

12. Агулова Л. П. Принципы адаптации биологических систем к космогеофизическим факторам // Биофизика. 1998. Т. 43. Вып. 4. С. 561—564.

13. Леднев В. В. Биоэффекты слабых комбинированных, постоянных и переменных магнитных полей // Биофизика. 1996. Т. 41. Вып. 1. С. 224.

14. Агаджанян Н. А. и др. Медико-биологические эффекты геомагнитных возмущений. М.: ИЗМИРАН, 2001.

15. Холодов Ю. А. Мозг в электромагнитных полях. М.: Наука, 1982.

16. Холодов Ю. А., Лебедева Н. Н. Реакции нервной системы человека на электромагнитные поля. М.: Наука, 1992.

17. Макеев В. Б. и др. Физиологически активные инфранизкочастотные магнитные поля // Электромагнитные поля в биосфере. М.: Наука, 1984. Т. 2. С. 62—72.

18. Гриднев В. И. и др. Влияние внешних периодических стимулов на вариабельность сердечного ритма у здоровых лиц и у пациентов с ишемической болезнью сердца // Физиология человека. 2006. Т. 32. № 5. С. 74—83.

Дата принятия 11.02.2014 г.

ФИЗИКО-ХИМИЧЕСКИЕ ИЗМЕРЕНИЯ

543.08

Повышение точности количественного хроматографического анализа с использованием нейронных сетей

Т. З. ХАБУРЗАНИЯ*, А. В. МЕШКОВ **

* С.-Петербургский государственный политехнический университет, С.-Петербург, Россия, e-mail: timur.kh@iit.icc.spbstu.ru

** Всероссийский научно-исследовательский институт метрологии им. Д. И. Менделеева, С.-Петербург, Россия, e-mail: amesh@b10.vniim.ru

Предложен метод повышения точности хроматографического анализа путем разделения пересекающихся пиков хроматограмм с помощью аппарата нейронных сетей, сопряженного с регуляризирующим методом решения некорректных задач. Приведены примеры кратного снижения погрешности анализа.

Ключевые слова: перекрывающиеся хроматографические пики, нейронные сети, регуляризация.

A method of increasing the accuracy of chromatographic analysis by separating of overlapping peaks of chromatograms at application of neural networks combined with a regularization method for incorrect problems solution is suggested. The examples of multiple reduction of the analysis error are presented.

Key words: overlapping chromatographic peaks, neural networks, regularization.

Точность результатов количественного химического анализа сложных смесей органических соединений, таких как нефтепродукты и природный газ, существенно снижается при пересечении близких пиков их компонентов. Такие пересечения происходят вследствие недостаточной разрешающей способности применяемых хроматографических колонок и детекторов. Поскольку зарегистрированная хроматограмма точно так же, как и спектрограмма, представляет свертку аппаратной функции прибора с реальной характеристикой вещества, Рэлей еще в конце XIX в. предположил, что разделения пересекающихся пиков можно достичь решением интегрального уравнения со сверточным ядром относительно свертываемой функции. Это интегральное уравнение является уравнением Фредгольма, решение которого — некорректно поставленная задача. Общий подход к решению подобных задач был предложен лишь в конце XX в.

А. Н. Тихоновым, после чего появилась возможность добиться желаемого результата [1, 2].

Аппаратная функция хроматографа априори неизвестна и изменяется в процессе регистрации хроматограммы с увеличением времени ее выхода, значительно осложняя решение задачи. Можно сделать предположение лишь о виде функции, описывающей хроматограмму и аппаратную функцию. По мнению большинства специалистов, эти функции имеют вид экспоненты с неизвестными параметрами [3]. Однако некоторые специалисты полагают, что они описываются кривыми Лоренца или Коши. В настоящей работе будем придерживаться мнения большинства.

