

УДК 615.471:616-071

И.Г. Перова

Харьковский национальный университет радиоэлектроники, Харьков

ОЦЕНКА СОСТОЯНИЯ БОЛЬНЫХ С НАРУШЕНИЕМ ФУНКЦИИ НОСОВОГО ДЫХАНИЯ С ПОМОЩЬЮ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ

В статье рассмотрен подход, позволяющий подвергать процедуре нечеткой кластеризации дискретные временные ряды путем их преобразования в вид таблицы «объект-свойство». Преобразование заключается в замене дискретных временных отсчетов параметрами модели авторегрессии второго порядка. Для проверки адекватности модели была проведена ее апробация на дискретных временных рядах показателей риноманометрии, таких как дифференциальное давление между хоаной и подмасочным пространством и расход воздушного потока, которые были зарегистрированы синхронно с частотой дискретизации 200 Гц.

Ключевые слова: нечеткая кластеризация, авторегрессия, риноманометрия.

Введение

Нарушение функции носового дыхания в настоящее время является крайне сложной проблемой современной оториноларингологии [1]. К заболеваниям, приводящим к затрудненному носовому дыханию, относят острые и хронические воспалительные заболевания носа и околоносовых пазух, вазомоторный, аллергический и полипозный ринит, различные опухолевые процессы в полости носа, посттравматические искривления носовой перегородки. Несмотря на большое количество работ, направленных на поиск способов объективной диагностики в данной области, до сих пор не было найдено информативного и эффективного метода диагностики нарушений носового дыхания. Согласно рекомендациям международного комитета по объективной оценке носового дыхания (ISOANA) среди различных методов исследования, включающих компьютерную и магниторезонансную томографию, наиболее клинически значимым является метод активной передней риноманометрии (ПАРМ), основным диагностическим параметром которой является величина носового сопротивления R [2], вычисляемая как отношение внутриносового дифференциального давления Δp к величине расхода воздушного потока $\dot{V}$ согласно формуле $R = \Delta p / \dot{V}$ [3].

При использовании подобного подхода все равно остается большое количество проблем, связанных с проведением объективной диагностики в данной области, таких как невозможность использовать стандартные подходы математической статистики в связи с тем, что заранее не известен закон распределения данных в пространстве признаков, слабая субъективная оценка симптомов самим пациентом, сильная зависимость показателей риноманометрии от антропометрических данных пациента, его возраста и расы [4 – 6], малых объемов выборки.

Целью данной работы является поиск новых решений при обработке данных риноманометрических исследований.

Формирование диагностических признаков

Для исследования в рамках данной работы были взяты данные риноманометрического исследования, проведенного с помощью прибора «Optimus» [7], которые представляют из себя временные ряды. Измеряемыми параметрами являлись дифференциальное давление между хоаной и подмасочным пространством и расход воздушного потока, которые были зарегистрированы синхронно с частотой дискретизации 200 Гц. Пример отрезка временных рядов представлен в табл. 1. Следует отметить, что временной ряд для одного из признаков составляет от 5000 до 11000 отсчетов. То есть каждый пациент характеризуется как минимум 10000 наблюдений.

Таблица 1

Параметры риноманометрии
для одного пациента

Дифференциальное давление	Расход воздушного потока
50,547	-356,738
55,278	-356,738
59,511	-399,5799
62,997	-406,748
62,25	-413,416
60,756	-415,2497
61,503	-412,9159
59,511	-412,9159

Учитывая все сложности, связанные с особенностью подобных данных, следует отметить необходимость использования для их обработки методов, которые были бы нечувствительны к закону распределения данных и не ориентированы на эталонные объ-

екты в рассматриваемой выборке, то есть методы, основанные на принципах самообучения. Среди подобных методов следует отметить адаптивный метод нечеткой кластеризации данных [8, 9], который может быть применен для сгустков данных различных форм. Подобный метод является частично нечувствительным к соотношению количества признаков к количеству объектов исследования, в отличие от искусственных нейронных сетей (ИНС), которые также могут быть применены в данном случае.

Данная задача относится к задачам, при которых невозможно получение больших объемов выборки. Основной причиной является большое количество признаков, характеризующих объект исследования (большая частота дискретизации при проведении измерений), сложность измерительных процедур и прочие факторы, которые могут ограничивать число объектов исследования (пациентов). В такой ситуации есть несколько вариантов сокращения размерности входного пространства признаков.

Первым вариантом является использование метода главных компонент (PCA – Principal component analysis) [10 – 12] или автоассоциативного трехслойного персептрона типа «бутылочное горлышко», («Bottle Neck») [10, 13]. Но подобные подходы могут использоваться в случае, когда параметры каждого объекта не являются функциями времени. В той ситуации, когда важен порядок следования одного объекта за другим, то есть исходные данные – это дискретный временной сигнал, необходимо использовать другой подход к сокращению размерности входного пространства признаков и обработке подобных данных. При анализе временных рядов были построены

математические модели каждого из них на основе подхода Дж. Бокса и Г. Дженкинса [14].

