

13. Sijl J., Vos H.J., Rosental T., de Jong N., Lohse D., Versluis M. Combined optical and acoustic detection of single microbubble dynamics // J. Acoust. Soc. Am. 2011. Vol. 130. Iss. 5. PP. 3271-3281.
14. Змиевской Г.Н., Крылов Ю.В., Скворцов С.П. Моделирование процессов при оптическом зондировании ультразвуковой кавитационной области // Биомедицинские технологии и радиоэлектроника. 2005. № 9. С. 55-60.
15. Змиевской Г.Н., Крылов Ю.В., Скворцов С.П. Исследование возможностей оптического зондирования кавитирующей жидкости при ультразвуковом воздействии // Биомедицинская радиоэлектроника. 2006. № 10. С. 32-36.
16. Zmievskoy G., Skvortsov S. Ultrasound Cavitation Detection by Means of Optical Probing / Proceedings of the 6th Russian-Bavarian Conference on Biomedical Engineering. Moscow, 2010. PP. 121-122.
17. Скворцов С.П. Модель светорассеяния в ультразвуковой кавитационной области // Наука и образование. 2015. № 3. PP. 102-119.

Сергей Павлович Скворцов,
канд. техн. наук, доцент,
Никита Сергеевич Масленков,
дипломник,
Виталий Игоревич Нечаев,
дипломник,
Артем Петрович Кравченко,
дипломник,
кафедра БМТ-1 «Биомедицинские
технические системы»,
МГТУ им. Н.Э. Баумана,
г. Москва,
e-mail: skvortsov@bmstu.ru

В.В. Манойлов, Л.В. Новиков, И.В. Заруцкий, А.Г. Кузьмин, Ю.А. Титов

Методы обработки масс-спектрометрических сигналов выдыхаемых газов для медицинской диагностики

Аннотация

Состав выдыхаемых человеком газов может быть проанализирован на малогабаритном квадрупольном масс-спектрометре. По результатам такого анализа возможна диагностика состояния здоровья. Математическая обработка масс-спектрометрических сигналов позволяет разделить обрабатываемые масс-спектры по классам: здоров или болен, а также по группам болезней. Представлены описания алгоритмов первичной обработки данных и алгоритмов классификации на основе многомерных статистических методов: дискриминантного и кластерного анализов.

Введение

Анализ состава выдыхаемого воздуха является одним из важных направлений неинвазивной медицины. Диагностика на основе анализа выдыхаемых газов имеет ряд преимуществ по сравнению с традиционными лабораторными методами. Анализ газовой смеси безопасен для персонала, так как не связан с работой с химическими и биологическими жидкостями. Он является относительно дешевым, занимает немного времени и позволяет обнаруживать летучие компоненты в выдыхаемом воздухе на уровне следовых концентраций вещества в реальном масштабе времени. В статье рассматриваются алгоритмы обработки масс-спектров выдыхаемых газов, полученные на квадрупольном масс-спектрометре MS7-200, с электронной ионизацией и с прямым капиллярным вводом пробы [1]-[3]. Анализируемый газ при атмосферном давлении через капиллярный ввод подается в камеру ионизации источника ионов с электронным ударом. Получившиеся ионы вводятся в масс-анализатор квадрупольного типа. Полученные в процессе регистрации масс-спектрометрические сигналы обрабатываются с помощью специализированного программного обеспечения и сравниваются со спектрами в библиотеке стандартных масс-спектров, затем проводятся идентификация отдельных компонент спектра и определение их концентрации. Капиллярная система ввода пробы в масс-спектрометр, прогреваемая до 50 °С, позволяет проводить анализ на расстоянии до 5 м от прибора. Для анализа используется до 1 мл пробы в минуту. В вакуумной системе применяются диффузионный или турбомолекулярный насосы.

Целью рассматриваемой работы является анализ возможностей алгоритмов обработки информации для классификации пациентов по группам здоровья и патологий.

Традиционные методы анализа данных масс-спектрометров, как правило, включают в себя предварительную (первичную) обработку с целью обнаружения пиков и вторичную об-

работку с целью извлечения нужной для исследователя информации о количественном и качественном составе анализируемого вещества.

