ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Calculation of EEG correlation dimension: Large massifs of experimental data

Alexey Mekler *

Institute of Human Brain Russian Academy of Sciences, Acad. Pavlov str. 9, 197376 Sankt-Petersburg, Russia

ARTICLE INFO

Article history:
Received 6 May 2008
Received in revised form
14 June 2008
Accepted 19 June 2008

Keywords:
EEG
Human brain
Deterministic chaos
Dimensional complexity
Correlation dimension

ABSTRACT

Correlation dimension of reconstructed attractor (D_2) is one of the specific values for human electroencephalogram (EEG). It makes it possible to evaluate variability of human brain functioning.

There are some requirements, made for time series to be analyzed by Grassberger–Procaccia method, and EEG does not meet them perfectly. Also, realization of this algorithm uses subjective evaluation procedure and needs setting of additional parameters for calculation, such as lag and embedding dimension.

Method of calculation, suggested in the article, meets specifics of EEG and allows calculating without using subjectivity. This makes it possible to process large amounts of experimental data. The value, calculated by this method, is not D_2 in strict sense, but it is useful for evaluated variability of brain functioning.

1. Introduction

Electroencephalogram (EEG) analysis methods, implementing theory of dynamical systems and, particularly, deterministic chaos have been widely adopted in the last two decades.

One of the values that could be calculated using these methods is correlation dimension of reconstructed attractor of EEG (D_2). This value allows judging generalization of brain processes or their variability.

Calculation of this value involves certain difficulties. First, corresponding mathematical theorems and algorithms developed from these theorems are theoretically valid for endless and noiseless chaotic processes. The EEG does not meet these conditions. Second, different researchers may use slightly different calculation procedures. Also, there is no consensus on selection of calculation parameters. This ends in difficulties in comparison of results achieved by different laboratories. Finally, third, two stages of calculations (finding proper scaling region and region of saturation on intermediate graphs) are held in subjective manner. As a result, processing of big volumes of experimental data becomes very time-expensive. This problem forced researchers to modify classic algorithm to make it possible to carry on calculations in automatic mode.

Algorithm, proposed in this article, also allows automatic calculation of D_2 and processing of big massifs of experimental data. This algorithm is adopted for EEG signals specifics. It implements calculation procedures, which are represented in quite popular TISEAN project [1], making it available to a wide circle of researchers. Another advantage of TISEAN algorithms is availability to keep calculations in fast mode (see description in Ref. [1]). This is quite important when researcher deals with large volumes of data. Resulting value, calculated by this method, is not D_2 in strict sense (as well as results of other automated algorithms), but it is useful for evaluation of brain functioning variability.

* Tel.: +7 812 444 04 19.
E-mail address: mekler@narod.ru.
0169-2607/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2008.06.009

2. General terms

It is known, that calculation of D_2 is based on Takens' theorem [2]. That theorem permits one to judge about the evolution of the whole system by a single time series derived therefrom. This can be done by constructing a pseudo attractor with metrical characteristics of the system's attractor in the phase space. The correlation dimension is then calculated using the pseudo attractor without having to reconstruct the dynamic system that had generated the time series in point. Calculation procedure is as follows.

Let $X_0(t)$ be a sequence of instantaneous readouts of x. We can map each point of this sequence $x(t_i)$ into point of n-dimensional space with the following coordinates—$\{x(t_i), x(t_i+\tau),\ldots,x(t_i+(n-1)\tau)\}$. According to Takens' theorem there are such lag τ and a depth of embedding n resulting in set of points of the same metrical properties as attractor of the system under study. τ can be chosen arbitrary for infinite series. This n-dimensional space is called an embedding or a lag space, n — embedding dimension and set of points—reconstructed attractor.

According to Grassberger and Procaccia, calculation of D_2 is based on the following principle [3].

The proportion of points in reconstructed attractor that were at a distance not greater than ε from x_i belonged to the same set is calculated. Then, averaging this value for all points of reconstructed attractor, the following relation (so-called correlation integral) is obtained:

$$C(\varepsilon)=\frac{1}{N^2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\Theta(\varepsilon-|\vec{x}_i-\vec{x}_j|)$$

where $\Theta(x)$ is Heaviside's function: $\Theta(x)=\begin{cases}0, & x<0\\ 1, & x>0\end{cases}$ and $|\vec{x}_i-\vec{x}_j|$ —the distance between images of i-th and j-th points of original time series in reconstructed attractor.

