

ВЛИЯНИЕ ФИЛЬТРАЦИИ НА КЛАССИФИКАЦИЮ СПЕКТРАЛЬНЫХ ДАННЫХ

Э.Э. Ильясова^{1,2}, А.И. Князькова^{1,2}, Е.А. Сандыкова²

Научный руководитель: старший преподаватель, В.С. Заседатель

¹Национальный исследовательский Томский государственный университет,

Россия, г. Томск, пр. Ленина, 36, 634050

²Сибирский государственный медицинский университет

Россия, г. Томск, Московский тракт, 2, 634050

E-mail: elenor.93@mail.ru

INFLUENCE OF FILTRATION ON CLASSIFICATION OF SPECTRAL DATA

E.E. Ilyasova^{1,2}, A.I. Knyazkova^{1,2}, E.A. Sandykova²

Scientific Supervisor: senior lecturer, V.S. Zasedatel

Tomsk State University, Russia, Tomsk, Lenin Ave., 36, 634050

Siberian State Medical University, Russia, Tomsk, 2 Moskovsky trakt, 634050

E-mail: elenor.93@mail.ru

Abstract. *In the present study, we considered the influence of filtration on classification of model absorption spectra of exhaled air of healthy volunteers and patients with obstructive chronic lung disease with additive noise. The classification was carried out by the Support Vector Machine (SVM). According to the analysis the median filter negatively affects on data classification, the Savitzky–Golay filter increases specificity, however reduces sensitivity, the Second Order Butterworth Filter increases specificity and sensitivity.*

Введение. Одним из наиболее важных процессов в организме является газообмен с окружающей средой, основой которого является поглощение кислорода и выделение углекислого газа. Кроме этого в выдыхаемом человеком воздухе содержится до 600 летучих соединений, в том числе молекулы-биомаркеры эндогенного (продуцируемого в организме) происхождения, механизмы образования или выделения которых обладают специфичностью, достаточной для исследования как нормальных, так и патологических процессов.

Типичными маркерами для бронхо-легочных заболеваний являются молекулы этан и пентан преимущественно генерируются культурами клеток под действием активных форм кислорода. При анализе конденсата выдыхаемого воздуха больных ХОБЛ обнаружено следующие маркеры: H₂O₂, эйкозаноиды (лейкотриены, простаноиды, изопростаны), производные NO (S-нитросотиолы, нитриты и нитраты), альдегиды [1].

Появление в выдыхаемом воздухе указанных метаболитов является потенциальной основой для создания методов неинвазивной диагностики бронхо-легочных заболеваний на основе спектрального анализа выдыхаемого воздуха [2].

В рамках данной задачи представляет интерес анализ влияния процесса предварительной обработки экспериментальных спектров на примере лазерного оптико-акустического ИК газоанализатора на точность диагностики.

Процесс измерения спектров поглощения связан с воздействием на работу оптико-акустического газоанализатора линейных и нелинейных искажений, которые, в общем случае, являются случайными и моделируются в результирующем сигнале шумовой составляющей. В зависимости от применяемых методов обработки сигналов шум может моделироваться в виде аддитивного или мультипликативного вклада в сигнал.

В общем случае анализ сигналов затруднен наличием шумов, имеющих случайный характер с априорно неизвестными статистическими характеристиками. Уменьшение вклада шумовой составляющей в результаты обработки сигнала достигается алгоритмами фильтрации и сглаживания первичного экспериментального сигнала [3].

Чтобы выбрать адекватные методы сглаживания и фильтрации спектров поглощения следует принять во внимание физические условия измерения спектров, согласно которым термодинамические параметры газовой пробы и состояние измерительного прибора (оптико-акустического газоанализатора) в процессе измерения остаются неизменными. Учет этих особенностей сигнала (спектров поглощения) позволяет выбирать методы фильтрации сигналов со стационарным шумом. Для таких шумов, в частности, частотный спектр имеет одинаковый состав как на всем интервале измерения спектра поглощения, так и на его частях.

