

УДК 544.021

БЫСТРАЯ МОЛЕКУЛЯРНАЯ РЕКОНСТРУКЦИЯ ХИМИЧЕСКОГО СОСТАВА СЛОЖНЫХ УГЛЕВОДОРОДНЫХ СМЕСЕЙ

© 2024 г. Н. А. Глазов*, А. Н. Загоруйко

Институт катализа СО РАН, Новосибирск, Россия

*e-mail: glazov@catalysis.ru

Поступила в редакцию 30.01.2024

После доработки 26.10.2024

Принята в печать 30.10.2024

Предложен новый эвристический подход для проведения стохастической молекулярной реконструкции значительно быстрее. За основу взят двухступенчатый метод, объединяющий стохастическую реконструкцию и реконструкцию максимизацией энтропии. В предложенном методе поиск оптимальных параметров распределений осуществляется при решении нескольких сравнительно простых оптимизационных задач. Предложенный метод позволил реконструировать состав образца вакуумного газоляя как минимум в 100 раз быстрее классического подхода с генетическими алгоритмами.

Ключевые слова: молекулярная реконструкция, стохастическая реконструкция, моделирование состава, тяжелые нефтяные фракции, максимизация энтропии

DOI: 10.31857/S0040357124060137 **EDN:** VGYEMI

ВВЕДЕНИЕ

Математическое моделирование является важной и неотъемлемой частью во многих сферах, включая нефтепереработку. Адекватные кинетические модели помогают в создании и совершенствовании процессов переработки. Создание кинетической модели для процессов с тяжелыми нефтяными фракциями является особенно сложной задачей из-за ограниченной информации. Очень часто данные сводятся к весьма общим свойствам смеси.

Методы псевдокомпонентов и агрегирования использовались и продолжают использоваться для решения этой задачи на протяжении долгого времени, а также входят во многие пакеты программного обеспечения [1]. Однако такие подходы не лишены недостатков, поскольку они сильно полагаются на предположение, что вещества, объединенные в единый групповой компонент, реагируют с одинаковой скоростью, что выполняется далеко не всегда.

Ряд альтернативных подходов был разработан за последние несколько десятилетий [2]. Многие из них часто называют методами “молекулярной реконструкции”. Центральная идея заключается в том, чтобы представить состав сырья с помощью хорошо изученных суб-

станций (индивидуальных веществ или смесей).

Стохастическая реконструкция [3] сама по себе или в комбинации [4] с методом реконструкции максимизацией энтропии [5] является часто используемым подходом для описания состава сложных углеводородных смесей. Одним из существенных недостатков этого метода является длительность вычислений. Для представления сырья необходимо найти оптимальный набор параметров распределений структурных атрибутов. Из-за стохастической природы метода набор методов оптимизации весьма ограничен. Наиболее часто используемыми подходами являются генетические алгоритмы [6] и имитация отжига [4, 7].

Попытки обойти [8] и решить [9] эту проблему были сделаны ранее, включая использование искусственных нейронных сетей [10]. В этой работе предложен новый эвристический подход.

МЕТОДЫ МОЛЕКУЛЯРНОЙ РЕКОНСТРУКЦИИ

Поскольку предложенная эвристика основывается на методах стохастической реконструкции и реконструкции максимизацией энтропии, имеет смысл кратко описать основные идеи этих методов.

Статистическая реконструкция. Несмотря на то, что количество индивидуальных компонентов в тяжелых нефтяных фракциях велико, зачастую возможно найти достаточно небольшой перечень структурных атрибутов (длина цепи, количество ароматических колец и т.д.), которых достаточно для описания большой доли веществ. С помощью метода Монте-Карло возможно “собирать” молекулы из структурных атрибутов.

Для реконструкции этим методом выбирается набор структурных атрибутов, позволяющий описать строение молекул, вызывающих интерес при моделировании кинетики. Каждому атрибуту присваивается определенный тип распределения. Дискретное и гамма-распределения наиболее популярны из-за их гибкости. Конкретные значения параметров распределений определяются в дальнейшем в ходе оптимизации. Изучаемые параметры выбираются зачастую случайно.