1 Предварительная обработка данных

Первый этап обработки масс-спектров включает в себя обнаружение пиков на фоне шума, «выбросов», изменения базовой линии и эффекта неполного разделения пиков с целью превращения непрерывного спектра в линейчатый спектр. Для этой задачи применяется метод поиска на основе согласованной фильтрации или метод сравнения величины производных в трех точках скользящего окна данных [4]. Оба подхода могут использоваться как самостоятельно, так и последовательно (первый за вторым), дополняя друг друга.

Масс-спектр можно представить в виде аддитивной смеси K пиков $X_k(t)$, $k = 1, \dots, K$, например гауссовой формы и шума $n(t)$, базовой линии $f(t)$:

$$y(t) = \sum_{k=1}^K x_k(t) + n(t) + f(t), \quad (1)$$

где

$$x_k(t) = A_k e^{-\frac{t-t_k}{2w_k^2}};$$

t – непрерывное или дискретное время, $t = i\delta t$, $i = 1, 2, \dots, N$, N – общее число отсчетов данных масс-спектра; δt – интервал дискретизации; A_k , t_k , w_k – интенсивность, положение и среднеквадратическая ширина k -го пика соответственно; $n(t)$ – шум; $f(t)$ – базовая линия.

В первом методе используется скользящая свертка исходного сигнала $y(t)$ с функцией, описывающей форму пика:

$$s(t) = \int_{t-T}^{t+T} y(t_1) g(t-t_1) dt_1, \quad (2)$$

где t – независимая переменная, время; $g(t)$ – форма пика; T – размер половины скользящего окна данных.

В качестве формы пика в рассматриваемой работе использовались нормированные к единице две функции:

- первая – на основе гауссианы:

$$g_k(t) = e^{-\frac{t^2}{2w^2}},$$

где w – полуширина пика;

- вторая – на основе функции Эрмита:

$$\Psi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2^n n! \sqrt{\pi}}} H_n(x),$$

где

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

– полином Эрмита порядка n при значении $n = 2$.

В результате фильтрации отношение сигнала к шуму увеличивается в 6...7 раз для гауссовых функций и в 7...10 раз для функций Эрмита. Однако функции Эрмита имеют преимущество в повышении разрешения плохо разделенных пиков по сравнению с функциями Гаусса. Процедура свертки позволяет надежно обнаружить пики спектра и определить их положение.

Полученный сигнал $s(t)$ сравнивается с порогом h , с помощью которого обнаруживаются пики и определяются их положения. Если величина $s(t)$ превышает порог h , то эту точку t считаем принадлежащей пику. Более подробное описание рассмотренных алгоритмов можно найти в [5], [6].

Во *втором методе* используется традиционный подход обнаружения вершины пиков по пересечению первой производной нулевой линии. С целью повышения надежности обнаружения начала, конца, вершины и седловины между пиками предлагается сравнивать величины производных в трех точках скользящего окна данных. Различные методы оценки параметров пиков с использованием производных описаны в работах [4], [7]-[9].

2 Вторичная обработка данных

Следующим шагом обработки является проведение классификации обрабатываемых масс-спектров. На основе методов дискриминантного и кластерного анализов проводится классификация по группам. Масс-спектры групп здоровых и

больных пациентов заметно отличаются прежде всего наличием во второй группе дополнительных компонент с массами 55, 59, 64 и 71 дальтон. Этот признак свидетельствует о наличии изменений в организме, вызванных определенной болезнью. Можно предположить, что соотношение интенсивностей компонент масс-спектров содержит информацию о виде патологии, что позволяет, в свою очередь, после набора определенной статистики эффективно диагностировать патологию и, более того, классифицировать ее по группам болезней.

Набор данных «здоровые пациенты – интенсивность компонент их масс-спектра» образует матрицу «объект-признак» $XS = |x_{ij}|$, где $i = 1, \dots, N$ – номера объектов (пациентов); $j = 1, \dots, M$ – номера признаков – массовых чисел; x_{ij} – интенсивность j -й спектральной компоненты i -го здорового пациента.

