859	0,0015	0,0011
67	48,898	0,8236	48,898	0,8232	0,0005	0,0004
68	49,192	0,5041	49,193	0,5032	0,0011	0,0009
69	49,461	0,5744	49,462	0,5755	0,0010	0,0011
70	50,542	0,9589	50,541	0,9586	0,0006	0,0003
71	51,661	1,6237	51,661	1,6238	0,0003	0,0001
72	52,598	0,7642	52,597	0,7635	0,0007	0,0007
73	53,341	1,2418	53,342	1,2405	0,0006	0,0013
74	54,428	0,6598	54,427	0,6589	0,0014	0,0009
75	55,305	1,0829	55,304	1,0841	0,0006	0,0012
76	55,541	0,9412	55,540	0,9422	0,0014	0,0010
77	56,324	0,6344	56,324	0,6353	0,0004	0,0009
78	56,845	0,7829	56,846	0,7842	0,0012	0,0013
79	57,157	0,6744	57,156	0,6754	0,0007	0,0010
80	58,021	0,4721	58,022	0,4722	0,0014	0,0001
81	58,644	0,6871	58,645	0,6856	0,0006	0,0015

Из табл. 1 следует, что максимальная величина модуля разницы между истинными и восстановленными значениями составила 0,0017 для времени выхода и 0,0015 для энергии выхода – справедливо будет заметить, что эти значения погрешности остаются максимальными только для выборки данных, соответствующей набору известных «истинных» значений. Однако получить числовые значения координат для незарегистрированных хроматографом локальных максимумов можно только с помощью аппроксимации, поэтому величину среднеквадратического отклонения для всех значений на данном этапе вычислить не представляется возможным без привлечения стороннего метода аппроксимации.

В рамках предметной области нефтехимического анализа минимальным шагом времени выхода (параметра оси абсцисс) для перехода от одного химического соединения к другому является пороговое значение 0,025 [11]: погрешность разработанного метода в 1,47 раза меньше обозначенной величины, поэтому он может применяться для решения задачи конвертации данных изображения в формат числового представления координат.

Использование векторного представления числовых восстановленных данных позволило получить существенный прирост в скорости обучения (табл. 2) интеллектуальной модели. В качестве персонального компьютера, на базе которого производилось обучение модели и тестирование ее точности, был использован ноутбук Apple MacBook Pro 16" выпуска 2023 г. в своей максимальной комплектации [12] с установленной версией macOS Sonoma 14.2.1 (23C71).

Таблица 2

Значение критериев оценки эффективности применения метода
для решения прикладной задачи

Table 2

Significance of criteria for evaluating the effectiveness of a method
in solving an applied problem

Тип модели	Время обучения (мин.)	Время, требуемое для предсказания (мин.)	Максимальное значение величины средн. отклонения при получении прогноза	
			Время выхода	Энергия выхода
Базовая модель	1230	8,2	0,0940	0,0610
Применение метода	148	4,3	0,0017	0,0015

Из табл. 2 следует, что время, требуемое для обучения модели и получения от нее предсказаний, сократилось в 8,3 и 1,9 раза соответственно.

Заключение

В рамках проведенного исследования был разработан алгоритм (и метод на его основе) для восстановления значений функции сигнала хроматографа, представленной в графическом виде. Исследована величина погрешности фиксации числовых значений координат для данных с хроматограммы тяжелой нефти. Обоснована достаточная точность восстановления значений для прикладного применения разработанного метода в рамках решения прикладной задачи нефтяной области. Выполнена векторизация значений восстановленных координат и модифицирована базовая интеллектуальная модель для возможности обработки нового представления входной информации. Зафиксировано сокращение величины среднеквадратического отклонения минимум на 0,0923 (по критерию времени выхода) при выполнении прогноза перераспределения углеводородных компонентов (решение прикладной задачи) в случае применения метода, что свидетельствует о формировании лучшей обобщающей способности интеллектуальной модели. Сокращение времени предсказания в 1,9 раза доказало целесообразность предварительной обработки входных данных хроматограмм тяжелой нефти.

Список источников

1. **Шадрина А. В., Крец В. Г.** Основы нефтегазового дела. М.: Нац. открытый ун-т «ИНТУИТ», 2016. 214 с.
2. **Басарыгин Ю. М., Будников В. Ф., Булатов А. И.** Исследование факторов и реализация мер долговременной эксплуатации нефтяных и газовых скважин. М.: Просвещение-Юг, 2004. 242 с.
3. **Шевченко Д. В., Васильева Л. Х.** Математическое моделирование вытеснения тяжелых нефтей горячей водой в тонком пласте. Казань: Познание, 2013. 60 с.
4. **Хуснутдинов И. Ш., Копылов А. Ю., Гончарова И. Н.** Разработка и совершенствование сольвентных технологий переработки тяжелого органического сырья. Казань: КГТУ, 2009. 265 с.

