UDC 004.93

Alina S. Nechyporenko

METHOD OF BIOMEDICAL TIME SERIES PROCESSING FOR PATHOLOGY CLASSIFICATION

This paper presents the new approach of biomedical time series processing for feature extraction based on chaos theory. Methods of nonlinear dynamics for processing of time series allow us to obtain significant features of physiological signals.Using of the F-transform for phase space diagrams approximation provides more high accuracy of classification. It has been demonstrated that signals of breathing have a fractal properties. Application of fractal analysis will allow developing new approaches forpathology identification.

Keywords: Signal processing, Feature extraction, F-transform, Approximation, Classification.

Introduction

Medical diagnosis is very complex process. It is a complex of objective and subjective methods for estimation of diagnostic parameters. Most objective methods are based on time series processing for diagnostics task.

Time series processing is a well-studied task. The data scientists usually employ two approaches: analysis of global integral statistical properties of signals or analysis of significant parts of signal [1]. A lot ofpapers use linear models, where time series are transformed in stationary series through differentiation. However, physiological processeshave high complexity, all of them are nonstationary and nonlinear. Some of them have a chaotic bechavior and many types of components, such a trends, impulses, other uncontrolled features. So,representation of physiological time series by linear model is not suitable.

In recent years, there has been a dramatic increase in the use of computation-intensive methods to analyze biomedical signals. The general approach falls under the methods of artificial intelligence or machine learning for decision-making in medicine.Such methods require a dataset ofsignificant features that will be fully representative of underlying biological processes.Several techniques using non-linear chaos features of the signal have been proposed for classification and prediction purposes in [2-4]. The list of applications includes automated electrocardiogram (ECG) or electroencephalogram (EEG) analysis for cardiovascular or neurological disorder diagnosis [5, 6]. The principal task of nonlinear approach is to examine the possible existence of chaotic properties in the signal [7].

Previous work [8] is devoted to identification of fractal properties of time series using signals of nasal breathing. These properties allow identification hidden features in time series. The main task of current paper is to find significant features of time series ofbreathing signals derived from the data itself for pathology classification task.

Materials and methods

A wide variety of methods for evaluating the parameters of nasal breathing is used in rhinology [9].All of them have common disadvantage: lack of objective criteria for differential diagnosis of rhinological-pathologies. Active Anterior Rhinomanometry remains one of the most used methods for evaluating the nasal breathing function. This method is based on simultaneous registration of two parameters: differential pressure Δp and an airflow rate Q through a nasal cavity. Rhinomanometric data were recorded by a system for rhinomanometric measurements [10]. The measurements result are the time series of the differential pressure Δp [Pa] and the airflow rate Q [cm3/s]presented in Fig.1.

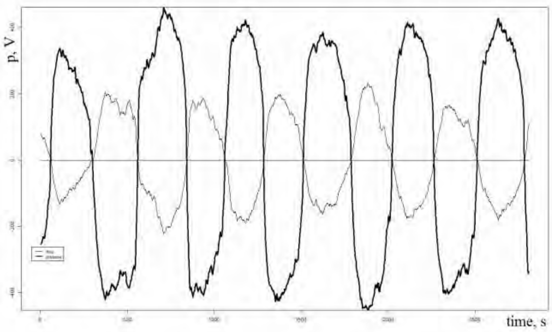

Fig.1 – The dependence of pressure and airflow rate on a time

Rhinomanometric signals are quasi periodic, nonstationary and nonlinear[8]For validation of fractal properties of signals, we propose a calculation of Hurst parameter H and fractal dimensionD according to [11]. The Hurst parameter is suggested as a measure of the degree of self-similarity of a data series because it allows the evaluation of the bursty nature of a data.Processed dataset contains 1076 measurements of rhinomanometric signals. Result of calculation is within the range of [0.16,0.20] and the range of [1.80,1.84] for H and D respectively. It means signals have a fractal properties and process is anti-persistent.