Набор данных больных пациентов образует $I_0 \times J$ – матрицу «объект-признак», каждая строка которой содержит интенсивности масс компонент выдыхаемого воздуха одного пациента. Эту матрицу обозначим как $XI = |x_{ij}^0|$, где $i = 1, \dots, I_0$ – номера объектов (пациентов); $j = 1, \dots, J$ – номера признаков – массовых чисел; x_{ij}^0 – интенсивность j -й спектральной компоненты i -го больного пациента. Из матриц «пациент – масс-спектр здоровых и больных» образуем единую матрицу X данных размера $(I_0 + I) \cdot J$: $X = (|XS, XI|)$.

Используя эти данные, сформируем пространство главных компонент (ГК) и вычислим матрицу счетов T .

Применим к матрице T алгоритмы дискриминантного и иерархического агломеративного кластерного анализов.

2.1 Дискриминантный анализ

Дискриминантный анализ является разделом многомерного статистического анализа, который включает в себя методы классификации многомерных наблюдений по принципу максимального сходства анализируемых наблюдений с наблюдениями, отнесенными к определенным классам по результатам обучения. При нахождении коэффициентов дискриминантной функции используют QR-разложение или алгоритмы, описанные в [10]. На этапе предварительной обработки выполняется отбор масс-спектров здоровых людей для их включения в обучающую выборку.

Необходимость выполнения этого шага объясняется тем, что даже в масс-спектрах практически здоровых людей могут содержаться спектральные линии с амплитудами, характерными для масс-спектров больных людей. Выполнение этого шага осуществляется путем выявления масс-спектров, содержащих

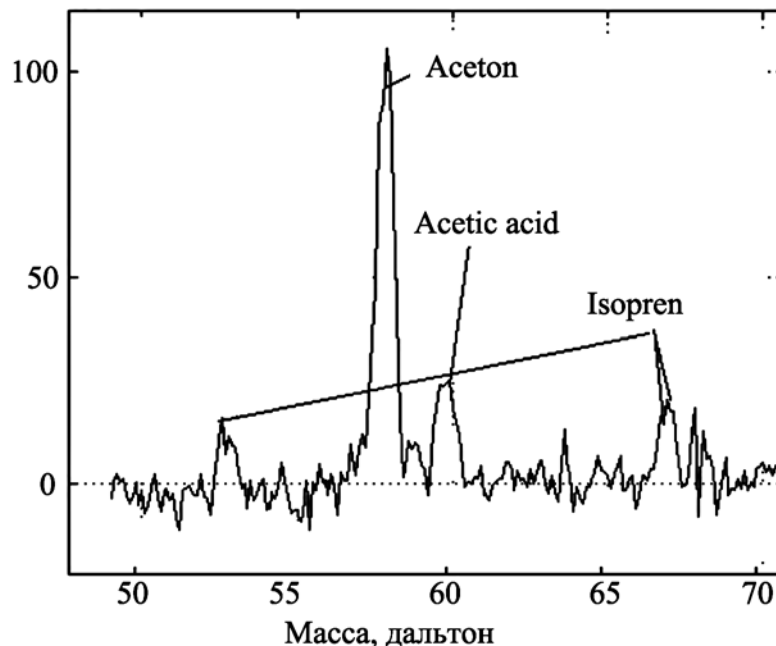


Рис. 1. Эталонный масс-спектр (по вертикальной шкале – интенсивность пиков в относительных единицах)

наименьшее отклонение от эталонного масс-спектра, т. е. усредненного масс-спектра здоровых людей. На рис. 1 представлен эталонный масс-спектр.

Результатом проведения обучения является набор коэффициентов дискриминантной функции, которые вычисляются по одному из алгоритмов, рассмотренных в [10]. В качестве переменных в алгоритмах многомерной статистики использовались не исходные значения амплитуд пиков на определенных массах, а главные компоненты, полученные путем преобразования исходных данных по методу главных компонент (Principal Component Analysis, PCA).