Correlation integral behaves as a power of ε for small values of ε:

$$C(\varepsilon)\propto\varepsilon^{D_2} \quad \text{and} \quad D_2\approx\frac{\log C(\varepsilon)}{\log\varepsilon}$$

where D_2 is correlation dimension of reconstructed attractor.

Thus, dimension of reconstructed attractor can be estimated as a slope of linear region of $C(\varepsilon)$ function, plotted in the double logarithmic scale (Fig. 1). This region corresponds to plateau at graph of local slopes of $C(\varepsilon)$—so-called Rapp's graph (Fig. 4a). D_2 may be estimated as an average value of Rapp's graph on the plateau—so-called scaling region. For correct estimation of D_2 it is necessary to set correct τ and n (or, as it is usually designated, D_{emb}).

D_{emb} could be set as value where curve D_2 (D_{emb}) saturates, successively calculating D_2 for increasing values of D_{emb}. If the process under study is stochastic, then there would be no saturation (Fig. 2).

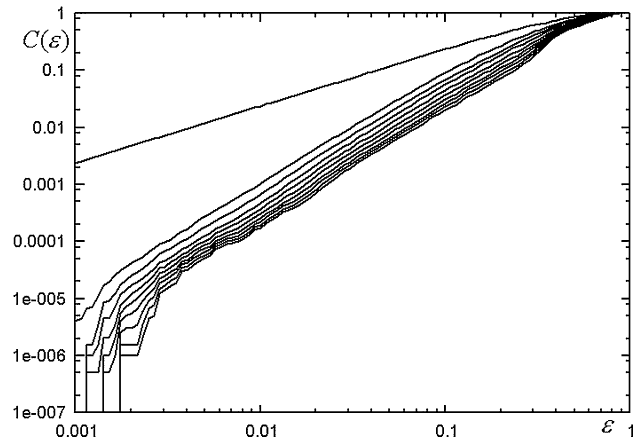

Fig. 1 – Correlation integral of Lorentz' system attractor, calculated for D_{emb} from 1 to 10.

3. Automatic calculation of the EEG's reconstructed attractor correlation dimension

Modifying of classical D_2 calculation algorithm in order to automate it always leads to more or less spurious results. Another source of inaccuracy is the specificity of time series under study (EEG signal) that was mentioned in Section 1). This is the price, we pay for such a signal specifics and the way, we calculate D_2. Our task is to find the some balance between three reciprocal aims:

- Accuracy.
- Automation.
- Tolerance to EEG specifics.

D_2 calculation procedure could be divided into three stages:

- Calculation of $C(\varepsilon)$ for different values of D_{emb}.
- Finding of scaling region and calculation of average slope of $C(\varepsilon)$ in it—for every value of D_{emb}.
- Estimation of optimal D_{emb} and corresponding value of D_2.

In order to reach declared aims we meet a set of problems at each stage. As a matter of presentation convenience there is a reason to analyze the first stage at least.

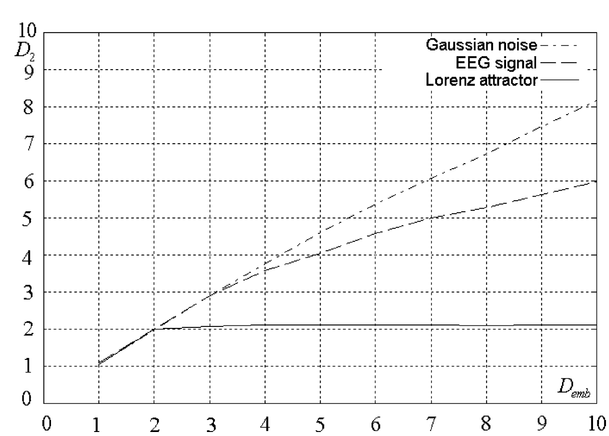

Fig. 2 – Correlation dimensions, calculated at different values of D_{emb}.