Recurrence Quantification Analysis: A Promising Method for Data Evaluation in Medicine

Jakub Schlenker\textsuperscript{1}, Tomáš Nedělka\textsuperscript{1,2}, Lucie Riedbauchová\textsuperscript{3}, Vladimír Socha\textsuperscript{1}, Karel Hána\textsuperscript{1}, Patrik Kutílek\textsuperscript{1}

\textsuperscript{1} Czech Technical University in Prague, Faculty of Biomedical Engineering, Kladno, Czech Republic
\textsuperscript{2} Department of Neurology, University Hospital Motol and 2\textsuperscript{nd} Faculty of Medicine of the Charles University, Prague, Czech Republic
\textsuperscript{3} Clinic of Cardiology, University Hospital Motol and 2\textsuperscript{nd} Faculty of Medicine of the Charles University, Prague, Czech Republic

Abstract

Introduction: This paper describes principles of a promising method for medical data analysis called recurrence analysis, which is based on the chaos theory that better describes processes in the living organisms.

Methods: Phase space reconstruction and recurrence analysis are explained in brief. The main part of the work focuses on recurrence plot, which is a basic tool for recurrence analysis.

Conclusion: Possible clinical applications of recurrence analysis in the field of medicine are discussed and a short overview of pilot projects performed by our team on this topic is given.

Keywords
Nonlinear analysis, recurrence analysis, recurrence plot, phase space, heart rate variability, autonomic dysfunction.

Correspondence to:
Jakub Schlenker
Joint Department of Biomedical Engineering CTU and Charles University in Prague, Czech Republic
Address: Studničkova 7/2028, 128 00 Prague 2, Czech Republic
E-mail: jakub.schlenker@gmail.com

©2014 EuroMISE s.r.o.
EJBI – Volume 10 (2014). Issue 1

1 Introduction

Usage of nonlinear methods for data analysis is becoming increasingly popular in medicine due to the fact that they seem to be able to describe selected processes occurring in living organism more effectively than it is these days \cite{1}.

To increase the probability of full recovery or to minimize the health damages, it is important to detect diseases in their early or even in their subclinical stages. Since specific methods of nonlinear analysis seem to be sensitive enough to uncover these early phases of the disease development, their application in the data analysis may improve health care and help the physicians to understand better the physiological and pathophysiological processes occurring in the human body.

One such nonlinear method recently applied in medicine is the recurrence analysis. Method of so called recurrence analysis is derived from the chaos theory which describes the basic dynamics of a system with chaotic behaviour that can be found in every biological system \cite{2}. Recurrence analysis has been successfully used in pilot projects in cardiology \cite{3,4} and neurology \cite{5,6,7,8,9} where it was mainly used to describe dynamics of the heart rate and blood pressure regulation. These physiologic variables are under permanent control of the autonomic nervous system which may be viewed as an example of nonlinear deterministic system since the autonomic nervous system instantaneously changes its tone based on the actual demands and needs of the organism \cite{11,2}. Impaired function of the autonomic nervous system is therefore often associated with reduced variability of functions that this system controls, i.e. reduced heart rate variability. Due to this fact, the system (heart rate) tends to recur to a similar state and exhibit only limited changes in response to outer inputs when its control through the autonomic nervous system is harmed.

The effort of our team has recently focused on the evaluation of possible role of the recurrence analysis in the diagnostic of various diseases in their early phases (diseases origin of which is associated with the autonomic dysregulation) especially in the field of neurology \cite{9,10} and cardiology.

2 Methods

Recurrence analysis is one of the non-linear analysis of data derived from chaos theory. Recurrent graphs, ba-
drial tool recurrent analysis allow to visualize repetitive behaviour of non-linear analysis of short-term and non-stationary data [12].

- The first step of analysis, as with most nonlinear techniques [12], is the reconstruction of the phase space.
- The second step is the formation of recurrent plot using the threshold distance.
- The last step in the analysis is to calculate the measures of recurrent plot.