Математически это означает, что если аппаратная функция $k(t) = Be^{-\beta^2 t^2}$, то зарегистрированная хроматограмма $f(t)$ — свертка:

$$f(t) = \sum_{i=1}^n \int_{-\infty}^{\infty} B e^{-\beta^2(t-z)^2} \varphi_i(z-t_i) dz, \quad (1)$$

где $\varphi_i(z-t_i)$ — базисные функции, описывающие хроматографический пик i -го компонента; B — амплитуда аппаратной функции (в форме гауссианы); β — ширина аппаратной функции; t_i — центры базисных функций φ_i ; t , z — переменные времени элюирования компонентов.

Для сохранения неизменности площади пиков, пропорциональной массовой доле компонентов исследуемого вещества, должно выполняться условие

$$\int_{-\infty}^{\infty} B e^{-\beta^2 t^2} dt = 1. \text{ Оно}$$

выполняется, если $B = \beta / \sqrt{\pi}$.

В качестве базисных функций, описывающих разделяемые пики, выберем гауссианы вида

$$\varphi_i(z-t_i) = C_i e^{-c_i^2(z-t_i)^2},$$

где C_i , c_i — амплитуда и характеристика ширины пика, подвергшегося расширению; t_i — время удерживания i -го компонента (время от начала анализа до выхода максимума пика из колонки).

Как предположил Рэлей, разделение пиков можно осуществить решением (1) относительно параметров базисных функций C_i , c_i , а также времен удерживания t_i . Попытки прямого решения этого уравнения путем регуляризованного обратного преобразования Фурье для некоторых частных случаев были вполне удачными, но в других ситуациях они давали неудовлетворительные результаты [4, 5].

С целью расширения возможностей разделения пересекающихся пиков авторы применили нейросетевые технологии, обобщающие как полезную информацию, так и шум и искажения, что в итоге позволило получить устойчивое решение обратной задачи при неизвестной аппаратной функции и в присутствии влияющих факторов.

Одно из первых применений нейронных сетей в хроматографии для разделения пиков описано в 1996 г. [6]. В 2001 г.

было предложено использовать радиально-базисную нейронную сеть и генетический алгоритм для определения количества пиков и оптимизации их параметров — времени выхода, ширины и амплитуды [7]. Применительно к методу капиллярного электрофореза в 2003 г. хроматографические пики были разделены с помощью многослойного персептрона, обучение которого велось алгоритмом обратного распространения ошибки в сочетании с методом главных компонентов [8]. В 2005 г. Д. В. Медянцева использовал генетический алгоритм для разделения пиков на этапе количественного анализа [9]. Несмотря на достигнутые успехи в конкретных случаях, в этих работах не учитывалась некорректность решаемой задачи.

В 2003 г. А. А. Севастьянов предпринял попытку разделить спектрометрические пики методом статистической регуляризации с применением байесовского подхода, требующего обширной априорной информации, отсутствующей в хроматографии [10].

Для получения физически обоснованного решения математически некорректной задачи разделения пересекающихся хроматографических пиков авторы объединили нейросетевые технологии с регуляризирующими алгоритмами. Нейросетевую реализацию метода регуляризации А. Н. Тихонова применили в следующих вариантах:

поиск решения задачи разделения пересекающихся хроматографических пиков в пространстве гладких функций, образованных семейством гауссиан, гладкость которых целиком определяется их шириной;

использование радиально-базисных нейронных сетей с целью адекватного интегрирования неразделенных хроматографических пиков и представление решения в виде линейной комбинации гауссиан [11, 12];

формализация и использование естественных физических ограничений, которым должно удовлетворять решение задачи [12].

Ниже описан эксперимент и нейросетевая регуляризирующая обработка его результатов.