Определившись с набором распределений, можно генерировать наборы чисел из них, и интерпретируя эти значения как структурные атрибуты, получать “молекулы”. Сгенерировав достаточно много молекул (как минимум 5000-50000), можно воспринимать этот набор молекул как эквимолярную смесь (т.е. при генерации N молекул каждой из них присваивается мольная доля $1/N$).

Свойства получаемых молекул зачастую неизвестны достоверно, и для описания свойств индивидуальных веществ прибегают к методам групповых вкладов. Зная свойства каждой из молекул и какое количество раз она была сгенерирована, можно оценить свойства всей смеси.

Таким образом, имея набор параметров распределений, можно сгенерировать смесь и оценить любые ее свойства, которые возможно оценить, зная детальный состав. После чего возможно сопоставить измеренные свойства реального образца и расчетные свойства сгенерированного состава. Методы оптимизации позволяют найти такой набор параметров, при котором свойства сгенерированного состава будут неотличимы от реального.

Поскольку целевая функция статистична по своей природе, не удается эффективно использовать методы оптимизации, основывающиеся на производной, из-за чего чаще всего для такой задачи применяются генетические алгоритмы или метод имитации отжига.

Реконструкция максимизацией энтропии. Другой подход к реконструкции основывается на применении принципа максимума энтропии. Такой подход предполагает, что качественный состав образца уже известен (поэтому он

редко применяется в чистом виде для тяжелых нефтяных фракций). Решая задачу условной оптимизации, можно найти набор концентраций, удовлетворяющих наложенным условиям. В наиболее общей форме задача записывается следующим образом:

$$E(x, \sigma, v, \lambda) = -\sum_{i=1}^N x_i \ln x_i + \mu \left(1 - \sum_{i=1}^N x_i \right) - \frac{1}{2} \sum_{k=1}^K \frac{\left(f_k - \sum_{i=1}^N x_i f_{i,k} \right)^2}{\sigma_k^2} + \sum_{m=1}^M v_m \left(\sigma_m^2 - \sum_{i=1}^N x_i (f_m - f_{i,m})^2 \right) + \sum_{j=1}^J \lambda_j \left(f_j - \sum_{i=1}^N x_i f_{i,j} \right), \quad (1)$$

где λ , v и μ – множители Лагранжа, x_i – концентрация i -го компонента, f – свойства смеси, f_i – свойства i -го компонента. Первое слагаемое представляет собой информационную энтропию. Второе слагаемое обеспечивает выполнение материального баланса (сумма концентраций равна 1). Третье слагаемое вводит “не жесткие” ограничения на известные свойства, что помогает при реконструкции добиться того, чтобы расчетные значения смеси были близки к заданным с известным среднеквадратичным отклонением. Четвертое слагаемое позволяет задать дисперсию и сделать распределение концентраций более близким к нормальному. Пятое слагаемое позволяет задать “жесткие” ограничения на свойства, т.е. после реконструкции расчетное свойство будет в точности совпадать с заданным.

Представленная задача может показаться сложной, поскольку неизвестных не меньше, чем количество компонентов смеси. Если говорить о тяжелых нефтяных фракциях это число может превышать десятки тысяч. Однако если правила смешения линейны (т.е. свойства всей смеси представимы как взвешенная сумма свойств индивидуальных компонентов), то существует полуаналитическое решение, позволяющее заметно упростить задачу.

$$\frac{\partial E}{\partial x_i} = -\ln x_i - 1 - \mu + \sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \frac{\left(f_k - \sum_{i=1}^N x_i f_{i,k} \right)}{\sigma_k} - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j}. \quad (2)$$

Выражение в четвертом слагаемом принимают за еще одну переменную, и после этого довольно легко получить выражение для концентрации i -го компонента, приравняв производную нулю

$$\ln x_i = -1 - \mu + \sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \varepsilon_k - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j}, \quad (3)$$

$$x_i e^{1+\mu} = \exp \left(\sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \varepsilon_k - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j} \right). \quad (4)$$

