Journal of Neuroscience Methods 172 (2008) 27–33

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Properties of multivariate data investigated by fractal dimensionality

Danko Nikolić[a,b,*], Vasile V. Moca[a,b,c], Wolf Singer[a,b], Raul C. Mureşan[a,c]

[a] *Department of Neurophysiology Max-Planck-Institute for Brain Research, Deutschordenstr. 46, 60528 Frankfurt am Main, Germany*
[b] *Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany*
[c] *Center for Cognitive and Neural Studies (Coneural), Str. Cireşilor 29, 400487 Cluj-Napoca, Romania*

ARTICLE INFO

Article history:
Received 11 December 2007
Received in revised form 31 March 2008
Accepted 7 April 2008

Keywords:
Fractal
Data model
Multivariate analysis
Dimensionality

ABSTRACT

Elaborated data-mining techniques are widely available today. Nevertheless, many non-linear relations among variables remain undiscovered in multi-dimensional datasets. To address this issue we propose a method based on the concept of fractal dimension that explores the structure of multivariate data and apply the method to simulated data, as well as to local field potentials recorded from cat visual cortex. We find that with changes in the analysis scale, the dimensionality of the data often changes, indicating first that the data are not simple fractals with one unique dimension and second, that, at a certain scale, important changes in the geometric structure of the data may occur. The method can be used as a data-mining tool but also as a method for testing a model's fit to the data. We achieve the latter by comparing the dimensionality of the original data to the dimensionality of the data reconstructed from a model's description of the data (here using the general linear model). The method provides indispensable help in estimating the complexity of non-linear relationships within multivariate datasets.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Data analysis requires investigation of relations between data points, and by any type of such analysis, irrespectively of whether explorative or hypothesis-driven, only a limited subset of all possible relations can be addressed. Despite the elaborated data-mining procedures, many such relations in many datasets remain hidden. We can only guess how many (important) scientific insights have been missed just because patterns could not be easily detected in otherwise, perfectly reliable and legitimate sets of data. Therefore, we should welcome every new analysis method that is able to probe new relationships and present the results in an elegant and easily interpretable way. For such methods, reduction of dimensionality plays an important role (Brand, 2003; Levina and Bickel, 2004; Tenenbaum et al., 2000). In the present study we propose a method that is designed to explore relations across multivariate data points and that is based on the concept of fractal dimension.

1.1. *The concept of fractal dimension*

A fractal is an object with a high degree of self-similarity, whereby globally the object looks very similar to its details (Falconer, 2003). We show three example fractals in Fig. 1a–c and for one of them we illustrate how it is created by a simple iterative procedure (Fig. 1a) (Falconer, 2003, pp. xviii–xx). Perhaps the most common quantitative description of a fractal is the measure of its dimension. As the dimensionality of standard geometric objects, e.g., triangles and cubes, can be grasped easily by intuition, it is also easy to acquire intuitive understanding of fractal dimensions. Depending on the space that they occupy, fractals have different dimensionality and they can be given by real numbers. For example, the Koch fractal in Fig. 1a occupies a $D = 1.26$-dimensional space. The Sierpinsky fractal in Fig. 1b and another fractal in Fig. 1c (Landau and Paez, 1997) appear visually to occupy gradually more space, and this is consistent with their calculated fractal dimensions (indicated in Fig. 1).

In the present study we use the concept of fractal dimension to address the common scientific issue of the dimensionality of data—even if the data are, strictly speaking, not fractals. Data dimensionality is usually investigated by principal component analysis (PCA) or factor analysis (Gorsuch, 1983), but not with fractal dimension. The latter is normally used only if the analyzed objects are already known to have (or are expected to have) fractal properties (e.g., a chaotic attractor) (Strogatz, 1994). However, this need not be the case. Much insight about datasets commonly used in scientific research (e.g., those that are described typically by the general linear model—GLM) can be gained by investigating the dimensionality of non-fractal data with fractal dimension (Lutzenberger et al., 1992; Pereda et al., 1998; Woyshville and Calabrese, 1994).

* Corresponding author at: Department of Neurophysiology Max-Planck-Institute for Brain Research, Deutschordenstr. 46, 60528 Frankfurt am Main, Germany. Tel.: +49 69 96769 736; fax: +49 69 96769 327.
E-mail address: danko@mpih-frankfurt.mpg.de (D. Nikolić).