Throught this paper, we use the Rhinology laboratory of ENT Clinic "Garyuk"/laboratory of intelligent software and hardware systems of Kharkiv National University of Radio Electronics rhinologic database (http://www.garyuk.com). Signals were recorded using software/hardware system for rhinomanometric measurements "Optimus" [10]. Database consists of 1076 records for several groups of patients. There are patients with different types of rhinitis and rhinosinusitis, cysts, polyps, septal deviation and other.Each record contains 3 to 5 breathing cycles.

The initial data set obtained from database can hold noises. On the one side the usage of high-sensitive sensors is always associated with registration of noises. On the other side during measuring procedure an inadequate seal of the mask, incorrect connecting tubes between mask and sensors could appear. Therefore, a preprocessing of signals is required. Data preprocessing stage includes several techniques like data cleaning, data reduction, data transformation, which are described in detail in [12, 13]. An important task of preprocessing is to remove a signal distortion and don't remove the significant parts of signal. Preprocessing for time series represents a filtering of noises and smoothing of signals. Survey of filtering techniques for biomedical data has been presented in the works [14].

We used "RStudio" for signal processing and feature extraction. First of all we implement an additional filtration for removing of signal distortion. Secondly, we cut the measurements to variety of periods, which equal to one breathing cycle. Sample of one breathing cycle is shown in Fig. 2. Then we built a phase space diagram of the rhinomanometric signal. The phase space diagram of the rhinomanometric signal is constructed as follows : the differential function $\dot{x} = \partial x / \partial t$ is plotted on the Y-axis, and the original function x - on the X-axis of the phase plane. Phase space diagrams for norm and septal deviation are shown in Fig. 3, 4.

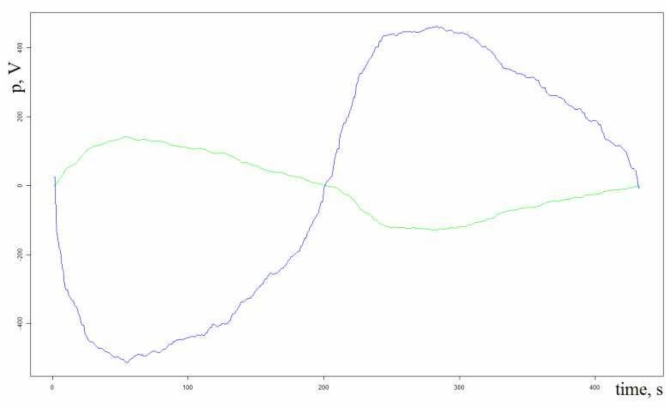

Fig. 2 – The breathing cycle

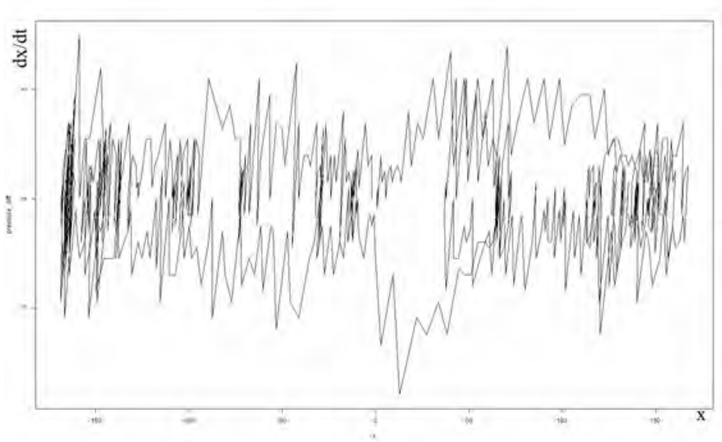

Fig. 3 – Sample of phase space diagram of differential pressure for “norm”

