Compound method of time series classification

Łukasz Korus\textsuperscript{a}, Michał Piórek\textsuperscript{b}

\textsuperscript{a}Department of Control Systems and Mechatronics, Wroclaw University of Technology, Janiszewskiego 11/17, 50-372 Wroclaw, Poland lukasz.korus@pwr.edu.pl
\textsuperscript{b}Department of Computer Engineering, Wroclaw University of Technology, Janiszewskiego 11/17, 50-372 Wroclaw, Poland michal.piorek@pwr.edu.pl

Received: July 13, 2014 / Revised: January 17, 2015 / Published online: September 9, 2015

Abstract. Many real phenomena preserve the properties of chaotic dynamics. However, unambiguous determination of belonging to a group of chaotic systems is difficult and complex problem. The main purpose of this paper is to present compound method of time series classification which is basically directed to the detection of chaotic behaviors. The method has been designed for differentiation of three types of time series: chaotic, periodic and random. Our approach assumes, that more reliable information about the dynamics of the system will provide the compilation of several methods, than any individual. This paper focuses on choosing a good set of methods and analysis of their results. In our investigation, we used the following methods and indicators: time delay embedding, mutual information, saturation of system invariants, the largest Lyapunov exponent and Hurst exponent. We checked the validity of the methods applying them to three kinds of basic systems which generate chaotic, periodic and random time series. As a summary of this paper, all selected methods and indicators computed for generated times series have been summarized in the table, which gives the authors a possibility to conclude about type of observed behavior.

Keywords: deterministic chaos, time series analysis, Takens theorem.

1 Introduction

In 1981, Floris Takens presents his theorem about Embedding Nonlinear Dynamical systems, which conceive an opportunity to reconstruct attractors of chaotic dynamical system. Reconstruction is based on a time series given by observations of one state of dynamical system [37].

Over the next few years many of embedding approaches and theirs applications were described, for example, by Sauer, Yorke and Casdagli [34], Stark applied Taken’s theorem to Forced Systems [36]. Abarbanel described the analysis of observed chaotic data in physical systems [1].
There are also many papers about observing chaos in real systems [4, 5, 6, 12, 18, 19, 21, 22, 26, 28, 32, 38, 41, 42] and about controlling chaos [3, 8, 15, 25, 31, 39].

Such a large number of methods from listed papers proves, that the field of analysis of chaotic systems is heavily explored. This is due to the difficulty and complexity of the subject.

The approach described in this article was the result of previous experimental studies. In Section 2, a set of time series analysis methods were chosen and described. Selection was preceded by a review of the literature. Section 3 describes a way for generating test time series for further experiments. Section 4 presents the numerical results of experiments carried out on the generated time series. Conclusions are described in Section 5.

2 Set of methods

In our investigations, methods described below were used.

2.1 Time delay embedding

Many of real phenomenon are the results of nonlinear systems dynamics evolution [1]. Observation of such kind of systems provides many difficulties due to the fact that only a limited set of information is accessible outside as a measurement of time series. The basic assumption of this paper is that one time series is formed by samples measured in the output of the system

\[ s(n) = x(t_0 + n\Delta t), \]  

where \( s(n) \) is the value of \( n \)th sample of measurement of physical process \( x \) in time \( t_0 + n\Delta t \) and \( t_0 \) is initial time. According to Takens’ embedding theorem, it is possible to reconstruct the state trajectory from single time series using below algorithm:

\[ y(n) = [s(n), s(n + T), \ldots, s(n + (d - 1)T)], \]  

where \( T \) is time delay and \( d \) is embedding dimension, which estimates a real dimension of the observed system. The main point of the state space reconstruction method is \( T \) and \( d \) estimation. To estimate time delay \( T \), an average mutual information \( I \) has been used, while for embedding dimension the saturation of system invariants method [1].

2.2 Time delay estimation: average mutual information method

There are three types of criterions of time delay \( T_d \) selection [29]:

- series correlation approaches (autocorrelation, mutual information [17] or high-order correlations [2]),
- approaches of phase space extension (fillfactor [11], wavering product [10] or average displacement [14]),
- multiple autocorrelation and nonbias multiple autocorrelation [27].
For the purposes of this paper, mutual information method was selected. This approach is based on information theory and transformation of linear autocorrelation to nonlinear systems. More precisely, this method consists of 2-dimensional adaptive histogram.