Зная, что сумма всех концентраций равна единице, можно найти выражение для $\exp(1+\mu)$

$$e^{1+\mu} \sum_{i=1}^N x_i = \sum_{i=1}^N \exp \left(\sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \varepsilon_k - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j} \right). \quad (5)$$

Таким образом, финальное выражение для концентрации i -го компонента

$$x_i = \frac{\exp \left(\sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \varepsilon_k - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j} \right)}{\sum_{i=1}^N \exp \left(\sum_{k=1}^K \frac{f_{i,k}}{\sigma_k} \varepsilon_k - \sum_{m=1}^M v_m (f_m - f_{i,m})^2 - \sum_{j=1}^J \lambda_j f_{i,j} \right)}. \quad (6)$$

Если подставить такое выражение в исходную задачу, удастся полностью избавиться от концентраций.

$$E(\varepsilon, v, \lambda) = \ln Z + \sum_{k=1}^K \left(\frac{1}{2} \varepsilon_k^2 - \frac{f_i}{\sigma_k} \varepsilon_k \right) + \sum_{m=1}^M v_m \sigma_m^2 + \sum_{j=1}^J \lambda_j f_j, \quad (7)$$

где Z – это знаменатель из выражения для концентрации i -го компонента.

Такая задача является сравнительно простой, поскольку количество неизвестных в ней равно количеству накладываемых ограничений, и решается с применением распространенных методов оптимизации (например, методом сопряженных градиентов). Стоит еще раз подчеркнуть, что такое упрощение возможно исключительно из-за линейности ограничений. Любое усложнение правил смешения значительно усложнит поставленную задачу.

Стохастическая реконструкция с максимизацией энтропии. Этот метод заключается в последовательном использовании вышеописанных методов. Такая комбинация решает некоторые проблемы индивидуальных методов. Стохастическая реконструкция предоставляет качественный состав, необходимый для работы метода на основе максимизации энтропии, а он, в свою очередь, позволяет сколько угодно точно приближать расчетные свойства к экспериментально измеренным.

Стоит, однако, заметить, что в таком виде методы сохраняют свою индивидуальность. Скорость реконструкции во многом лимитируется скоростью стохастической реконструкции.

Предлагаемая эвристика. Как понятно из описания метода реконструкции максимизацией энтропии, он позволяет подстроить концентрации веществ таким образом, чтобы они удовлетворяли заданным ограничениям. В связи с этим возникает соблазн использовать этот метод как можно раньше.

Применение максимизации энтропии с плохим начальным приближением может привести к ситуации с преобладанием малой группы веществ в смеси из-за экспоненциальной зависимости от лагранжевых множителей (если не вводятся ограничения на среднеквадратичное отклонение). Чаще всего этого пытаются избежать. Кроме того, даже если начальное приближение не так далеко от решения, метод максимизации энтропии вызывает искажения в сравнительно гладких распределениях, получаемых в стохастической реконструкции, что зачастую невозможно адекватно интерпретировать.

Свойства полученного состава, даже если начальное приближение достаточно далеко от решения, обычно ближе к реальному образцу, чем состав со стадии стохастической реконструкции. Этот состав может быть использован для оценки параметров распределений со стадии стохастической реконструкции. Повторяя эту процедуру несколько раз, можно найти набор параметров распределений, который хорошо описывает экспериментальные данные.

Оценка параметров. При оценке параметров распределений важно учитывать дискретность используемых данных. Гамма-распределение и его вариации являются непрерывными распределениями, в то время как структурные атрибуты обычно дискретны. Поэтому генерированные случайные значения часто округляются до ближайших/меньших/больших целых. В связи с этим используемые распределения в строгом смысле уже не являются вариациями гамма-распределения, а значит, и привычные процедуры для оценки параметров не работают корректно.

Для корректной оценки параметров стоит применять методы оптимизации. Кумулятивные распределения позволяют вычислять вероятность попадания случайной величины в заданный интервал значений. Зная вероятности появления всех возможных значений для структурного атрибута как функции от параметров модели, можно сформулировать задачу оптимизации, например таким образом:

$$\min_{\mathbf{p}} \sum_{i=1}^M (g_i(\mathbf{p}) - w_i)^2, \quad (8)$$

где w_i — наблюдаемая частота появления атрибута, g_i — расчетная вероятность появления указанного значения как функция от параметра. Суммирование происходит по всем возможным значениям атрибута.

