



АППРОКСИМАЦИОННЫЙ МЕТОД КОРРЕКЦИИ АППАРАТНОЙ ФУНКЦИИ ХРОМАТОГРАФИЧЕСКИХ СИГНАЛОВ

(Самарский государственный технический университет)

Введение

При обработке хроматографических сигналов наибольшей проблемой является разделение наложенных пиков. На практике для разделения наложенных пиков чаще всего используются методы перпендикуляра и триангуляции [1]. Для разделения асимметричных пиков использование данных методов дает большую погрешность, из-за чего актуальным является вопрос разработки новых математических методов для разделения наложенных пиков. Одним из путей снижения погрешности данных методов является предварительная коррекция аппаратной функции анализатора, что позволяет снизить асимметрию пиков, что позволяет применять даже наиболее простой метод перпендикуляра с существенно меньшей погрешностью. Коррекция аппаратной функции для хроматографических сигналов связана с вычислением производной, применение которой для реальных сигналов, содержащих помеху, усиливает влияние помехи на конечный результат. Одним из способов исключить влияние помехи на коррекцию аппаратной функции является предварительная аппроксимация сигнала в базисе функций Чебышева-Эрмита, которые обладают сглаживающим свойством [2].

В настоящей работе предлагается метод коррекции аппаратной функции для хроматографического сигнала с помощью аппроксимационного метода, основанного на использовании функций Чебышева-Эрмита

Аппроксимация сигналов и их производных в базисе функций Чебышева-Эрмита

Функции Чебышева-Эрмита определяются следующим образом:

$$\varphi_n(t) = \frac{1}{\alpha_n} \cdot e^{-\frac{t^2}{2}} \cdot H_n(t), \quad (1)$$

где $\alpha_n = \sqrt{2^n n! \sqrt{\pi}}$ – нормирующая константа;
 $H_n(t)$ – полином Эрмита n -го порядка [3]:

$$H_n(t) = (-1)^n e^{t^2} \frac{d^n e^{-t^2}}{dx^n}. \quad (2)$$

Набор функций $\varphi_n(t)$ (1) задает базис для аппроксимации (кодирования и декодирования) сигналов. Аппроксимация сигнала $f(t)$ с помощью N функций разложения (алгоритм кодирования) заключается в нахождении коэффициентов c_n . В силу локализации функций Чебышева-Эрмита нахождение данных коэффициентов можно свести к формуле:



$$c_n = \int_{-\tau_n}^{\tau_n} f(t) \cdot \varphi_n(t) dx, \quad n = 0, 1, \dots \quad (3)$$

где $[-\tau_n, \tau_n]$ – отрезок, на котором локализована функция $\varphi_n(t)$.

Алгоритм декодирования с помощью n коэффициентов, в результате которого восстанавливается аппроксимированный сигнал, выражается формулой:

$$\hat{f}(t) = \sum_{n=0}^N c_n \cdot \varphi_n(t). \quad (4)$$

За счет рекуррентных свойств полиномов Эрмита можно вычислять производные функций (1) с использованием самих функций (1) предыдущего и последующего порядков:

$$\frac{d\varphi_n(t)}{dt} = \sqrt{\frac{n}{2}} \varphi_{n-1}(t) - \sqrt{\frac{n+1}{2}} \varphi_{n+1}(t), \quad (5)$$

тем самым, можно вычислить производную сигнала, аппроксимированного в рассматриваемом базисе:

$$\frac{d\hat{f}(t)}{dt} = \sum_{n=0}^N c_n \left[\sqrt{\frac{n}{2}} \varphi_{n-1}(t) - \sqrt{\frac{n+1}{2}} \varphi_{n+1}(t) \right]. \quad (6)$$

Коррекция аппаратной функции хроматографического анализатора

Большинство хроматографических анализаторов обладают экспоненциальной аппаратной функцией, для устранения которой требуется сложить сигнал с его производной, умноженной на некоторый коэффициент, соответствующий постоянной времени экспоненты [4]:

$$f_0(t) = f(t) + \tau \frac{df(t)}{dt}. \quad (7)$$

Поскольку сигнал $f(t)$ раскладывается в базисе функций Чебышева-Эрмита, для вычисления производных воспользуемся выражением (6):

$$f_0(t) = \sum_{n=0}^N [c_n \varphi_n(t)] + \sum_{n=0}^N \left[c_{n-1} \sqrt{\frac{n}{2}} \varphi_{n-1}(t) - c_{n+1} \sqrt{\frac{n+1}{2}} \varphi_{n+1}(t) \right]. \quad (8)$$