Исходные данные. Всероссийский научно-исследовательский институт метрологии им. Д. И. Менделеева (ВНИИМ) предоставил для исследования серию из трех хроматограмм трех одинаковых проб образца нестабильного газового конденсата с Астраханского месторождения. Хроматограммы получены в одинаковых условиях на газовом хроматографе «Хроматэк Кристалл 5000.2» с кварцевой капиллярной колонкой длиной 10 м, диаметром 0,53 мм, толщиной полиметилсилоксановой фазы 2,65 мкм. Температура термостата колонки в диапазоне 0,20—160 °С изменялась со скоростью 15 °С/мин, а в диапазоне 160—250 °С — со скоростью 20 °С/мин. Хроматограф оснастили пламенно-ионизационным детектором (ПИД) температурой 250 °С; расход газа носителя (He) составил 5 мл/мин. Для анализа выбрали фрагмент хроматограмм между 2,5 и 3 мин, содержащий три не полностью разделенных пика (рис. 1).

Идентификация неразделенных пиков. Первичную идентификацию и разметку пиков выполнили в автоматическом режиме с помощью программного обеспечения (ПО) «Хроматэк Аналитик 2.6» с применением типовых хроматографических методик — перпендикуляра или тангенты, описанных в нормативных документах и реализованных в ПО производителей хроматографического оборудования.

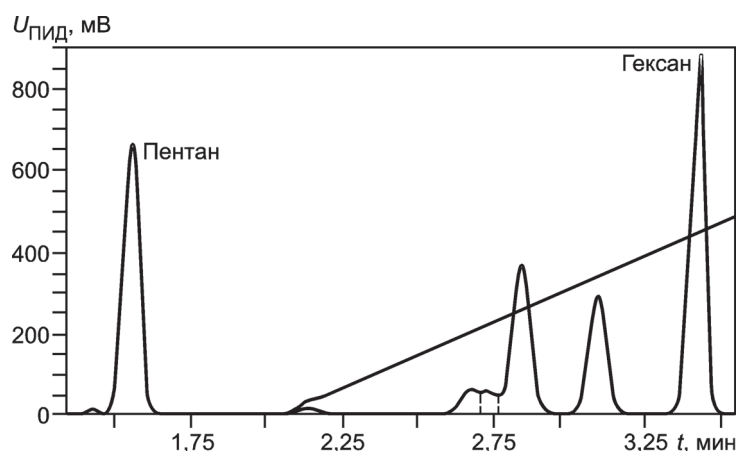


Рис. 1. Фрагмент исследуемой хроматограммы НГК: наклонная линия характеризует процесс изменения температуры в колонке

Исходя из априорной информации об анализируемой пробе эксперты ВНИИМ по базе данных NIST (США) индексов удерживания Ковача между опорными пиками пентана и гексана (значения индексов 500 и 600, соответственно) идентифицировали не полностью разделенные циклопентан (566,80), 2,3-диметилбутан (569,24) и 2-метилпентан (573,7). Первые два обладают весьма схожими физико-химическими свойствами, и их пики требуют разделения.

Определение границ пиков и разметку нулевой линии провели таким же образом с учетом параметров интегрирования, которое выполнено методом опускания перпендикуляра через так называемые «долины», т. е. через минимумы хроматограммы, находящиеся между предполагаемыми пиками. В качестве площадей неразделенных пиков принимали площади фигур между проведенными перпендикулярами. Результаты расчета площадей по трем исследуемым фрагментам хроматограмм приведены в сравнительных таблицах 1—3.

Разработанное ПО для обработки хроматограмм реализовано в среде Wolfram Mathematica семью алгоритмами: параметризации и обучения нейронной RBF-сети, аппроксимации и компенсации нулевой линии, вычисления характеристик восстановленных пиков и визуализации результатов в зависимости от ширины моделируемой аппаратной функции хроматографа, расчета площадей и массовых долей компонентов обрабатываемой хроматограммы [12].

Выбор параметров нейросети. На первом этапе выдвинули предположение о наличии пяти пиков в исследуемой группе. Поэтому в соответствии с технологией обучения нейронной RBF-сети в анализе использовали пять нейронов.