2.1 Phase Space Reconstruction

The $n$ state variables in time form vector (trajectory) in $n$-dimensional space (phase space). Trajectory in phase space represents all possible states of a system. Each state of the system corresponds to a specific point in the space phase. For $N$ state variables $N$-dimensional phase space is created. However, very often it is not possible to observe more than one state variable of a system in the field of clinical medicine, because they are not known or it is difficult to measure them [12]. Nevertheless we can reconstruct the phase space trajectory from a single observation by using the Takens’ theorem [12]. Takens’ theorem is one of the most commonly used method for reconstruction of the phase space trajectory [12,13]:

$$x_i = (y_i, y_{i+\tau}, \ldots, y_{i+(m-1)\tau}),$$

where $m$ is the embedding dimension, $\tau$ is the time delay and $y_i$ is a single observation.

An optimal set of embedding dimension and time delay is important for reconstruction of phase space that fully describes the system dynamics [14]. One of the frequently used approaches for choosing time delay is the first minimum of mutual information [15]. The mutual information measures mutual dependence of two variables $A$ and $B$ and can be defined using entropy as [15]:

$$I(A, B) = H(A) + H(B) - H(A, B),$$

where $H(A)$ and $H(B)$ are the entropies and $H(A, B)$ is the joint entropy of $A$ and $B$.

Entropy is generally, expressed in terms of a discrete set of probabilities $s_i$ as [13]:

$$H(S) = -\sum_{i=1}^{n} p_s(s_i) \log p_s(s_i),$$

where $H(S)$ is entropy of system $S$, which consists of a set of possible messages $(s_1, s_2, ..., s_n)$ and the associated probabilities $p_s(s_1), p_s(s_2), ..., p_s(s_n)$.

False nearest neighbour method (FNN) is often used for optimal embedding dimension setting. Embedding dimension is a parameter which tells us how many dimensional phase space will be reconstructed. FNN method is based on the fact that choosing of low embedding dimension causes crossing of phase space trajectory [14]. We use modified FNN method using ratio of Euclidian distances of two neighbour’s states in $m$ and $m+1$ dimensional space by Cao [16]:

$$a(i,m) = \frac{\| y_i(m+1) - y_{n(i,m)}(m+1) \|}{\| y_i(m) - y_{n(i,m)}(m) \|},$$

where $\| . \|$ is Euclidian distance, $y_i(m)$ is $i$-th reconstructed vector with embedding dimension $m$, $y_{n(i,m)}(m)$ is the nearest neighbor of $y_i(m)$.

The mean value of all $a(i,m)$ is introduced by Cao [16]:

$$E(m) = \frac{1}{N - m \tau} \sum_{i=1}^{N-m\tau} a(i,m).$$

The average $E(m)$ is dependent only on dimension $m$ and delay $\tau$. Finally, variable $E_1(m)$ is used to investigate variations from dimension $m$ to dimension $m+1$ [16]:

$$E_1(m) = \frac{E(m+1)}{E(m)}.$$ 

Value of $E_1(m)$ stops changing when dimension $m$ is greater than the dimension of attractor $m_0$.

2.2 Recurrence Analysis

Recurrences in the nature were observed a long time ago [12], but the method of recurrence analysis was introduced just after the development of computer science in 1980s [19]. The basic tool of recurrence analysis is recurrence plot (RP) see Fig. 1 which can visualize recurrences in system dynamics using two-dimensional graph [19]. RP shows all moments in times when the phase space trajectory of the dynamical system visited roughly the same area in the phase space [17]. It is possible to visualise multidimensional phase space using two-dimensional graph (RP). We can demonstrate recurrences in a dynamical system, find interrelations between several systems or detect transitions between different states with RPs [6,12,17]. Mathematical expression of RP see equation [1]

$$R_{i,j} = \Theta(\epsilon - \| x_i - X_j \|), \text{ for } i, j = 1, 2, \ldots, N, \quad (1)$$

where $N$ is the number of states, $\epsilon$ is a threshold distance, $\| . \|$ a norm and $\Theta(.)$ the Heaviside function.
Figure 1: On the top figure, there is the input signal for recurrence analysis. The input signal is represented by the length of R-R intervals. On the bottom figure, there is the recurrence plot. On both figures we can distinguish two phases of the orthostatic test, resting in supine position and stand-up position, each for 5 minutes. In the second phase, after the verticalisation of the patient, we can recognise lower heart rate variability.