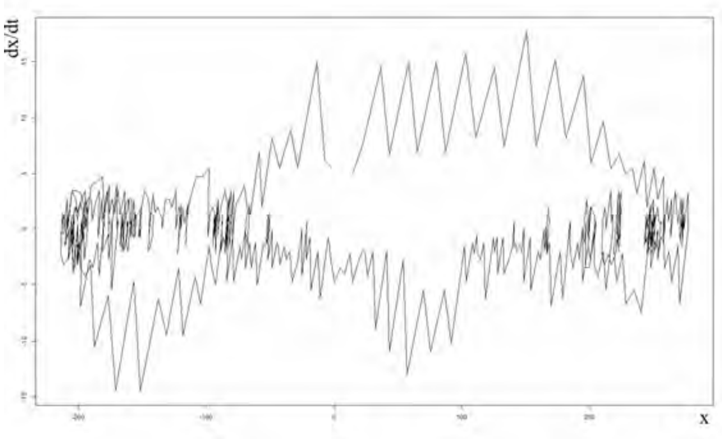

Fig.4 – Sample of phase space diagram of differential pressure for “septal deviation”

The next stage is approximation of phase space diagram using F-transform. This approachwas introduced by I. Perfilieva in [15].F-transform is good chose for time series. It has many advantages, among them excellent filtering properties and easy calculation. F-transform is stable with respect to the choice of points. It means that the resulting function will not be significantly changed.Let us remind the basic definitions and properties of F-transform.

The domain $D=[a,b]$ is partitioned by k fuzzy partition by fuzzy sets $\{f_1,\dots,f_k\}$.Let $P=\{a=t_1<t_2<\dots<t_n=b\}$ of $[a,b]$into $n-1$subintervals, where t – time of signal measurement.Each subinterval $[t_{k-1},t_k]$, $k=2,\dots,n$ and by a family $A=\{A_1,A_2,\dots A_n\}$of $n$fuzzy numbers (the basic functions), identified by their membership functions $A_1(t),A_2(t),\dots,A_n(t)$for $t\in[a,b]$and with the properties:

1. A_i is continuous on D;

2. A_i - strictly increase on $[t_{i-1},t_i]$ and strictly decrease on $[t_i,t_{i+1}]$

3. $A_i:[1..N]\to[0,1], A_i(t_i)=1$;

4. $A_i(t)=0$, if $t\notin(t_{i-1},t_{i+1})$, and assume that $t_0=t_1=1, t_{n+1}=t_n=N$

5. $\sum A_i(t)=1$ for all $t\in[1..n]$.

After that time series are transformed into k values $[F_1,\dots F_k]$ using (1):

$$F_j=\frac{\sum X_{t_i}A_j(t_i)}{\sum A_j(t_i)}, i=1,\dots,k \tag{1}$$

We have used triangle base function as $A(t)$for each of fuzzy partition.

This kind of data received from human breathing, and may have different number of values in each data measurement. It related with different time of breathing cycle for each person. On other side we have a lot of machine learning approaches which require fixed count of input features for the each measurement. We propose to set count of fuzzy values equivalent to K and receive length of fuzzy partition (N) from this equation: $N=floor\left(\frac{L}{K}\right)\sqrt{b^2-4ac}$, where L – number of data in current measurement. We useobtained K number of F-components as input data features for machine learning methods, for example SVM and Random Forest. For the comparing of approximation results we have used FFT and moving average methodes.

Let $i = 1..L$, where L - amount of calculated values of phase diagram. Represent each point with coordinates (x_i, y_i) of the shape as a complex number $z_i = x_i + y_i j$, where x_i as real part and y_i as imaginary part.

We will apply Discrete Fourier transform (DFT) to the vector $Z = z_1 ... z_N$ using method from [16]. As result we have received components $F = [F_0, F_1, ..., F_{N-1}]$.