Материалы и методы исследования. В соответствии с литературными данными о составе выдыхаемого воздуха у различных групп пациентов [4,5] были синтезированы с использованием базы данных Nitran спектры поглощения выдыхаемого воздуха для здоровых лиц и пациентов с хронической обструктивной болезнью легких. Состав был выбран следующим образом: 1% H₂O, 5% CO₂, 0,3 ppm N₂O, 1,5ppm CO, 1,7ppm CH₄ – для здоровых лиц; 0,9% H₂O, 3,5% CO₂, 1,1ppm N₂O, 5ppm CO и по 20ppm H₂O₂ и C₂H₆.

Сравнивая модельные сигналы здорового человека и пациента с ХОБЛ (рис. 1) можно заметить, что разница в амплитудах в низкочастотной части спектра около 10% от максимальной амплитуды спектра здорового человека и порядка 30% - для высокочастотной части спектра. Можно предположить, что если шум будет превышать этот порог, то диагностика таких данных может быть ошибочна.

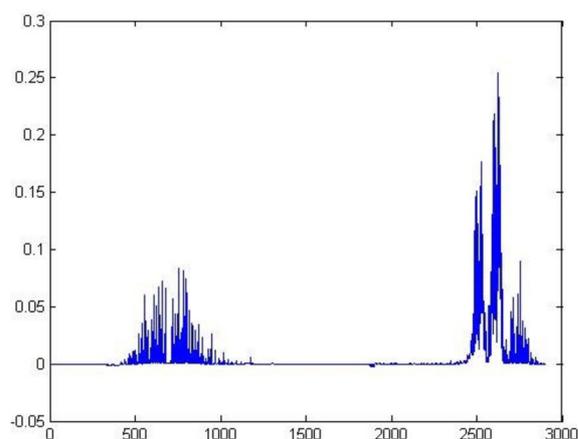


Рис. 1. Разница в амплитудах модельных спектров поглощения выдыхаемого воздуха здорового человека и пациента с ХОБЛ

В качестве алгоритма классификации был использован метод опорных векторов (SVM-метод, от англ. «support vector machine»). Основной идеей метода является перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей исследуемые классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

Для подавления шумов были использованы следующие фильтры: медианный, Баттерворта, дискретная фильтрация с помощью фильтра Савицкого—Голея. Для проверки работы фильтров модельный сигнал был зашумлен посредством аддитивного добавления белого шума разной амплитуды.

Результаты. Результаты анализа показывают, что медианный фильтр негативно сказывается на классификации данных, при увеличении окна медианного фильтра ошибка классификации возрастает. При использованных параметрах фильтр Савицкого-Голея – увеличивает специфичность, однако уменьшает чувствительность. При использованных параметрах фильтр Баттерворта 2-го порядка увеличивает специфичность и чувствительность.

Заключение. Результаты анализа показывают, что в задаче неинвазивной диагностики бронхолегочных заболеваний на основе спектрального анализа выдыхаемого воздуха методами ИК спектроскопии и хемометрики предварительная фильтрация экспериментальных данных фильтром Баттерворта 2-го порядка увеличивает специфичность и чувствительность.

Следует отметить, что шум со значениями более 10% не типичен для экспериментальных данных, однако различия в экспериментальных спектрах поглощения также могут быть существенно меньше. В соответствии с данными (рис. 1), определяющим являются не абсолютные значения амплитуды шума, а его величина по отношению к различиям в экспериментальных спектрах поглощения.

Работа выполнена при частичной финансовой поддержке РФФИ, грант № 17-00-00186.

СПИСОК ЛИТЕРАТУРЫ

1. Montuschi P. (2005). Exhaled breath condensate analysis in patients with COPD. Clin. Chim. Acta, no. 356, pp. 22 – 34.
2. Кузнецов А.И., Логачев А.П., Степанов Е.В. Анализ выдыхаемого человеком воздуха методами диодной лазерной спектроскопии // Известия АН СССР. Серия физическая. – 1990. – Т. 54. – № 10. – С. 15 – 21.
3. Айчифер Э.С., Джервис Б.У. Цифровая обработка сигналов. Практический подход. 2-е издание: пер. с англ. — М.: Издательский дом «Вильямс», 2004. – 992 с.
4. Amann A., Lacy Costello B. de, Miekisch W., Schubert J., et al. (2014). The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J. Breath Res. , no. 8, pp. 034001.
5. Vreman H.J., Mahoney B., Stevenson D.K. (1995) Carbon monoxide and carboxyhemoglobin. Adv. Pediatrics, no. 42, pp. 303–334.