The most important parameter of recurrence analysis is the threshold distance. If the distance between two states on the phase space trajectory is smaller than a given threshold, the recurrence point in RP arises. Value of the point in recurrence matrix is one; otherwise it is zero [12].

This technique is called pair test and it is a pairwise test of all states (for $N$ states we compute $N^2$ tests) [12].

With closer looking at the RP we can see single recurrence points, diagonal lines, vertical lines and horizontal lines. The structures formed by these elements are basis for recurrence quantification analysis (RQA). This analysis was introduced by Zbilut and Weber [19, 20, 21] to evaluate RPs quantitatively. RQA is a set of measures derived from diagonal and vertical structures of RP [17].

Diagonal line occurs when the trajectory visits the same area of the phase space at different times, it is when the system returns to the same or similar states at the different time points. Horizontal and vertical lines indicate how long the system stays in given state which does not change at all or just very slowly [17]. Single isolated recurrence point can represent a rare state [17].

The Recurrence rate is density of recurrence points in recurrence plot [17]:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}.$$

This value corresponds to probability that a specific state will recur.

Determinism is a measure which represents the percentage of recurrence points which form diagonal lines [17]:

$$DET = \frac{\sum_{l=l_{\text{min}}}^{N} l P(l)}{\sum_{i,j} R_{i,j}},$$

where $P(l)$ is the histogram of the lengths $l$ of the diagonal lines.

Laminarity is the amount of recurrence points which form vertical lines. This variable is related with amount of laminar states [17]:

$$LAM = \frac{\sum_{v=v_{\text{min}}}^{N} v P(v)}{\sum_{v=1}^{N} v P(v)},$$

where $P(v)$ is the histogram of the lengths $v$ of the vertical lines.

The average length of vertical lines is called Trapping time $TT$ and it is related with laminarity time. This value contains information about frequency and length of laminar states [17]:

$$TT = \frac{\sum_{v=v_{\text{min}}}^{N} v P(v)}{\sum_{v=v_{\text{min}}}^{N} P(v)}.$$

The low value of $LAM$ and $TT$ indicate high complexity of a system [6].

The Longest diagonal line $L_{\text{MAX}}$ is length of the longest diagonal line [17]:

$$L_{\text{MAX}} = \max(l_i; i = 1 \ldots N_i).$$

Divergence $DIV$ is the inverse of $L_{\text{MAX}}$ and it is related with Kolmogorov-Sinai entropy [17]:

$$DIV = \frac{1}{L_{\text{MAX}}}.$$

Other measures of RQA are average length of diagonal lines ($AVDL$), ratio ($RATIO$) between $DET$ and $RR$, Shannon entropy ($ENTR$) and maximal length of vertical line ($V_{\text{MAX}}$) can be found in [17].
The advantages of RP include the ability to capture the chaotic properties of a system without the need of a long data series and the fact that it is relatively immune to noise and nonstationarity [17]. RQA is a sensitive tool for detecting any dynamic changes; however, it can be easily affected by inappropriate settings especially setting of the threshold distance where even a small change can dramatically affect the results of RQA [22, 23]. There are several methods for threshold distance setup [17, 22, 23]. One of the widely used methods is setting of the threshold distance as a percentage of the maximum distance in phase space. Furthermore, there is setting where values should not exceed 10% or 15% of the average or maximum distance in phase space [24, 18]. Quite frequently used method is called a fixed percentage of recurrent points. This means that we set a threshold distance value that guarantees the exact percentage of recurrent points [6]. Often is this value 1% [22, 24], but we can also find other values, such as 5% [6, 9, 25]. Selecting the optimal threshold distance appears to be a task for data mining [26].