Let’s assume that there are two nonlinear systems: $A$ and $B$. Outputs of these systems are denoted as $a$ and $b$, while values of these outputs are represented by $a_i$ and $b_k$. Mutual information factor describes how many bits of $b_k$ could be predicted, where $a_i$ is known:

$$I_{AB}(a_i, b_k) = \log_2 \frac{P_{AB}(a_i, b_k)}{P_{A}(a_i)P_{B}(b_k)}. \quad (3)$$

Here $P_A(a_i)$ is probability that $a = a_i$ and $P_B(b_k)$ is probability that $b = b_k$ and $P_{AB}(a_i, b_k)$ is join probability that $a = a_i$ and $b = b_k$. Average mutual information factor can be described by

$$I_{AB}(T) = \sum_{a_i, b_k} P_{AB}(a_i, b_k) I_{AB}(a_i, b_k). \quad (4)$$

In order to use this method to assess correlation between different samples in the same time series, the average mutual information factor is finally described by the equation

$$I(T) = \sum_{n=1}^{N} P(S(n), S(n + T)) \log_2 \frac{P(S(n), S(n + T))}{P(S(n))P(S(n + T))}. \quad (5)$$

Fraser and Swinney [17] propose that $T_{m \text{ lag } T_d}$, where the first minimum of $I(T)$ occurs as a useful selection of time. This selection guarantees that the measurements are somewhat independent, but not statistically independent. In the case of absence of the average mutual information clear minimum, this criterion needs to be replaced by choosing $T_d$ as the time for which the average mutual information reaches $4/5$ of its initial value:

$$\frac{I(T_d)}{I(0)} \approx \frac{4}{5}. \quad (6)$$

2.3 Embedding dimension estimation: saturation of system invariants

As it was mentioned, delay coordinates are used to construct $d$-dimensional vector to state space reconstruction. Takens theorem guarantees that if chosen embedding dimension is large enough, properties of the attractor of dynamical system will be the same when computed on lagged coordinates and when computed in the physical coordinates. The main goal of state space reconstruction is to provide a Euclidean space $\mathbb{R}^d$ large enough so that the set of points of dimension $d$ can be unfolded without ambiguity. When all ambiguities are resolved, one says that the space $\mathbb{R}^d$ provides an embedding of the attractor of dynamical system and dimension $d$ is an embedding dimension $d_E = d$.

There are several methods of selecting embedding dimension. One of them is the method of false neighbors [9]. Very promising seems to be neural network method [30]. In this paper, saturation of system invariants method was used [1].
For properly reconstructed attractor, every its property depending on distances between points in the state space should become independent of the value of the embedding dimension once the large enough $d_E$ has been reached. Appropriate necessary embedding dimension can be established by computing such property for $d_E = 1, 2, \ldots $ until variation with $d_E$ cases. Mentioned property can be the correlation integral [20].

The average number of points on the attractor within a radius of $r$ of points $x$ in the state space, $n(r, x)$, is defined by

$$n(r, x) = \frac{1}{N} \sum_{i=1}^{N} \theta(r - |x(i) - x|), \quad \text{(7)}$$

where $\theta(u)$ is the Heaviside function: $\theta(u) = 0$, $u < 0$; $\theta(u) = 1$, $u > 0$. The average over all points in attractor of $n(r, x)$ is called correlation integral and is described below:

$$C(r) = \frac{1}{M} \sum_{j=1}^{M} \left[ n(r, x(j)) \right] = \frac{1}{M} \frac{1}{N} \sum_{j=1}^{M} \sum_{i=1}^{N} \theta(r - |x(i) - x(j)|). \quad \text{(8)}$$

We can evaluate $C(r)$ as a function of $d_E$ and determine when the slope of its logarithm as function of $\log r$ becomes independent of $d_E$.

In this method, the authors perceive the possibility to improve its efficiency through the use of correlation dimension.

### 2.4 Correlation dimension saturation

Fractal geometry and fractal dimension conception provide a general framework for the study of such irregular sets, like strange attractors. Fractals, which are irregular geometric objects, require a special meaning of dimension. Very roughly, fractal dimension provide a description of how much space a set fills [7, 16]. Of the wide variety of fractal dimensions in use, the definition of Hausdorff is the most important:

$$\dim_H F = \inf \left\{ s: H^s(F) = 0 \right\} = \sup \left\{ s: H^s(F) = \infty \right\}. \quad \text{(9)}$$

Here $H^s(F)$ is $s$-dimensional Hausdorff measure of set $F$ defined as follows:

$$H^s(F) = \lim_{\sigma \to 0} H^s_\sigma(F), \quad \text{(10)}$$

where $H^s_\sigma(F)$ is described by equation

$$H^s_\sigma(F) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s: \text{U}_i \text{ is } \sigma\text{-cover of } F \right\}. \quad \text{(11)}$$