В качестве базовой модели была принята модель авторегрессии 2-го порядка вида:

$$x(k) = a_0 + a_1 x(k-1) + a_2 x(k-2), \quad (1)$$

где k – дискретное время, для нашего случая 0,005с.

Исходные ряды предварительно были процентрированы с помощью процедуры:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x(k); \quad \tilde{x} = x(k) - \bar{x}.$$

За счет этого базовая модель приобретает вид уравнения (1) при $a_0 = 0$. Таким образом, для каждого из рядов была построена своя математическая модель вида:

$$x(k) = a_1 x(k-1) + a_2 x(k-2).$$

На основании рассчитанных коэффициентов авторегрессии сформированы векторы признаков, в которые помимо двух коэффициентов авторегрессии входили также среднее значение временного ряда, получаемое с помощью процедуры центрирования, и величина дисперсии, определяемая по формулам:

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^N (x(k) - \bar{x})^2 = \frac{1}{N} \sum_{k=1}^N \tilde{x}^2(k),$$

т.е. в качестве исходных данных для последующей обработки использовался набор из 4 признаков для каждого из временных рядов. Общее количество признаков для каждого пациента принято равным 8. Количество объектов исследования равно 6, 3 из которых (1, 2 и 3-й) относятся к норме и 3 (4, 5 и 6-й) – к патологии. Таким образом, была сформирована таблица «объект-свойство», вид которой представлен в табл. 2.

Таблица 2

Вид таблицы «объект-свойство»

№ пациента	Пр.1	Пр.2	Пр.3	Пр.4	Пр.5	Пр.6	Пр.7	Пр.8
1	2,44	10740,18	0,14	0,86	1,80	159024,76	-0,41	1,41
2	4,64	17688,92	-0,12	1,12	12,08	216409,96	-0,16	1,16
3	20,00	21791,22	-0,26	1,26	-19,82	120287,70	-0,44	1,44
4	13,59	44862,78	-0,49	1,48	4,36	62297,86	-0,11	1,11
5	30,31	156080,09	-0,13	1,13	-11,05	10493,33	-0,25	1,25
6	10,37	20588,75	-0,25	1,25	-20,59	65859,11	-0,22	1,21

Далее все признаки были преобразованы таким образом, чтобы они попали в гиперкуб $[-1;1]$, согласно процедуре:

$$\tilde{x} = 2 \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1.$$

После процедуры кодирования таблица «объект-свойство» поступает на вход адаптивной процедуры нечеткой кластеризации данных [9], которая, основываясь на степени «похожести» одного объекта на другой, формирует из данных однородные группы – кластеры, причем форма этих кластеров

варьируется в зависимости от решаемой задачи. На рис. 1 представлен результат кластеризации таблицы «объект свойство». На рисунке жирными точками отмечены центры кластеров, кружочками отмечены объекты, принадлежащие одному кластеру, звездочками – принадлежащие второму кластеру. Из рис. 1 хорошо видно, что объекты образуют кластеры эллипсовидной формы, которые в многомерном пространстве приобретают форму гиперэллипсоидов.

Также в ходе работы алгоритма была получена матрица уровней принадлежности каждого из объектов определенному кластеру, представленная в табл. 3.

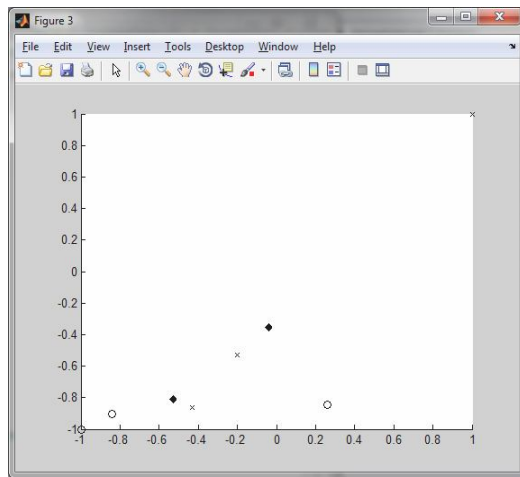


Рис. 1. Результат кластеризації таблиці «об'єкт-властивість»

Видно, що перших 3 пацієнта належать до першого класу (Норма), а останніх 3 – до другого (Патологія), що відповідає вихідним даним.

Таблиця 3

Матриця степенів приналежності

	1	2	3	4	5	6
Норма	0,88	0,60	0,55	0,20	0,30	0,20
Патологія	0,12	0,40	0,45	0,80	0,70	0,80

Висновки

Розроблений підхід дозволяє піддавати обробку медичні дані, що складаються з частини або повністю з дискретних часових рядів. При цьому для запропонованої процедури не потрібно наявності навчальної вибірки, що і було перевірено на даних риноманометричного обстеження.