С помощью вычисленных коэффициентов дискриминационных функций находят границы, которые разделяют исходные масс-спектры выдыхаемых газов на группы: здоровых людей и людей с патологиями. На рис. 2 показаны результаты проведения дискриминантного анализа соответственно для линейного (рис. 2а) и квадратичного (рис. 2б) методов.

Обработка масс-спектров выдыхаемых газов посредством методов дискриминантного анализа показала, что количество масс-спектров больных людей, отнесенных к классу здоровых, равно нулю.

2.2 Кластерный анализ

Обработка данных состоит из двух этапов: *обучения* и *классификации*. На этапе *обучения* по спектрам здоровых пациентов формируется пространство ГК: вычисление собственных

векторов P и собственных чисел $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_j]$ ковариационной матрицы K и вычисление проекций обучающих образцов на оси главных компонент матрицы: $T = X \overset{\circ}{S} \times P$, где $\overset{\circ}{X}$ – центрированный масс-спектр здоровых пациентов.

На этапе *диагностики* анализируется масс-спектр обследуемых пациентов.

1. Вычисляются проекции масс-спектра пациента на оси главных компонент: $T_c = X I \overset{\circ}{S} \times P$, где $\overset{\circ}{X} I$ – центрированный масс-спектр больного. Центрирование вектора $X I$ производится с помощью средних значений по столбцам матрицы X .

Применим к матрице T_c алгоритмы иерархической агломеративной кластеризации, необходимые для разделения объектов на кластеры. В результате получим несколько кластеров данных, характеризующих специфику заболевания пациентов. На рис. 3 показано разбиение здоровых (1...12) и больных (20...32) пациентов на кластеры. Больные пациенты разделены на восемь групп: 13, 14, 16, 17; 15; 18, 20, 28; 19; 21, 23, 29, 30, 31; 22, 26; 24, 25, 27; 32. PCA-1 и PCA-2 – первая и вторая оси главных компонент.

В ходе дальнейшей обработки определяется величина вероятности правильной классификации заболевания при заданной величине порога. Эта вероятность будет максимальной, если евклидово расстояние между центром «облака» спектров здоровых пациентов (центроидой) и спектральной точкой больного в пространстве ГК будет близким к нулю. Более подробное описание алгоритма приведено в работах [11], [12].