The attractor of chaotic system has a fixed fractal dimension determined by Hausdorff. Because of difficulties in its computations, several other fractal dimension estimators were introduced.

http://www.mii.lt/NA
Correlation dimension $D_2$ is one of the fractal dimension estimators particularly suited for relatively easy experimental determination. It is connected with correlation integral

$$C(r) \approx r^{D_2}. \quad (12)$$

Above equation after some transformations can be written as

$$D_2 = \lim_{r \to 0} \frac{C(r)}{\ln r}, \quad (13)$$

$$D_2 = \lim_{r \to 0} \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\theta(r - |x(i) - x(j)|)}{\ln r}. \quad (14)$$

For large $N$, Eq. (14) is useful estimator of $D_2$.

$$\ln(C(r)) \approx \ln(r^{D_2}) \approx D_2 \ln r. \quad (15)$$

In conclusion, estimation of $D_2$ value is a slope factor of regression function chart tangent to the most linear part of $\ln(C(r)) = f(\ln r)$ dependency.

The authors propose to saturate correlation dimension $D_2$ instead to correlation integral in saturation of system invariants method. As will be shown in Section 4.2, this improvement allows for a more clear and intuitive embedding dimension selection. Correlation dimension was chosen because of its intuitive implementation. There are several other fractal dimension estimators, such as box-counting dimension, information dimension or Higuchi dimension. Interesting approaches to fractal dimension estimation and its applications was presented [33, 35].

### 2.5 The largest Lyapunov exponent

Lyapunov exponents describe velocity of distance increasing between two initially neighboring orbits in the state space. The largest Lyapunov exponent describes the mean divergence between initially neighboring trajectories by the following formula:

$$d(t) = D_0^{L_1 t}, \quad (16)$$

where $d(t)$ is distance between orbits in time, $D$ is initial separation between neighboring points and $L_1$ is the largest Lyapunov exponent. In practice, value of the largest Lyapunov exponent is commonly used chaos detection factor, because its positive value clearly indicates chaotic behavior in the system. Wolf in his article [40] proposed to estimate the largest Lyapunov’s exponent for time series based on below equation:

$$L_1 = \frac{1}{t} \sum_{j=1}^{m} \log_2 \frac{L'(t_{j+1})}{L(t_j)}, \quad (17)$$

where $L(t_j)$ is a distance (in Euclidean sence) between pairs of trajectory points in time $t_j$, and $L'(t_{j+1})$ is a distance between pairs of trajectory points in time $t_{j+1}$.

Alternative methods of the largest Lyapunov exponent estimation are Rossenstein’s algorithm [13] or Kantz’s algorithm [24].
2.6 Hurst exponent

Hurst exponent $H$ can be counted by using $R/S$ analysis based on the following algorithm [23]:

1. $N$-length time series divide into $k$ $n$-length subsets, where $k \cdot n = N$.
2. For each subset $m = 1, \ldots, k$:
   - count average $E_m$ and standard deviation $S_m$,
   - rescale each value in subset $x_{i,m}$ by cutting of the average value counted for this subset
     \[ z_{i,m} = x_{i,m} - E_m \quad \text{for } i = 1, \ldots, n, \]     (18)
   - construct cumulated subset of rescaled values
     \[ y_{i,m} = \sum_{j=1}^{i} z_{j,m} \quad \text{for } i = 1, \ldots, n, \]     (19)
   - calculate range $R_m$
     \[ R_m = \max\{y_{1,m}, \ldots, y_{n,m}\} - \min\{y_{1,m}, \ldots, y_{n,m}\}, \]     (20)
   - rescale range
     \[ R_m \frac{S_m}{S_m}. \]     (21)
3. The average value of rescaled range for $n$-length subset can be described by
   \[ \left( \frac{R}{S} \right)_n = \frac{1}{k} \sum_{m=1}^{k} \frac{R_m}{S_m}. \]     (22)

Hurst exponent can be calculated from below equation:
   \[ \left( \frac{R}{S} \right)_n = (cn)^H, \]     (23)
where: $R/S$ – rescaled range, $n$ – number of measurements in subset, $c$ – constant value.

Finally, after logarithmization of both sides above equation can be rewritten as follows:
   \[ \ln \left( \frac{R}{S} \right) = H \ln n + H \ln c. \]     (24)

Slope factor value of linear regression function tangent to the most linear part of $\ln(R/S) = f(\ln n)$ dependency is an estimation of $H$.