Возможные сложности в применении эвристики. Предложенная эвристика может давать сбой, если начальное приближение слишком далеко от решения, особенно когда большое количество условий используются единовременно на стадии максимизации энтропии.

Для решения этой проблемы можно перебрать другие случайные начальные приближения. При небольшом количестве ограничений это может быть самым быстрым решением.

Более надежным решением может служить постепенное добавление налагаемых ограничений. На каждой последующей итерации метода вводятся новые ограничения, что позволяет решать задачу максимизации энтропии, даже если

начальное приближение слишком далеко от решения. Другая альтернатива связана с применением “не жестких” ограничений с постепенным уменьшением среднеквадратичных отклонений.

Четвертый подход является самым кардинальным из предлагаемых и заключается в полном уходе от стохастичности. Такое изменение открывает доступ к применению более широкого набора методов оптимизации. При игнорировании расположения заместителей (эффект которого на свойства зачастую и так потерян из-за грубоści методов групповых вкладов) количество соединений, которые могут получаться на стадии стохастической реконструкции, зачастую ограничивается несколькими десятками тысяч уникальных соединений. Концентрациями в этом случае могут служить вероятности появления молекул. Такой подход позволяет однозначно определить функцию свойств реконструированной смеси от параметров модели, что позволяет оценивать производные и впоследствии находить хорошие начальные приближения для использования эвристики.

Такой подход в теории можно использовать и самостоятельно, что может быть полезно, если решаемая задача требует нелинейных правил смешения, которые не удается эффективно включить в задачу максимизации энтропии.

Стоит так же заметить, что метод максимизации энтропии сравнительно легко интегрировать с “нестохастической” процедурой.

$$\begin{aligned} E(x, \sigma, v, \lambda) = & -\sum_{i=1}^{N^*} n_i x_i \ln x_i + \mu \left(1 - \sum_{i=1}^{N^*} n_i x_i \right) - \\ & - \frac{1}{2} \sum_{k=1}^K \frac{\left(f_k - \sum_{i=1}^{N^*} n_i x_i f_{i,k} \right)^2}{\sigma^2} + \\ & + \sum_{m=1}^M v_m \left(\sigma_m^2 - \sum_{i=1}^{N^*} n_i x_i (f - f_i)^2 \right) + \\ & + \sum_{j=1}^J \lambda_j \left(f_j - \sum_{i=1}^{N^*} n_i x_i f_{i,j} \right). \end{aligned} \quad (9)$$

В выражении (1) для вычисления информационной энтропии учитывались отдельно все молекулы в базе данных. Однако, очевидно, что при росте N количество дубликатов в базе молекул будет значительным. Если учесть количество раз, когда каждая молекула появилась в базе данных (n_i), можно превратить суммирование

N слагаемых в суммирование по уникальным компонентам (N^*). Дальнейший вывод остается неизменным. При этом при устремлении N к бесконечности финальное выражение для концентрации имеет такую же форму, как и (6), за исключением того, что экспоненты доминируют на вероятность появления молекулы.

ПРИМЕР РАБОТЫ МЕТОДА

Продемонстрируем предложенную эвристику на примере реконструкции образца вакуумного газоля.

Основные сведения об используемой схеме распределений, во многом повторяющей схему из [11], приведены в табл. 1.

Стоит заметить, что распределения 5 и 11 имеют плавающий верхний предел допустимых значений. Это связано с тем, что количество заместителей не может быть больше суммарной длины всех заместителей. Это делает оценку параметров распределений более сложной. Параметры распределения 4 и 5 (10 и 11) находятся совместно при решении одной оптимизационной задачи.

Можно убедиться, что предложенная схема генерирует сравнительно небольшое количество соединений, если не учитывать взаимное расположение заместителей.

Распределение 2 создает 30 вариантов основной цепи парафинов. Помимо нормального строения, каждая из 30 основных цепей порождает 9 изопарафинов (4 варианта длины цепи, 3 варианта количества заместителей за вычетом 3 невозможных комбинаций). Суммарно 300 разных парафинов.