Кодируя полученный сигнал в базисе функций (1), получим новый набор коэффициентов c_n^0 , соответствующий сигналу со скорректированной аппаратной функцией:

$$c_n^0 = c_n + \tau \left(c_{n-1} \sqrt{\frac{n}{2}} - c_{n+1} \sqrt{\frac{n+1}{2}} \right). \quad (9)$$

Следовательно, для коррекции аппаратной функции в используемом базисе не требуется вычислять саму производную, только сложить соседние отсчеты с нужными весами. Выбор τ необходимо осуществлять итерационно, увеличивая это значение до тех пор, когда спад сигнала не пересечет нулевую линию (следует использовать предшествующую пересечению итерацию в качестве оптимальной).



Экспериментальное исследование

Для проверки работоспособности метода был получен сигнал с существенной асимметрией, извлеченный с помощью программы Хроматэк Аналитик 3.1 из хроматограммы углеводородов, записанной посредством хроматографа Кристалл-5000 с пламенно-ионизационным детектором. Для сигнала были вычислены коэффициенты кодирования (3), по которым вычислялись коэффициенты (9) сигнала с скорректированной аппаратной функцией, и уже по последним, с помощью декодирования (4), аппроксимировался искомый сигнал. Итерационно было подобрано оптимальное значение $\tau = 1,5$. Аппроксимируемый сигнал изображен на рис. 1, а (измеренный сигнал – дискретный, но для удобства отображения представлен в виде зависимости значений от времени). Приведенная погрешность аппроксимации сигнала в базисе функций (1), представленная на рис. 1, б, была вычислена по выражению:

$$\gamma(t) = \frac{f(t) - \hat{f}(t)}{\max[f(t)]} \cdot 100\%, \quad (10)$$

где $f(t)$ – исходный сигнал;

$\hat{f}(t)$ – аппроксимированный по (4) сигнал;

$\max[f(t)]$ – максимум исходного сигнала.

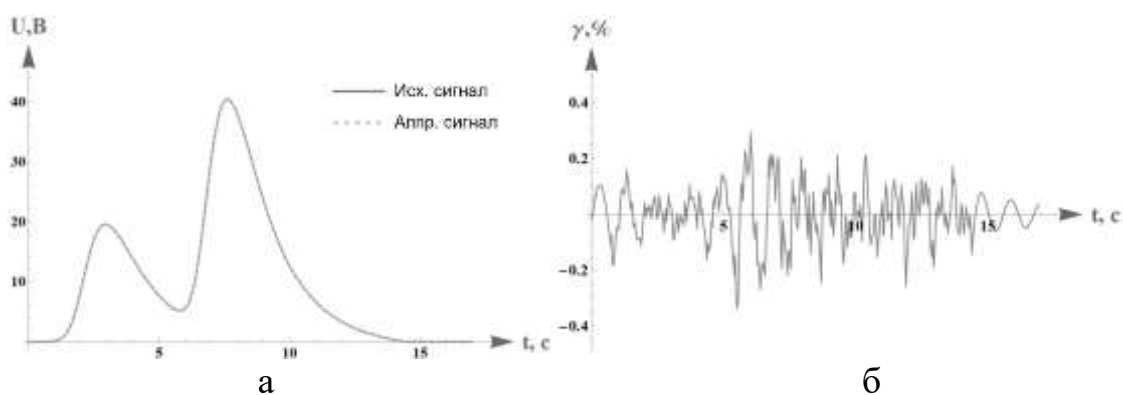


Рис. 1. Аппроксимация сигнала:

а – аппроксимация сигнала, б – приведенная погрешность аппроксимации

Значение озвученной приведенной погрешности не превышает 0,29%, что говорит о высоком качестве аппроксимации. По результатам аппроксимации, с учетом полученного значения $\tau = 1,5$ был восстановлен сигнал со скорректированной аппаратной функцией, представленный на рис. 2. Для сравнения погрешностей предложенного метода с традиционными методами перпендикуляра и триангуляции, была вычислена погрешность определения площади двух пиков в отдельности по каждому методу. Погрешность площади вычисляется по выражению:

$$\delta = \frac{|S - \tilde{S}|}{S} \cdot 100\%, \quad (11)$$



где S – площадь исходного пика;
 \tilde{S} – площадь пика, полученная одним из методов.