Количество итераций подбора параметров нейронов (положение максимумов пиков, их амплитуды и ширины) задавали равным 2000, так как эмпирическим путем установлено, что в данной реализации для числа дискретных отсчетов порядка 100—200 процедура разделения пиков занимает непродолжительное время — от одной до нескольких минут — и достаточна для суждения о нулевой линии сложной формы.

Таблица 1

Сравнительные результаты обработки первой хроматограммы

Идеально разделенные пики			Хроматограмма					
			зарегистрированная			восстановленная		
Площадь, мВ · мин								
1-й	2-й	3-й	1-й	2-й	3-й	1-й	2-й	3-й
619,83	619,83	619,83	791,12	642,09	5029,88	641,21	911,65	4908,03
Массовое содержание компонентов, %								
1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т
9,59	14,52	75,88	12,24	9,93	77,82	9,92	14,11	75,97
Относительная погрешность, %								
—	—	—	27,64	–31,58	2,60	3,45	–2,85	0,11

Таблица 2

Сравнительные результаты обработки второй хроматограммы

Идеально разделенные пики			Хроматограмма					
			зарегистрированная			восстановленная		
Площадь, мВ · мин								
1-й	2-й	3-й	1-й	2-й	3-й	1-й	2-й	3-й
671,68	671,68	671,68	769,64	635,12	4937,06	677,32	829,07	4835,23
Массовое содержание компонентов, %								
1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т
10,59	13,19	76,22	12,14	10,01	77,85	10,68	13,07	76,25
Относительная погрешность, %								
—	—	—	14,58	−24,07	2,14	0,84	−0,88	0,04

В соответствии с общим принципом регуляризации, разработанным А. Н. Тихоновым, решение нашли в пространстве гладких функций путем последовательного расширения аппаратной функции, поскольку она априори неизвестна. При этом ее подбор начинали с самой узкой, так как при выборе аппаратной функции в виде δ -функции (т. е. экстремально узкой) решением (1) будет зарегистрированная хроматограмма. Постепенно увеличивали ширину аппаратной функции и, тем самым, ухудшали гладкость решения, но останавливались в пространстве гладких функций, согласуясь с интуитивными представлениями об удовлетворительном разделении пиков, не позволяя получаемому решению пересекать горизонтальную координатную ось. С некоторого момента решение уравнения свертки может потерять устойчивость, ведь незначительное увеличение ширины аппаратной функции приводит к резкому росту амплитуды пика. При этом в решении могут появиться такие пики, которые окажутся ложными, если время их вы-

Таблица 3

Сравнительные результаты обработки третьей хроматограммы

Идеально разделенные пики			Хроматограмма					
			зарегистрированная			восстановленная		
Площадь, мВ · мин								
1-й	2-й	3-й	1-й	2-й	3-й	1-й	2-й	3-й
605,94	605,94	605,94	782,48	636,72	4993,55	633,71	903,73	4875,43
Массовое содержание компонентов, %								
1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т	1 к-т	2 к-т	3 к-т
9,45	14,63	75,92	12,20	9,93	77,87	9,88	14,09	76,03
Относительная погрешность, %								
—	—	—	29,13	−32,12	2,56	4,58	−3,65	0,13

хода не будет отвечать химическому компоненту, соответствующему стандартным индексам удерживания Ковача.

Получили следующие параметры ширины пиков, т. е. значения безразмерных параметров a_i в уравнении (1): 0,58; 0,52; 0,59; 2,81; 5,20. Три пика примерно одинаковой ширины (0,58; 0,52; 0,59) отнесли к действительным хроматографическим. Ширина двух других оказалась настолько незначительной, что их можно объяснить шумом. Кроме того, если бы зарегистрированная хроматограмма содержала нулевую линию сложной формы, то в решении оказался бы ложный пик очень большой ширины. Полученные результаты свидетельствуют об отсутствии такой линии, и можно заключить, что нулевая линия не сложнее, чем прямая, удаляемая вычитанием.