Распределение 6 порождает 6 возможных нафтеновых структур (из-за пренебрежения взаимным расположением). Каждая из таких структур может принимать одну из $36 \times 4 - 6 = 138$ вариантов заместителей. Суммарно 828 нафтеновых структур.

Распределение 7 дает 5 вариантов ароматического ядра. Распределение 9 утверждает это значение. Распределение 8 утраивает суммарное количество структур. И каждая из $5 \times 4 \times 3 = 60$ структур может иметь каждый из 138 вариантов заместителей, генерируя 8280 уникальных ароматических структур.

Суммарно распределения 1-11 генерируют 9408 возможных структур. Распределения 12 и 13 применяются к каждой из этих структур, порождая 3 новых структуры, доводя суммарное количество соединений до 37632 молекул.

Финальное количество молекул может показаться внушительным, но, если вспомнить что при применении Монте-Карло число генерируемых молекул часто берется порядка 50000, становится понятно, что количество уникальных соединений сравнительно мало.

Для оценки физических свойств индивидуальных соединений использовались два метода групповых вкладов: [6] для соединений с серой и азотом и [12] для углеводородов без гетероатомов.

Правила смешения были выбраны линейными, поскольку стадия максимизация энтропии в любом случае будет требовать линейных правил смешения.

Экспериментально измеренные свойства приведены в табл. 2. Данные из кривой имитированной дистилляции приведены на рис. 1.

Таблица 1. Схема распределений

№ п/п	Структурный атрибут	Цель	Значения	Распределение
1	Тип молекулы	все	Парафин, нафтен, ароматика	Дискретное
2	Длина цепи	парафины	10-40	Гамма
3	Изопарафин	парафины	Да, нет	Дискретное
4	Суммарная длина заместителей	изопараф.	1-4	Показательное
5	Число заместителей	изопараф.	1-min(3, длина)	Дискретное
6	Нафтеновые кольца	нафтены	1-6	Гамма
7	Ароматические кольца	ароматика	1-5	Гамма
8	Нафтеновые кольца	ароматика	0-2	Дискретное
9	Тип гетероцикла	ароматика	Нет, тиофен, пиррол, пиридин	Дискретное
10	Суммарная длина заместителей	ароматика, нафтены	1-36	Гамма
11	Количество заместителей	ароматика, нафтены	1-min(4, длина)	Дискретное
12	Амин	все	Да, нет	Дискретное
13	Тиол	все	Да, нет	Дискретное

Таблица 2. Свойства образца

Свойство	Метод	Значение
Плотность(d), кг/м ³	ASTM D7042-11a	879.1
Массовая доля серы ($\omega(S)$), ppm	ASTM D4294-10	35
Массовая доля азота ($\omega(N)$), ppm	ASTM D5762-10	24
Массовая доля ароматических углеводородов (Ar), масс. %	ASTM D7419-07	32.2