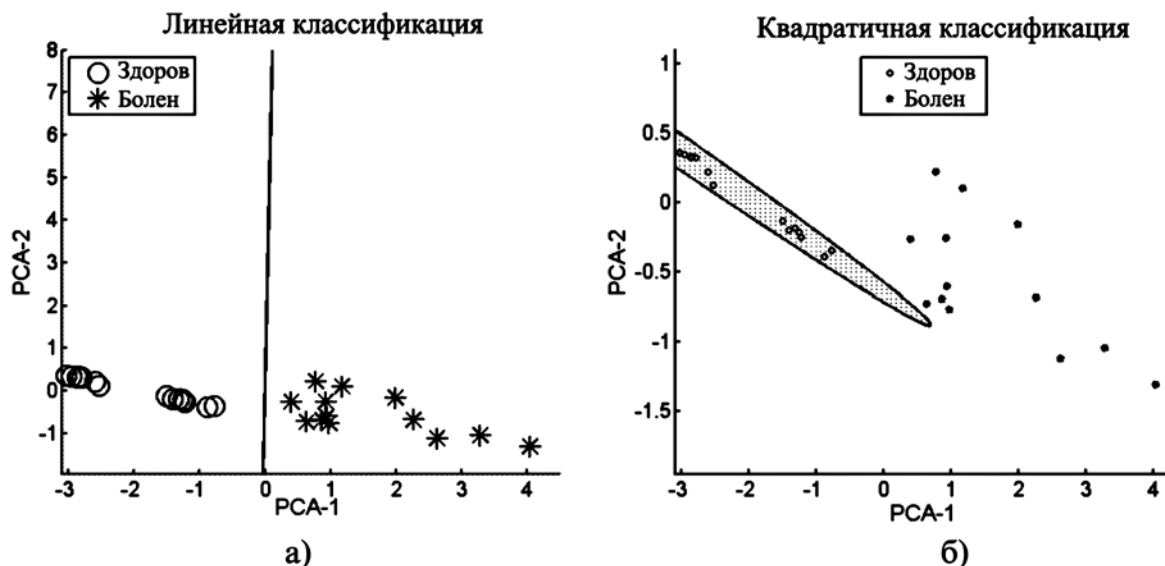


Рис. 2. Результат дискриминантного анализа: по горизонтальной оси – значения переменных по первой главной компоненте (PCA-1); по вертикальной оси – значения переменных по второй компоненте; а) результат линейного дискриминантного анализа; б) результат квадратичного дискриминантного анализа

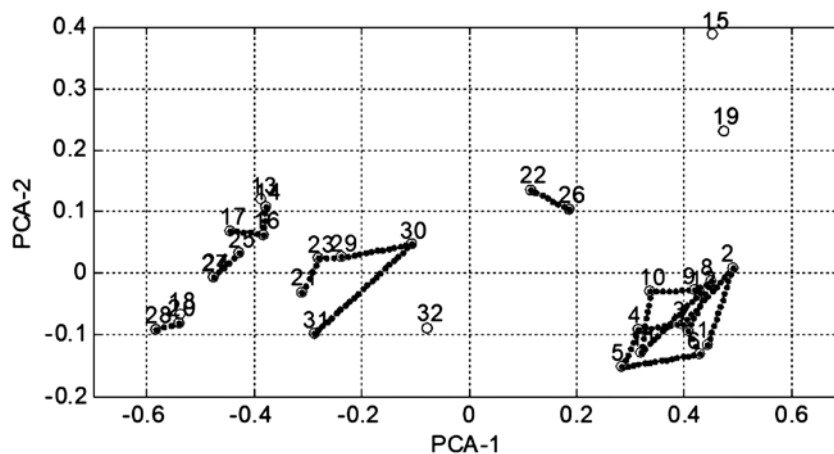


Рис. 3. Классификация масс-спектров по состоянию здоровья посредством кластерного анализа. Цифрами на рисунке обозначены номера масс-спектров выдыхаемых газов здоровых и больных пациентов

Заклучение

Рассмотренные алгоритмы предварительной обработки и классификации на основе дискриминантного и кластерного анализов дают возможность автоматического принятия решений о различии масс-спектров без визуального анализа информации, представленной в графическом виде. Такое автоматическое принятие решений может быть полезным при массовых экспресс-анализах. При незначительных отклонениях вычисленных значений дискриминантных функций целесообразным является повторение измерений с дополнительным анализом информации. Кроме того, данные методы целесообразно использовать параллельно с другими методами анализа патологий.

Рассмотренные алгоритмы просты в реализации, дают возможность автоматического принятия решения о принадлежности проверяемого сигнала к определенному классу.

Алгоритмы имеют практическое значение для специализированных малогабаритных бюджетных масс-спектрометров, в которых особенно целесообразно использовать автоматизированную классификацию по группам здоровья математическими методами.

Работа поддержана государственным заданием № 075-00780-19-00.

Список литературы:

1. Кузьмин А.Г. Квадрупольный масс-спектрометр / Патент на полезную модель № 94763. РФ. 27.05.2010 г.
2. Кузьмин А.Г., Титов Ю.А. Малогабаритные масс-спектрометры для динамических исследований состава выдыхаемого воздуха / Труды I Международной научно-практической конференции «Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине», СПб., 23-26 ноября 2010 г. Ч. 3. С. 266-270.
3. Кузьмин А.Г., Ткаченко Е.И., Орешко Л.С., Титов Ю.А. Перспективы метода масс-спектрометрической ароматодиагностики по составу выдыхаемого воздуха / Тезисы докладов X Евразийской научной конференции «ДОНОЗОЛОГИЯ-2014». СПб., 18-19 декабря 2014 г. С. 229-231.
4. Новиков Л.В., Куркина В.В. Метод оценки параметров спектральных пиков // Научное приборостроение. 2017. Т. 27. № 3. С. 99-106.
5. Vivó-Truyols G., Torres-Lapasió J.R., van Nederkassel A.M., van der Heyden Y.V., Massart D.L. Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals. Part I: Peak detection // Journal of Chromatography A. 2005. Vol. 1096. PP. 133-145.