Список літератури

1. Thulesius H.L. *Rhinomanometry in clinical use. A tool in the septoplasty decision making process.: doctoral dissertation, clinical sciences* / H.L. Thulesius – 2012. – 67 p.
2. Hilberg O. P.A.R. *The objective assessment of nasal patency* / O. Hilberg, P.A. Clement, A.S. Jones, D.E. Phillips,

F.J.M. Hilgers // *Diseases of the head and neck, nose and throat*. – 1998. – P. 719-742.

3. Clement P.A. *Committee report on standardization of rhinomanometry* / P.A. Clement // *Rhinology*. – 1984. – №22(3). – P. 151-155.

4. Canbay E.I. *A comparison of nasal resistance in white, Caucasians and blacks* / E.I. Canbay, S.N. Bhatia // *Am J Rhinology*. – 1997. – № 11. – P. 73-75.

5. Crouse U. *Effects of age, body mass index, and gender on nasal airflow rate and pressures* / U. Crouse, M.T. Laine-Alava // *Laryngoscope*. – 1999. – № 109. – P. 1503-1508.

6. Ерохин А.Л. *Выбор диагностического параметра при выполнении риноманометрических измерений* / А.Л. Ерохин, И.П. Захаров, А.С. Нечипоренко, И.Г. Перова, О.Г. Гартюк // *Метрологія та прилади*. – 2014. – №1 II(45). – С. 66-69.

7. Нечипоренко А.С. *Технические аспекты риноманометрии* / А.С. Нечипоренко // *Восточно-Европейский журнал передовых технологий*. – 2013. – Т.4, № 9(64). – С. 11-14.

8. Чурюмова И.Г. *Применение методов нечеткой кластеризации для анализа медицинских данных в режиме реального времени* [Текст] / И.Г. Чурюмова, Н.П. Мустецов // *Электроника и связь. Тематический выпуск «Проблемы электроники»*. – 2007. – Ч. 2. – С. 118-121.

9. Перова И.Г. *Адаптивная обработка данных медико-биологических исследований методами вычислительного интеллекта* [Текст] / И.Г. Перова // *Восточно-Европейский журнал передовых технологий*. – 2014. – 1/4 (67). – С. 24-28.

10. Тесленко Н.А. *Нечеткая кластеризация массивов биомедицинских данных в условиях избыточности информации* [Текст] / Н.А. Тесленко, И.Г. Чурюмова // *Бионика интеллекта*. – 2006. – №1 (64). – С. 92-95.

11. O. Nelles *Nonlinear System Identification: from classical approaches to neural networks and fuzzy models*, Springer, 2001.

12. Айвазян С.А. *Прикладная статистика. Классификация и снижение размерности* / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.

13. Bishop Christopher. *Pattern recognition and machine learning* / Bishop Christopher. – Berlin: Springer. 2006. – ISBN 0-387-31073-8.

14. Бокс Дж. *Анализ временных рядов* / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 406 с.

Поступила в редколлегию 31.10.2014

Рецензент: д-р техн. наук, проф. Е.А. Винокурова, Харьковский национальный университет радиоэлектроники, Харьков.

ОЦІНКА СТАНУ ХВОРИХ З ПОРУШЕННЯМ ФУНКЦІ НОСОВОГО ДИХАННЯ ЗА ДОПОМОГОЮ НЕЧІТКОЇ КЛАСТЕРИЗАЦІЇ

І.Г. Перова

У статті розглянуто підхід, що дозволяє піддавати процедурі нечіткої кластеризації дискретні часові ряди шляхом їх перетворення в вид таблиці «об'єкт-властивість». Перетворення полягає в заміні дискретних тимчасових відліків параметрами моделі авторегресії другого порядку. Для перевірки адекватності моделі була проведена її апробація на дискретних часових рядах показників риноманометрії, таких як диференціальний тиск між хоаною і підмасочним простором і витрата повітряного потоку, які були зареєстровані синхронно з частотою дискретизації 200 Гц.

Ключові слова: нечітка кластеризація, авторегресія, риноманометрія.

EVALUATION OF PATIENT CONDITIONS WITH IMPAIRED NASAL BREATHING BY FUZZY CLUSTERING

I.G. Perova

The article describes an approach that allows the procedure to apply fuzzy clustering to discrete time series by converting them to the table view "object-property". Conversion is the replacement of discrete time samples by autoregression model parameters of the second order. Testing of discrete time series of rhinomanometry indicators, such as differential pressure between choana and space under the mask and air flow, which were recorded synchronously with the sampling frequency of 200 Hz was carried out to test the adequacy of the model.

Keywords: fuzzy clustering, autoregression model, rhinomanometry.