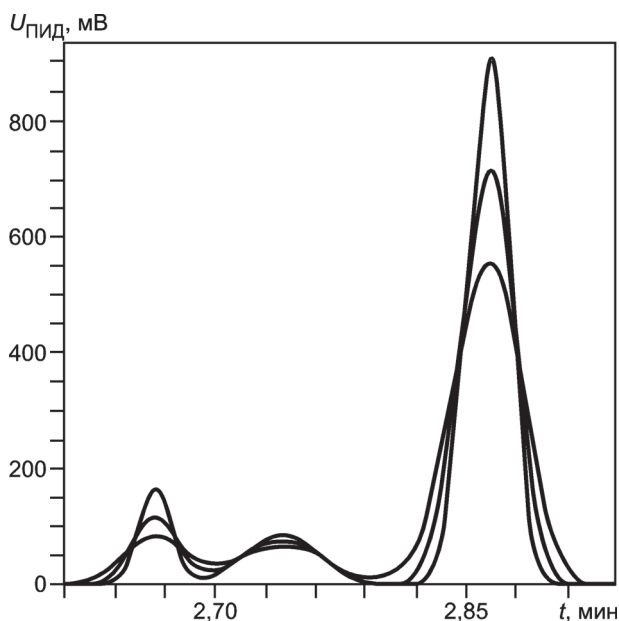


Рис. 2. Результаты разделения пиков на участке хроматограммы при разных значениях ширины аппаратной функции

представлены на рис. 2. По окончании процесса разделения определили времена удерживания и площади разделенных пиков.

Для сравнения погрешностей определения площадей пиков при типичном применении метода перпендикуляра к неразделенным и разделенным пикам в соответствии с [13] применяли численное моделирование зарегистрированной хроматограммы. Это моделирование выполняли путем свертки идеально разделенных пиков, площади которых и продолжительность удерживания равны площадям и продолжительности удерживания разделенных пиков. При этом использовали аппаратную функцию, на которой процесс ее расширения остановился. Путем незначительной подгонки ширины аппаратной функции, сворачивания идеальных смоделированных пиков и их суммирования добивались удовлетворительного соответствия результата свертки зарегистрированной хроматограмме. Так как интеграл от аппаратной функции равен единице, то свертка с ней не изменит площадь пиков.

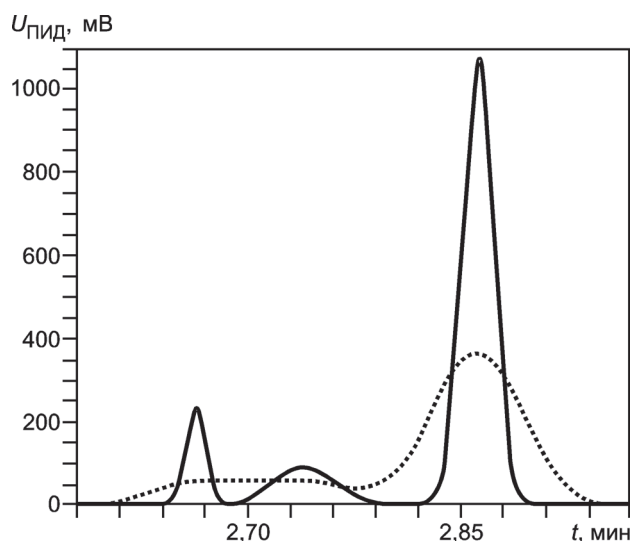


Рис. 3. Результаты моделирования участка хроматограммы

Таким образом, определили следующие результаты сверки: максимальная погрешность отклонения смоделированной хроматограммы от зарегистрированной 1,07 мВ, среднее квадратическое отклонение 0,39 мВ. Модели пиков и зарегистрированной хроматограммы представлены на рис. 3.