Let the K - count of pairs of Fourie components F, which will be used in reduced Fourie component list $F_r : F_r = [F_0, F_1, F_2 ... F_K, F_{N-K}, ..., F_{N-2}, F_{N-1}]$

If we perform Inverse Discrete Fourie Trasform (IDFT) to this reduced components list, we will receive approximated representation of the phase diagram. It allows creating visualization.

Experiments and results

The classifiers implemented in this research are Support Vector Machine (SVM) [17] and Random Forest approach (RF) [18].Set of features consists of phase space diagrams. Phase space diagrams were approximated using F-transform and FFT.Amount of elements ofinput data set is 1076 measurements, which were classified by otolaryngologists to the ‘norm’ and ‘pathology’ classes.Each measurement stores information about differential pressure and airflow rate.

For each classification method the set of features has been performed such that the optimal classification results are achieved.

We propose to compare a techniques based on F-transform and FFT for approximation of obtained phase space diagrams. Grafical visualization of implemented approximation techniques is shown on Fig. 5.

 ISSN 1562-9945

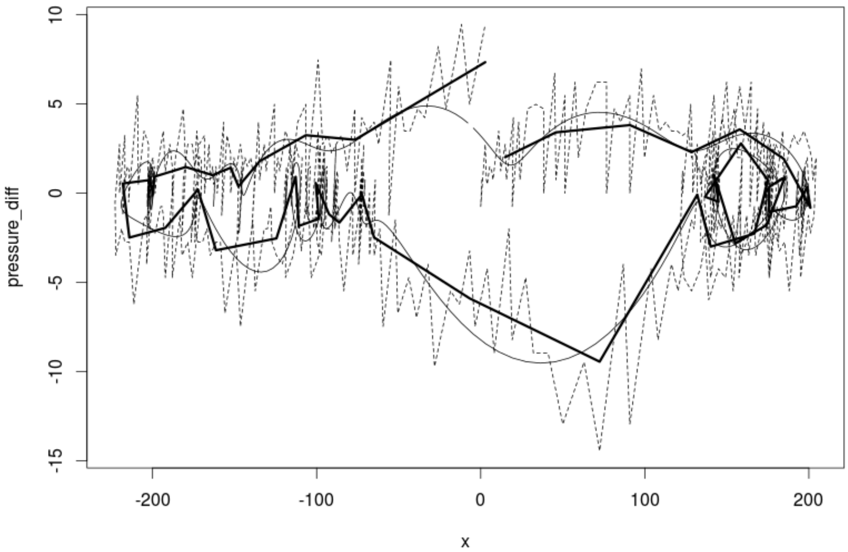

Fig. 5 - Result of approximation of phase space diagram: bold curve – F-transform approximation; faint curve – FFT approximation

The mean absolute error of approximation for FFT approach is 1.26, for F-transform approach is 0.63.Learning set takes 85% from all amount, test set takes 15% from all amount of measurements. Approximation result should be smooth or more close to initial data depends on value of the K variable. This value is discreet. Selection of this value should be preformed to receive best learning values using learn/test/validation data set. K valuewas selected equal to 43.

Results of classification are presented in Table 1.

Table 1

Error rates for different learning methods

Method	Learning RF	Learning SVM	Test RF	Test SVM
FFT	4,4	3,6	11,8	11,7
F-transform	5,0	4,3	10,53	10,53

Conclusions

The paper demonstrates a potential of using the methods of nonlinear dynamics for processing of biomedical time series.Theproposed method based on phase space diagram approximation to finding significant features of rhinomanometric signals. Approach is based on usage of the F-transform approximation of time seriesfor generation of fixed amount of initial features of signals. It returns the value of mean absolute error less than the FFT.Supervised learning algorithms SVM and

RF were used for classification on two classes 'norm' and 'pathology'. The classification accuracy with using F-transform is higher than with using FFT. Future investigation can be related with usage of 2D fuzzy transformation for approximation of rhinomanometric signals.