It can be observed that, Hurst exponent values can be divided into three sets:

- $H = 0.5$ – samples in examined time series is random and not correlated (i.i.d. series),
- $0 < H < 0.5$ – examined time series is antipersistent and ergodic, which means that the distribution parameters are constant (if the system trajectory in some period of time increases, then it is highly probable that it will decrease in upcoming period,
• $0.5 < H < 1$ – examined time series is persistent, which means that samples create some trends (if the system trajectory in some period of time follows a particular direction, it is highly probable that in the upcoming period the trajectory will keep the same direction).

3 Test time series

In this article, times series generated by four systems described below are examined.

3.1 Lorenz system

Considered Lorenz system is defined by differential equations

\[
\begin{align*}
\frac{dx(t)}{dt} &= \sigma(y(t) - x(t)), \\
\frac{dy(t)}{dt} &= -xz(t) + rx(t) - y(t), \\
\frac{dz(t)}{dt} &= xy(t) - bz(t).
\end{align*}
\]

For $\sigma = 10$, $r = 28$ and $b = 8/3$, this system generates chaotic time series. Figure 1 presents Lorenz attractor drawn for $2^{14}$ points. Figure 2 presents time series generated by state variable $x$.

3.2 Henon system

Henon system is described by differential equations

\[
\begin{align*}
\frac{dx(t)}{dt} &= 1.4 + 0.3y(t) - x(t)^2, \\
\frac{dy(t)}{dt} &= x(t).
\end{align*}
\]

Figure 3 presents Henon attractor drawn for $2^{14}$ points. Figure 4 presents time series generated by state variable $x$.
3.3 Random and periodic systems

Random time series are generated by uniform distributed system. Periodic time series are generated by the system described by the following equations:

\[ x(n) = \sin(n) + \cos(n). \]  

(27)

4 Numerical results

4.1 Time delay estimation

Based on \( I = f(T) \) dependency charts presented below, time delay values \( T \) for test systems were estimated.

Optimal time delay value \( T_{opt} \) is estimated either as a first minimum of dependency \( I = f(T) \) or from the equation \( I(T_{opt})/I(0) \approx 4/5 \). Analyzing Fig. 5, it can be assumed that \( T_{opt} \approx 13 \), because the first minimum is located in this area. In order to estimate
Compound method of time series classification

Fig. 7. $I = f(T)$ dependency chart for random system.

Fig. 8. $I = f(T)$ dependency chart for periodic system.

$T_{\text{opt}}$ for the Henon system, the second criteria was used. It means that the optimal time delay value was taken as 80% of initial value of mutual information $I(0)$. In this case, $I(0) \approx 7.62$, $I(T_{\text{opt}}) = 6.1$ and in consequence $T_{\text{opt}} \approx 1$. Analyzing dependency $I = f(T)$ for random system, it can be assumed that all samples for $T > 0$ are not correlated. For periodic system, a mutual information value $I$ is high in wide range of $T$. It confirms a deterministic nature of this system.

4.2 Embedding dimension estimation

In order to estimate embedding dimension of examined systems, saturation of system invariants have been used.

Above charts express dependency between correlation integral and radius used to calculate this invariant. This method was described in Section 2.3 and it assumes control of correlation integral $C_2$ value for increasing value of radius $e$.

Analyzing Fig. 9, it can be stated that the biggest change can be observed for dimensions lower than $d_E \leq 3$. Further increase of embedding dimension value doesn’t have significant impact on examined invariant $C_2$. In conclusion estimated embedding dimension value was taken as $d_E = 3$.

In Fig. 10 it can be noticed that the correlation integral $C_2$ is saturated for $d_E \geq 2$. For random time series, correlation integral is never saturated which can be observed in Fig. 11. Analyzing Fig. 12, it can be stated that saturation is present for wide range of $d_E \geq 1$.

In the second approach, system’s dynamics can be described by correlation dimension which is one of the fractal dimension and can be also treated as an invariant in saturation method.

For instance, analyzing Fig. 13 for Lorenz system, it can be observed that if embedding dimension is greater than three ($d_E \geq 3$), correlation dimension value $C_2$ seems to be saturated and doesn’t change to much.

For Henon time series saturation is reached for $d_E \geq 2$, which can be easily seen in Fig. 14.
As it was mentioned before, for random time series saturation is never observed, which can be also assumed from Fig. 15. It means that the distance between pairs of points in state space never became a constant value. Figure 16 presents that saturation is reached for $d_E = 1$. 

http://www.mii.lt/NA
4.3 Hurst exponent analysis

Below figures allow to estimate Hurst exponent using linear regression method.