3 Results

Our team currently evaluates possible applicability of the recurrence analysis in the field of medicine, specifically in neurology and cardiology. Our pilot study was focused on the early detection of the autonomic dysfunction in subclinical stages of diabetic cardiovascular neuropathy [9]. We found a significant increase in measures $DET$, $LAM$, $LMAX$ and $TT$ in diabetic patients compared to control group. Based on the promising results of this pilot study, a bigger project on this topic has started, the aim of which is to confirm high sensitivity of the recurrence analysis in the early detection of diabetic autonomic dysfunction and to identify typical values of RQA measures both in patients with diabetes and manifest autonomic dysfunction and in the healthy control group. One of the goals is also the identification of optimal value of threshold distance for diagnosis of autonomic neuropathy. For now we reached promising results using a fixed percentage of recurrent points and our experimental setup using standard deviation:

$\epsilon = \frac{0.02}{od}$,

where $\epsilon$ is threshold distance, $od$ is a standard deviation of input signal.

This study is a joint project of Faculty of Biomedical Engineering CTU and University Hospital Motol.

Next application of recurrence analysis in the field of neurology is the assessment of vasovagal syncope. Results of a pilot project on this topic were recently presented by our team at the international conference in Slovakia [10]. We found reduced complexity of the heart rate control in young patients with vasovagal syncope using the recurrence analysis. RQA measurement showed significant differences in $DET$, $LMAX$, $DIV$ and $ENTR$ see Fig. 2. However, future studies are needed in this particular field because of the limitations of our study - no gender match and small number of subject in groups.

Another cardiologic application of recurrence analysis that we are working on is the prediction of onset and termination of paroxysmal atrial fibrillation. It is a similar project to that of Mohhebi and Ghassemian [4] with the difference that we use real patient’s Holter monitor data recorded at the clinic of cardiology in the University hospital Motol. We hope, that our project will confirm capability of recurrence analysis in this field.

Currently a perspective applications of recurrence analysis focused on evaluation of psycho-physiological condition of pilots [27, 28]. When elevating pilots condition in stress situations, a change in physiological functions takes place, such as the change in heart rate, respiratory rate, blood pressure etc. The preliminary results show high potential of RQA for data evaluation of this type of measurement.

Figure 2: Boxplot showing the comparison between patients with syncope and the control group. A – Determinism ($p = 0.04225$). B – Divergence ($p = 0.02612$). C – The length of the longest diagonal line ($p = 0.006$). D – The Shannon entropy ($p = 0.034$) [10].

Boxplots show maximum, minimum, median, first and third quartile, cross points are outliers.
4 Conclusion

Detecting diseases in their early or even subclinical stages is important for increasing the probability of full recovery. Nonlinear methods seem to be useful for describing selected processes occurring in the living organism effectively. One of such methods is the recurrent analysis. Our pilot studies showed high potential of its use in medicine. However, there is a question about optimal setup of threshold distance and normative data are needed and results of pilot studies need to be confirmed in larger trials yet.

Acknowledgements

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant no. SGS14/169/OHK4/2T/17 and no. SGS14/170/OHK4/2T/17.

References


[27] Kaľavský P., Socha L.. Výskum tréningových metód pilotov s
využitím leteckých. In: FLYIN: aviation industry exhibition :
FLYIN letecká konference VŠB-TUO 2013 : sborník příspěvků.
1st. ed. Ostrava: Vysoká škola báňská - Technická univerzita

[28] Socha V., Szabo S., Socha L., Kutílek P. 2014, "Evalua-
tion of the variability of respiratory rate as a marker of
stress changes", Proceedings of 18th International Conference.
Transport Means, KAUNAS UNIVERSITY OF TECHNO-
LOGY, 23-24 October 2014 (in press)