5. **Хадавимогаддам Ф., Мищенко И. Т., Мостаджеран М.** Применение методов искусственного интеллекта в прогнозировании основных свойств нефти // Газовая промышленность. 2019. Вып. 12(794). С. 28–32.
6. **Азиева Р. Х.** Искусственный интеллект в добыче нефти и газа: возможности и сценарный прогноз // Проблемы экономики и управления нефтегазовым комплексом. 2022. Вып. 3(207). С. 38–46. DOI: 10.33285/1999-6942-2022-3(207)-38-46
7. **Байбаров Д. А.** Оценка продуктивности и экономической эффективности технологий искусственного интеллекта для автоматизации процессов разведки и добычи нефти и газа // XXI век: итоги прошлого и проблемы настоящего плюс. 2021. Т. 10, вып. 3. С. 100–105. DOI: 10.46548/21vek-2021-1055-0019
8. **Ломакин Н. И., Дженифер О. Ч., Голодова О. А., Сычева А. В., Кабина В. В.** AI-система «Персептрон2 для прогноза финансового результата деятельности предприятия нефтяной отрасли РФ // Фундаментальные исследования. 2019. Вып. 12-1. С. 98–103. DOI: 10.17513/ft.42629
9. **Овсеенко Г. А., Козелков О. В., Кашаев Р. С.** Использование нейронных сетей в механическом устройстве представительного отбора и анализа проб // Приборостроение и автоматизированный электропривод в топливно-энергетическом комплексе и жилищно-коммунальном хозяйстве: Мат. VII Национальной науч.-практ. конф. 2022. С. 92–96.
10. **Vaswani A., Shazeer N. et al.** Attention is all you need. In Proc. of the 31st Conference on Neural Information Processing Systems (NIPS). 2017. 11 p.
11. **Dosovitskiy A., Beyer L. et al.** An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proc. of the International Conference on Learning Representations (ICLR). 2021. 21 p. DOI: 10.48550/arXiv.2010.11929
12. **Arik S. Ö., Pfister T.** TabNet: Attentive Interpretable Tabular Learning // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. Vol. 35(8). P. 6679–6687. DOI: 10.1609/aaai.v35i8.16826

References

1. **Shadrina A. V., Krets V. G.** Fundamentals of Oil and Gas Engineering. Moscow, National Open University of INTUIT, 2016, 214 p. (in Russ.)
2. **Basarygin Y. M., Budnikov V. F., Bulatov A. I.** Research of factors and implementation of measures for long-term operation of oil and gas wells. Moscow, Prosveshchenie-Yug publ., 2004, 242 p. (in Russ.)
3. **Shevchenko D. V., Vasilieva L. H.** Mathematical modelling of displacement of heavy oils by hot water in a thin reservoir. Kazan, Poznanie Publishing House, 2013, 60 p. (in Russ.)
4. **Khusnutdinov I. Sh., Kopylov A. Yu., Goncharova I. N.** Development and improvement of solvent technologies for processing of heavy organic raw materials. Kazan, KSTU, 2009, 265 p. (in Russ.)
5. **Hadavimoghaddam F., Mishchenko I. T., Mostajeran M.** Application of artificial intelligence methods in predicting basic oil properties. *Gas Industry*, 2019, vol. 12(794), pp. 28–32 (in Russ.)
6. **Azieva R. H.** Artificial intelligence in oil and gas production: opportunities and scenario forecast. *Problems of economics and management of oil and gas complex*, 2022, vol. 3(207), pp. 38–46 (in Russ.). DOI: 10.33285/1999-6942-2022-3(207)-38-46
7. **Baybarov D. A.** Assessment of productivity and economic efficiency of artificial intelligence technologies for automation of oil and gas exploration and production processes. *XXI century: Resumes of the Past and Challenges of the Present plus*, 2021, vol. 10, iss. 3, pp. 100–105 (in Russ.). DOI: 10.46548/21vek-2021-1055-0019
8. **Lomakin N. I., Jennifer O. C., Golodova O. A., Sycheva A. V., Kabina V. V.** AI-system “Perseptron” for forecasting the financial result of the enterprise of the oil industry of the Rus-

- sian Federation. *Fundamental Research*, 2019, vol. 12-1, pp. 98–103 (in Russ.). DOI: 10.17513/fr.42629
9. **Ovseenko G. A., Kozelkov O. V., Kashaev R. S.** Use of neural networks in mechatronic device of representative sampling and analysis. Instrumentation and automated electric drive in fuel and energy complex and housing and communal services. *Proceedings of VII National Scientific and Practical Conference*, 2022, pp. 92–96. (in Russ.)
 10. **Vaswani A., Shazeer N. et al.** Attention is all you need. In *Proc. of the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017, 11 p.
 11. **Dosovitskiy A., Beyer L. et al.** An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2021, 21 p. DOI: 10.48550/arXiv.2010.11929
 12. **Arik S. Ö., Pfister T.** TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 35(8), pp. 6679–6687. DOI: 10.1609/aaai.v35i8.16826

Сведения об авторах

Пылов Петр Андреевич, аспирант Кузбасского государственного технического университета им. Т. Ф. Горбачева; старший разработчик высоконагруженных интеллектуальных систем на позиции Senior Computer Vision Engineer
WoS Researcher ID: JMQ-9360-2023

Майтак Роман Вячеславович, магистрант Кузбасского государственного технического университета им. Т. Ф. Горбачева; Middle+ NLP Data Scientist
WoS Researcher ID: JMQ-9434-2023

Копылов Даниил Евгеньевич, магистрант Иркутского государственного университета

Information about the Authors

Petr A. Pylov, PhD Student at the T.F. Gorbachev Kuzbass State Technical University. He combines his studies with his work as a Senior Computer Vision Engineer

Roman V. Maitak, Master Student at the T.F. Gorbachev Kuzbass State Technical University. He combines his studies with her work as a data scientist at Middle+ NLP

Daniil E. Kopylov, Master Student of Irkutsk State University

*Статья поступила в редакцию 26.01.2024;
одобрена после рецензирования 07.05.2024; принята к публикации 07.05.2024*
*The article was submitted 26.01.2024;
approved after reviewing 07.05.2024; accepted for publication 07.05.2024*