REFERENCES

1. Bogunovic N. Processing and analysis of biomedical nonlinear signals by data mining methods / N. Bogunovic, A. Jovic // Proceedings of 17th Conference on Systems, Signals and Image Processing IWSSIP, 2010.- P. 276-280.
2. Costa M. Multiscale entropy analysis of complex physiologic time series / M. Costa, A.L. Goldberger, C.K. Peng // Phys. Rev. Lett, no.6 - 2002. – P.89.
3. Wessel N. at al. Short-term forecasting of life threatening cardiac arhytmias based on symbolic dynamics and finite-time growth rules, Phys. Rev. no. 61(1) – 2000, P. 733-739.
4. Braun C. et al. Demonstration of nonlinear components in heart rate variability of healthy persons, Am. J. Physiol. Heart Circ. Physiol., no. 275. – 1998, P. 1577-1584.
5. Fainzilberg L.S. Computer amalysis and recognition of cognitive space electro-cardio graphic image / L.S. Fainzilberg, T.P. Potapova // Proc. of the 6th Intern. Conf. on Computer analysis of Images and Patterns (CAIP'95) Prague, - 1995. – P. 668 -673.
6. Saprykin G.Using the characteristics of chaotic dynamics in the knowledge base of decision support system / G. Saprykin, J. Kobytska, L. Kirichenko //Proc. of the International Conference on Computer Sciences and Information Technologies, Lviv, Ukraine, 14-17 September, - 2015. – P. 162-165.
7. Fraktal'nyj analiz processov, struktur i signalov, R.И. Paŝenko. – Har'kov: НООО NИО «Иkoperspektiva», 2006. – 348 s. (in Russian).
8. Nechyporenko A.Rhinomanometric signal processing for selection of formalized diagnostic criterion in rhinology / A. Nechyporenko // Telecommunications and Radio Engineering, Vol 74,no. 14, 2015. - P. 1285-1294.
9. Wallace D. The diagnosis and management of rhinitis: an updated practice parameter / D. Wallace, M. Dykewicz, D. Berstein at al. // J Allergy Clin Immunol, no 122, - 2008. – P. 1–84.

 ISSN 1562-9945

10.Yerokhin A. A New Intelligence-Based Approach for Rhinomanometric Data Processing / A. Yerokhin, A. Nechyporenko, A. Babii, O. Turuta // 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 19-21 April, - 2016. – P. 198-201.
11.Yerokhin A. Usage of phase space diagram to finding significant features of rhinomanometric signals / A. Yerokhin, O. Turuta, A. Babii, A. Nechyporenko, I. Mahdalina // 2016 IEEE XI International Scientific and Technical Conference on Computer Science and Information Technologies, Lviv, Ukraine, 6-10 September, - 2016. – P. 198-201.
12.Data mining:concepts and techniques. Second edition,J. Han, M. Kamber, Morgan Kaufmann, Elsevier Inc. – 2006. – P. 137.
13.Data mining: practical machine learning tools and techniques – 2nd edition. I. H. Witten, E. Frank, - Amsterdam etc.: Morgan Kaufman. – 2005. – P. 214
14.Goras M. Fira Preprocessing method for improving ECG signal classification and compression validation / M. Fira Goras // PHYSCON 2009, Catania, Italy, September, 1–September, 4, - 2009, P.114-119.
15.Perfilieva I. Fuzzy transformation and its applications / I. Perfilieva and E. Chaldeeva // 4th Czech - Japan Seminar on Data Analysis and Decision Making under Uncertainity, - 2001. – P. 116–124.
16.Fundamentals of signals and systems using Mathlab,E. Kamen, B. Heck, -*N.Y.* Prentice Hall, 1997. – 219.
17.Data mining: practical machine learning tools and techniques – 2nd edition”, I. H. Witten, E. Frank, - Amsterdam etc.: Morgan Kaufman. – 2005. – P.139.
18.Data preparation for data mining, D. Pyle, San-Francisco etc.: Morgan Kaufmann, - 1999, P.217.