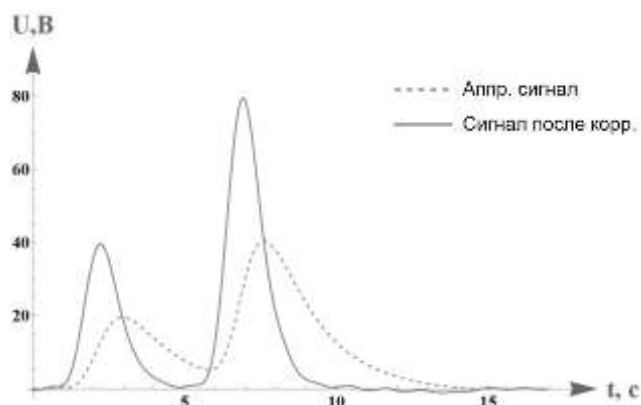


Рис. 2. Коррекция аппаратной функции сигнала

Результаты представлены в таблице 1. Можно видеть, что после коррекции аппаратной функции погрешность применения метода перпендикуляра снижается в 7-10 раз.

Таблица 1 – Погрешности разделения пиков

Метод	$\delta_{S1}, \%$	$\delta_{S2}, \%$
Перпендикуляра	10,8	9,3
Триангуляции	8,5	9,2
Перпендикуляра с коррекцией АФ	1,1	1,27

Заключение

Полученные результаты говорят о целесообразности применения разработанного метода при обработке наложенных хроматографических сигналов, в случае существенной их асимметрии. Метод позволяет автоматизировать процесс разделения наложенных пиков за счет простоты условия остановки итерационного процесса поиска τ .

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-31-90014

Литература

1. Felinger A. Data Analysis and Signal Processing in Chromatography / A. Felinger. – Amsterdam, Netherlands: Elsevier Science B.V., 1998. – 413 p.
2. Балакин Д.А., Штыков В.В., Построение ортогонального банка фильтров на основе преобразований Эрмита для обработки сигналов. // Журнал радиоэлектроники, №9, 2014. – с. 1-15.
3. Суетин П. К. Классические ортогональные многочлены. - 3-е изд., перераб. и доп. -М.: ФИЗМАТЛИТ, 2007. – 480 с.



4. Felinger A. Deconvolution of Overlapping Skewed Peaks / A.Felinger // Anal. Chem.–1994.–vol. 66.–pp. 3066-3072.

О.К. Головнин, А.В. Курганов

АВТОМАТИЗИРОВАННАЯ СИСТЕМА АНАЛИЗА КОММЕНТАРИЕВ ПОЛЬЗОВАТЕЛЕЙ В СЛУЖБЕ МГНОВЕННОГО ОБМЕНА СООБЩЕНИЯМИ TELEGRAM

(Самарский университет)

На текущий момент системы мгновенного обмена сообщениями превращаются в полноценные социальные сети со своей спецификой. В работе представлена разрабатываемая автоматизированная система анализа комментариев пользователей к постам в групповых чатах в службе мгновенного обмена сообщениями Telegram на предмет нарушения правил сообщества или законодательства страны размещения, построенная на основе нейронной сети с LSTM-ячейками. Предложенная система позволяет выявлять не только конкретные запрещенные слова или словосочетания, но и выделять и понимать контекст, в котором они использованы.

Разрабатываемая система включает в себя несколько подсистем: подсистему получения информации из службы мгновенного обмена сообщениями Telegram, подсистему обработки информации на основе нейронной сети с LSTM-ячейками, подсистему генерации и доставки уведомлений модераторам о нахождении подозрительного сообщения, подсистему суммаризации текстовой информации. Особенностью работы системы является возможность обучаться на основе решений модератора относительного того или иного подозрительного сообщения. Система разрабатывается на языке программирования Python.

Подсистема получения информации из службы мгновенного обмена сообщениями Telegram организуется в виде робота, имитирующего действия пользователя [1]. Найденная информация (посты и комментарии) сохраняются в виде json-файлов. Список анализируемых групп задается модератором системы.

Подсистема обработки информации реализуется на основе нейронной сети с LSTM-ячейками [2], где в качестве функции активации используется Softmax. Подсистема считывает данные из поступающих json-файлов. Разбиение входящего сообщения на слова выполняется с помощью text_to_word_sequence библиотеки Keras. В подсистеме используется метод One-Hot Encoding для преобразования категориальных данных в числовые, что позволяет преобразовать текстовые сообщения в закодированное 0 и 1 сообщение, которое обрабатывается с помощью метода Embedding [3], уменьшающего размерность сообщений. Алгоритм добавления нового Embedding-слоя в нейронную сеть при указании модератором его принадлежности к тому или иному классу представлен на рисунке, для чего