Fig. 15. $D_2$ values versus $d_E$ for random time series. Fig. 16. $D_2$ values versus $d_E$ for periodic time series.

Fig. 17. $\ln(R/S)$ versus $\ln(n)$ for Lorenz time series.

Fig. 18. $\ln(R/S)$ versus $\ln(n)$ for Henon time series.

Fig. 19. $\ln(R/S)$ versus $\ln(n)$ for random time series.

Fig. 20. $\ln(R/S)$ versus $\ln(n)$ for periodic time series.
Analyzing Fig. 17, it can be stated that examined time series is persistent and $H \approx 0.8$. Based on results of Hurst analysis presented in Fig. 18, it can be stated that Henon time series is antipersistent and $H \approx 0.42$. Results of $R/S$ analysis for random time series can be found in Fig. 19. Analyzing this figure, it can be concluded that $H \approx 0.5$, which confirms randomness of this time series. Figure 20 confirms a deterministic character of examined time series ($H \approx 1$).

4.4 Lyapunov exponents

Below figures present the largest Lyapunov exponent estimations for the test time series computed by using Wolf’s algorithm [40]. Based on below figures it can be stated that for chaotic times series generated by Lorenz and Henon system, the largest Lyapunov exponent are positive. It means that these systems are highly sensitive to initial conditions. Small change of initial state, generates exponential increasing distance between trajectories.

Unfortunately, Lyapunov exponents can not be used to differ chaotic and random time series, because both of them give positive values. The largest Lyapunov values for random times series are presented on Fig. 23.

![Fig. 21. The largest Lyapunov exponent for Lorenz time series.](http://www.mii.lt/NA)

![Fig. 22. The largest Lyapunov exponent for Henon time series.](http://www.mii.lt/NA)

![Fig. 23. The largest Lyapunov exponent for Random time series.](http://www.mii.lt/NA)

![Fig. 24. The largest Lyapunov exponent for Periodic time series.](http://www.mii.lt/NA)
As it can be seen on Fig. 24, the largest Lyapunov exponent for periodic system is negative, which confirms the existence of non-chaotic attractor.

5 Conclusions

In this article, three types of times series were examined: chaotic, random and periodic. To differ type of behavior, the following methods were used: mutual information to estimate time delay $T_{opt}$, saturation of system’s invariants for estimation of embedding dimension $d_E$, correlation dimension to estimate fractal dimension $D_2$, Hurst exponent to examine memory effect $H$ and the largest Lyapunov exponent to check dependency on initial conditions.

Values of all of mentioned above parameters for all examined systems are gathered in Table 1. Analyzing $T_{opt}$ for random system, it can be stated that information in the system is lost very quickly and it is exactly opposite to periodic system, where this dependency is continuously observable. Analyzing fractal dimension, it can be noticed that for chaotic time series, saturated values are non-integer, while for periodic are very close to integer and infinity for random. Hurst analysis allows to differ mainly random and periodic behaviors. Lyapunov exponents are positive for chaotic time series (Lorenz and Henon) and random. For periodic time series the largest Lyapunov exponent is negative.

Based on above summary, we can conclude that below set of indicators are properly selected to differ behaviors in time series. If there is a need to distinguish chaotic and periodic behavior in time series, the largest Lyapunov estimation is suggested method. On the other hand, if the goal is to differ random and periodic times series, Hurst exponent estimation could be useful. Fractal dimension estimations can provide information about attractor, which in consequence allows simple method to differ chaotic from periodic systems. To sum it up, it can be stated that there is none common method which gives a possibility to fully differ three types of behavior: chaotic, random and periodic. It is necessary to use wider set of methods, which provide more complex information about different properties of examined times series.

Additional conclusion which can be drawn from this paper, is that correlation dimension seems to be very efficient invariant in the saturation of system’s invariant method. Saturation of correlation dimension instead of correlation integral is more precise and intuitive, which can be observed on Figs. 13–16. This approach allows to define embedding dimension more clearly.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Lorenz</th>
<th>Henon</th>
<th>Random</th>
<th>Periodic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{opt}$</td>
<td>13</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$d_E$</td>
<td>3</td>
<td>2</td>
<td>$\infty$</td>
<td>1</td>
</tr>
<tr>
<td>$D_2$</td>
<td>1.98</td>
<td>1.19</td>
<td>$\infty$</td>
<td>1</td>
</tr>
<tr>
<td>$H$</td>
<td>0.80</td>
<td>0.42</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>$L$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>
References


http://www.mii.lt/NA