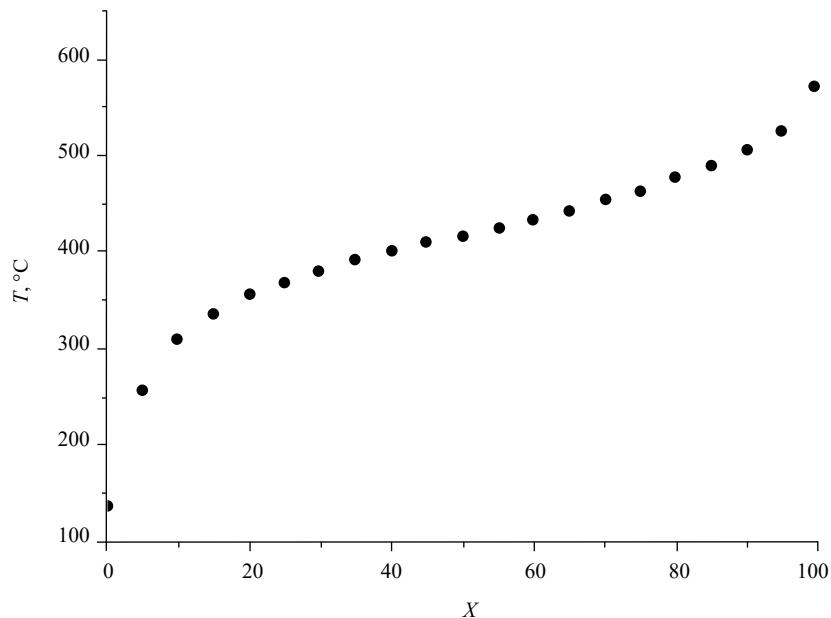


Рис. 1. Имитированная дистилляция (ASTM D2887-97a) образца.

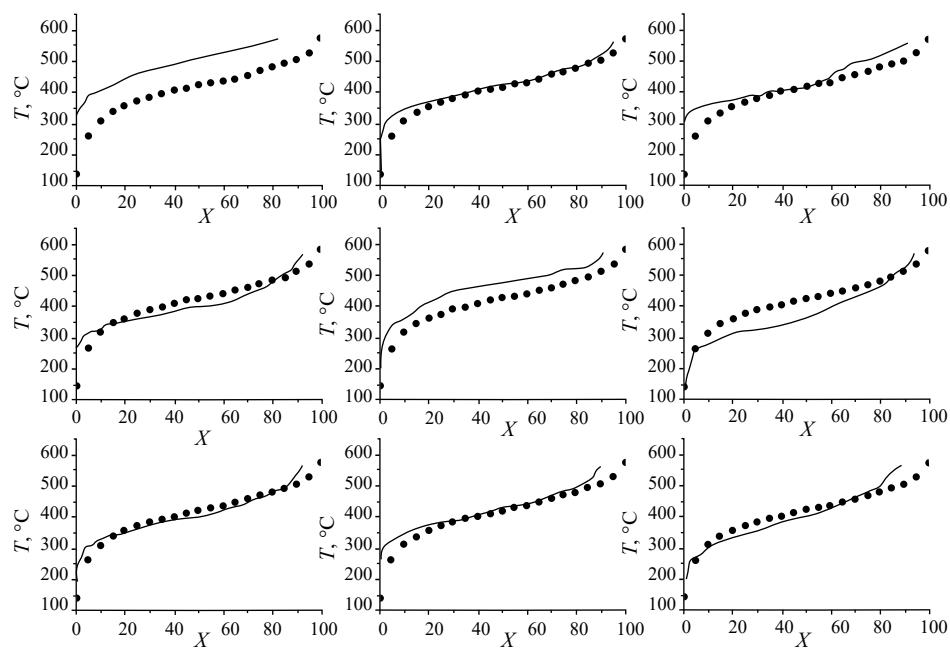


Рис. 2. Примеры получаемых кривых имитированной дистилляции для реконструированных составов на основе "нестохастического" метода. Точки — эксперимент, линия — расчет. Каждый график соответствует разному начальному приближению.