Каждая из трех хроматограмм независимо друг от друга была обработана следующим образом. Методом перпендикуляра выполнили расчет площадей пиков зарегистрированной хроматограммы и пиков, полученных в результате разделения. За истинные значения приняли площади идеально разделенных пиков, использованных при моделировании. Результаты сравнения полученных результатов приведено в табл. 1—3, из которого следует, что выполненное разделение пиков в рассмотренном случае дает шестикратный выигрыш в относительной погрешности определения площади пиков: от 32 до 5 %. Разумеется, реальный выигрыш зависит от конкретного вида хроматограммы. Часто встречаются ситуации, когда пики пересекаются настолько, что между ними отсутствует «долина» и провести касательные невозможно. В этом случае разделение пиков — единственное средство определения их площадей и времени удерживания.

Заключение. Предложенное в статье объединение нейросетевых RBF-технологий с применением методов регуляризации решения некорректных задач позволяет получить удовлетворительное разделение пересекающихся пиков и достичь кратного уменьшения погрешности определения площадей разделяемых пиков. Особенно эффективно использование нейронных сетей в совокупности с регуляризацией при очень сильном пересечении пиков, когда для определения площадей невозможно применить методы перпендикуляра и тангенты.

Л и т е р а т у р а

1. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1986.
2. Василенко Г. И. Теория восстановления сигналов. О редукации к идеальному прибору в физике и технике. М.: Сов. радио, 1979.
3. Русинов Л. А. Автоматизация обработки хроматографической информации. М.: Энергия, 1973.

4. Серегина Н. И., Солопченко Г. Н. Простой регуляризующий метод компенсации влияния аппаратной функции на результат измерения // Изв. АН СССР. Техническая кибернетика. 1984. № 2. С. 166—172.

5. Серегина Н. И., Солопченко Г. Н. Коррекция искажений хроматограмм и разделение перекрывающихся хроматографических пиков // Журнал аналитической химии. 1995. Т. 50. № 2. С. 187—189.

6. Miao H., Yu M., Hu S. Artificial neural networks aided deconvolving overlapped peaks in chromatograms // J. Chromatog. A. 1996. V. 749. N. 1. P. 5—11.

7. Li YB, Huang XY, Sha M, Meng XS. Resolution of overlapping chromatographic peaks by radial basis function neural network // Chinese J. Chromatog. 2001. N. 19 (2). P. 112—115.

8. Sentellas S. e. a. Quantitation in multianalyte overlapping peaks from capillary electrophoresis runs using artificial neural networks // J. Chromatog. Sci. 2003. N. 41. P. 145—150.

9. Медянцев Д. В., Пустовалов Д. С., Замятин Н. В. Анализ хроматографической информации с использованием нейронных сетей и генетического алгоритма // Нейроинформатика — 2005: Сб. трудов. VII Всерос. науч.-техн. конф. 2005. Т. 1. С. 147—151.

10. Севастьянов А. А., Харинцев С. С., Салахов М. Х. Нейросетевая регуляризация решения обратных некорректных задач прикладной спектроскопии // Электрон. журнал «Исследовано в России». 2003. Т. 6. С. 2254—2266. [Электрон. ресурс]. <http://zhurnal.ape.relarn.ru/articles/2003/189.pdf> (дата обращения 17.11.2014 г.).

11. Васильев А. Н., Тархов Д. А. Нейросетевое моделирование. Принципы. Алгоритмы. Приложения. СПб.: Изд-во политех. ун-та, 2009.

12. Хабурзания Т. З. Использование нейронных сетей в задачах обработки аналитических сигналов // Нейрокомпьютеры: разработка, применение. 2009. № 11. С. 53—61.

13. Фаткудинова Ш. Р., Солопченко Г. Н. Применение цифровых моделей реальных хроматограмм для оценки погрешности результатов химического анализа // Журнал аналитической химии. 1998. Т. 53. № 11. С. 1158—1165.

Дата принятия 17.01.2014 г.