6. Манойлов В.В., Кузьмин А.Г., Заруцкий И.В., Титов Ю.А., Самсонова Н.С. Методы обработки и исследование возможностей масс-спектров выдыхаемых газов // Научное приборостроение. 2019. Т. 29. № 1. С. 106-109.
7. Fredriksson M.J., Petersson P., Axelsson B.-O., Bylund D. An automatic peak finding method for LC-MS data using Gaussian second derivative filtering // J. Sep. Sci. 2009. Vol. 32. PP. 3906-3918.
8. Gregoire J.M., Dale D., Bruce van Dover R. A wavelet transform algorithm for peak detection and application to powder x-ray diffraction data // Review of Scientific Instruments. 2011. Vol. 82. PP. 015105-015112.
9. Du P., Kibbe W.A., Lin S.M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching // Bioinformatics. 2006. Vol. 22. PP. 2059-2065.
10. Манойлов В.В., Титов Ю.А., Кузьмин А.Г., Заруцкий И.В. Алгоритмы дискриминационного анализа для классификации масс-спектров выдыхаемых газов // Научное приборостроение. 2017. Т. 27. № 3. С. 33-57.
11. Новиков Л.В., Куркина В.В. Многомерная обработка данных масс-спектрометрического анализа состава выдыхаемого воздуха / Сб. трудов XXIX Междунар. науч. конф. «Математические методы в технике и технологиях – ММТТ-29»; в 12 т. Т. 9. Под общ. ред. А.А. Большакова. – Саратов: Саратов. гос. техн. ун-т; СПб.: СПбГТИ(ТУ), СПбПУ, СПИИРАН; Самара: Самарск. гос. техн. ун-т, 2016. С. 32-38.
12. Бериков В.Б., Лбов Г.С. Современные тенденции кластерного анализа / Институт математики им. С.Л. Соболева СОРАН, 2008. 26 с / <https://reslib.org/books/642840>.

*Владимир Владимирович Манойлов,
д-р техн. наук, доцент,
гл. научный сотрудник, зав. лабораторией,
Лев Васильевич Новиков,
д-р физ.-мат. наук, ведущий научный сотрудник,
Игорь Вячеславович Заруцкий,
канд. техн. наук, ст. научный сотрудник,
Алексей Георгиевич Кузьмин,
канд. физ.-мат. наук, ст. научный сотрудник,
Юрий Алексеевич Титов,
научный сотрудник,
ФГБУН «Институт аналитического
приборостроения» РАН,
г. С.-Петербург,
e-mail: manoilov_vv@mail.ru*

*С.П. Серегин, Т.Н. Говорухина, С.Н. Родионова, И.М. Холименко,
С.В. Петров, Г.В. Сипливый, В.В. Дмитриева*

Система поддержки принятия решений при дифференциальной диагностике форм пиелонефрита у беременных женщин на основе иммунного и оксидантного статусов

Аннотация

Приведены современные математические методы и информационные технологии, которые на основе показателей иммунного и оксидантного статусов позволяют точно проводить дифференциальную диагностику форм пиелонефрита у беременных. Проверка результатов срабатывания полученных решающих правил на контрольной выборке, показал, что диагностическая эффективность предлагаемого метода достигает 93 %, что приемлемо для использования в медицинской практике.

Введение

Заболеваемость пиелонефритом, в том числе гестационным, в течение последних лет неуклонно увеличивается. Часто наблюдается атипичное течение пиелонефрита. Несмотря на высокоинформативные методы диагностики, применяемые при

выявлении пиелонефрита, дифференциальная диагностика серозных и гнойных форм пиелонефрита остается неразрешенной проблемой. Многочисленные исследования и опыт практической медицины свидетельствуют о том, что показатели иммунного и оксидантного статусов больных пиелонефритом могут служить информативными тестами для оценки формы