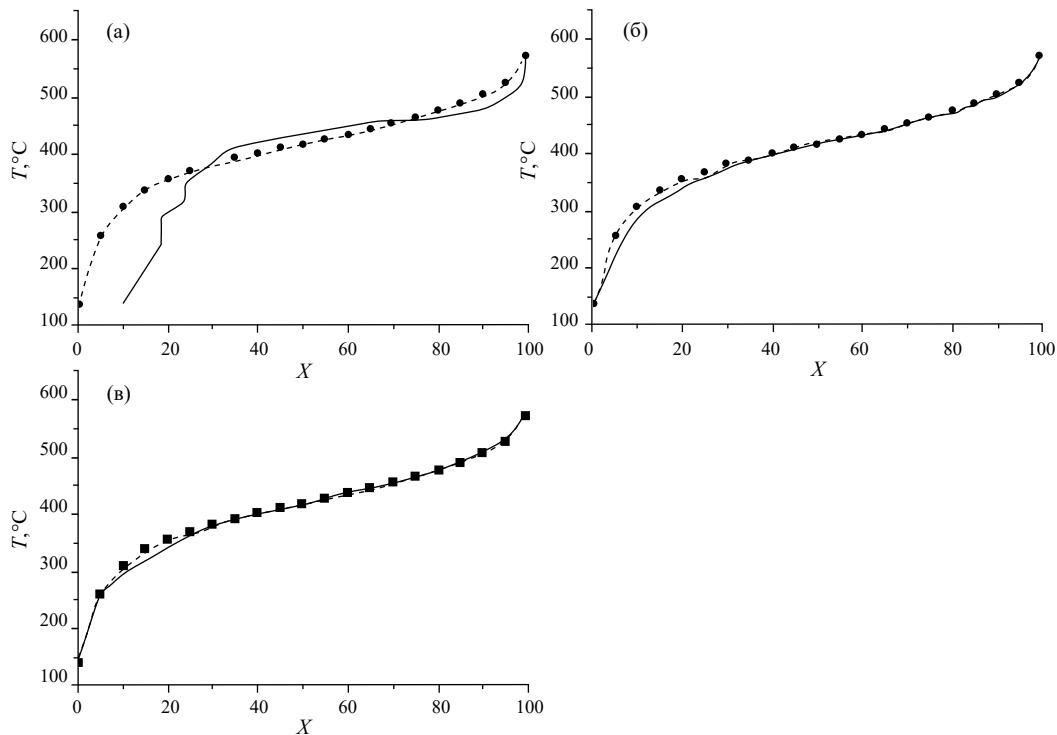


Рис. 3. Расчетные кривые имитированной дистилляции. а – исходное начальное приближение, б – после первой итерации, в – после пятой итерации. Точки – эксперимент, линия – расчет с помощью стохастической реконструкции, пунктир – результат расчета после максимизации энтропии.

Используя “нестохастический” вариант молекулярной реконструкции, удается получить начальные приближения, пригодные для применения предлагаемой эвристики за не более чем 2 итерации (порядка 10 секунд на домашнем компьютере с процессором AMD Ryzen 5 3600 6-Core Processor 3,59 ГГц) метода Ньютона. На рис. 2 можно видеть пример 9 начальных приближений, полученных таким способом.

На рис. 3 демонстрируется применение эвристики. Значительных изменений после 5-й итерации не наблюдается.

Можно видеть, что полученные кривые довольно хорошо описывают имеющиеся данные. Остальные свойства (плотность, содержание серы, азота и ароматических соединений) после максимизации энтропии фактически совпадают с экспериментально измеренными.

Реконструированная смесь на 32.2 мас. % состоит из ароматических соединений, доля нафтенов составила 30.4 мас. %, все остальное (37.4%) приходится на парафины. Стоит отметить, что экспериментальные данные о содержании нафтенов и парафинов отсутствуют, в связи с чем мы не можем говорить о том, насколько хорошо модель может экстраполировать за пределы имеющихся данных. В данном

случае представляется, что остальных данных недостаточно, чтобы однозначно определить содержание парафинов и нафтенов в смеси с большой точностью.

Суммарное время работы программы не превысило 1 минуты на используемом оборудовании, тогда как результаты классической процедуры гораздо менее стабильны по времени (в силу сильной зависимости от начального приближения), но даже в лучшем времени обычно требуется несколько часов, а в худшем оптимизация может не закончиться и за неделю.

И классический, и предлагаемый вариант были изначально реализованы автором статьи в виде программ на языке Python. Предлагаемый вариант затем был переписан на языке Julia для еще большего повышения производительности. Утверждение о как минимум 100-кратном ускорении было сделано без учета ускорения, вызванного использованием другого языка.

ЗАКЛЮЧЕНИЕ

В работе предложена эвристика, позволяющая проводить молекулярную реконструкцию составов сложных углеводородных смесей в сжатые сроки.

Предлагаемый метод позволяет описывать составы образцов с точки зрения реально существующих молекул, а не псевдокомпонентов. Он способен принимать любые аналитические данные, которые хоть в какой-то степени возможно предсказать на основании детального химического состава.

Предлагаемый метод может быть использован для построения детальных кинетических моделей.

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания для Института катализа им. Г.К. Борескова (проект FWUR-2024-0037).

ОБОЗНАЧЕНИЯ

Ar	массовая доля ароматических углеводородов в смеси, масс. %;
E	информационная энтропия Шеннона;
M	количество возможных исходов у распределения;
N	количество компонентов (с повторениями) в базе молекул;
N*	количество уникальных компонентов в базе молекул;
X	доля отгона, масс. %;
Z	знаменатель из выражения (6);
d	плотность, кг/м ³ ;
f _j	свойство смеси в j-м ограничении, разные;
f _{ij}	свойство i-го индивидуального компонента, участвующего в расчете j-го свойства смеси, разные;
g _i	вероятность появления i-го исхода как функция параметров модели;
n _i	количество раз i-й компонент оказался в базе молекул;
p	вектор параметров распределения
w _i	наблюданная взвешенная частота появления i-го исхода;
x	мольная доля;
ε _k	отношение разницы между экспериментально измеренным значением и экспериментом и среднеквадратичным отклонением в “нежестких” условиях;
λ	лагранжев множитель, отвечающий за среднее значение;
μ	лагранжев множитель, отвечающий за материальный баланс;
ν	лагранжев множитель, отвечающий за дисперсию;
σ _k	среднеквадратичные отклонения, разные;
ω	массовая доля, ppm.

ИНДЕКСЫ

i	номер компонента;
j	номер “жесткого” ограничения на среднее значение;
m	номер “жесткого” ограничения на дисперсию;
k	номер “не жесткого” ограничения на среднее.

СПИСОК ЛИТЕРАТУРЫ

1. *De Oliveria L., Hudebine D., Guillaume D.* A review of kinetic modeling methodologies for Complex Process- es // Oil Gas Sci. Technol. 2016. V. 71. P. 45.
2. *Ren Y., Liao Z., Sun J.* Molecular reconstruction: Recent progress toward composition modeling of petroleum fractions // J. Chem. Eng. 2019. P. 761.
3. *Neurock M., Nigam A., Trauth D.* Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms // Chem. Eng. Sci. 1994. V. 49. № 24A. P. 4153.
4. *Hudebine D., Verstrate J.* Molecular reconstruction of LCO gasoils from overall petroleum analyses// Chem. Eng. Sci. 2004. V. 59. P. 4755.
5. *Hudebine D., Verstraete J.* Reconstruction of Petroleum Feedstocks by Entropy Maximization. Application to FCC Gasolines // Oil Gas Sci. Technol. 2011. V. 66. P. 437.
6. *De Oliveria L., Vazquez Trujillo A., Verstrate J.* Molecular Reconstruction of Petroleum Fraction: Application to Vacuum Residues from Different Origins // Energy & Fuels. 2013. V. 27. P. 3622.
7. *Alvarez-Majmutov A., Chen J., Gieleciak R.* Molecular-Level Modeling and Simulation of Vacuum Gas Oil Hydrocracking // Energy & Fuels. 2016. V. 30. P. 138.
8. *Zhao G., Yang M., Du W.* A stochastic reconstruction strategy based on a stratified library of structural descriptors and its application in the molecular reconstruction of naphtha // Chin. J. Chem. Eng. 2022. V. 51. P. 153.
9. *Dantas T., Noriler D., Huziwar K.* A multi-populating particle swarm optimization algorithm with adaptive patterns of movement for the stochastic reconstruction of petroleum fractions // Comput. Chem. Eng. 2023. P. 174.
10. *Deniz C.U., Yasar M., Klein M.T.* Stochastic Reconstruction of Complex Heavy Oil Molecules using an Artificial Neural Network// Energy & Fuels 2017. V. 31. № 11. P. 11932.
11. *Alvarez-Majmutov A., Gieleciak R., Chen Jinwen.* Deriving the Molecular Composition of Vacuum Distillates by Integrating Statistical Modeling and Detailed Hydrocarbon Characterization// Energy & Fuels. 2015. V. 29. № 12. P. 7931.
12. *Skander N., Chitour C.E.* A new Group-contribution method for the estimation of Physical properties of Hydrocarbons// Oil Gas Sci. Technol. 2002. V. 57. № 4. P. 369.