0165-0270/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2008.04.007

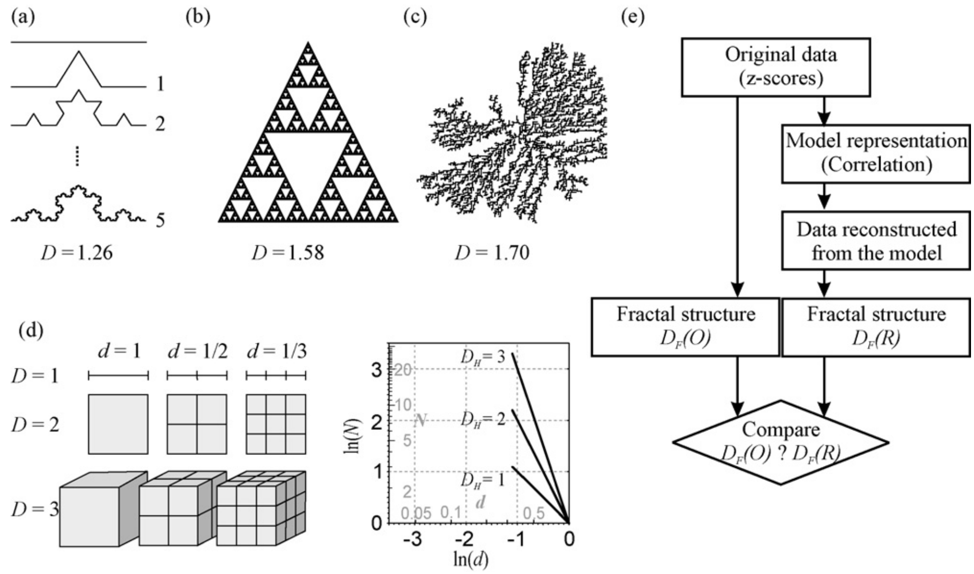

Fig. 1. The concept of fractal dimension, its measurement, and the proposed analysis steps necessary to investigate how well a model of a dataset accounts for the data's fractal structure. (a) Example Koch fractal and the process of its creation through iterative steps. (b) Sierpinsky (triangle) fractal with dimension larger than that in (a). (c) A fractal obtained through diffusion-limited aggregation of particles undergoing a random walk (Landau and Paez, 1997). The corresponding dimensions of fractals, D, are indicated. (d) The procedure for computing the fractal dimension is illustrated for three objects with true D-values = 1–3. The space occupied by the object is partitioned into 'boxes' of size d. The number of resulting boxes that intersect (or cover) the object, N, is counted. Finally, N is plotted against the size d in a log–log plot (right panel). If the object is a fractal, the plot results in a straight line and the slope of the line indicates the dimension of the object. (e) The dimensionality analysis of the data consists of two tracks. In one, the fractal structure of the original data is calculated and in the other, the fractal structure of data reconstructed from the model is calculated (in all examples we use GLM). Finally, the two structures are compared (e.g., by comparing their log–log plots).

1.2. Measuring fractal dimension

Fractal dimension is formally computed by one variant of the Hausdorf dimension, D_H (Falconer, 2003). The principles of this calculation are shown in Fig. 1d by a variant of Hausdorf dimension known as box-counting method. Here, with a change in the analysis scale, d, a different number of boxes, N, is needed to cover the object. For example, for one-, two-, and three-dimensional objects in Fig. 1d (left panel) the counts are 2, 4, and 8 for $d = 1/2$ and 3, 9, and 27 for $d = 1/3$, respectively. The dimensions of the objects are then calculated by plotting ln(d) versus ln(N) and calculating the absolute values of the slopes of the fitted straight lines (right panel). Therefore, the dimension D_H is the absolute value of the exponent in $N \approx d^{-D}$, describing how quickly the count N grows with the decrease in d.

In the present study we estimate fractal dimensions by a numerical procedure that is more computation-effective than box-counting methods and that is known as the correlation dimension, D_F (Camastra and Vinciarelli, 2002; Grassberger and Procaccia, 1983). In most cases, D_F produces identical results as D_H (within numerical limits) while in other cases $D_F < D_H$, the differences being very small. Thus, D_F can be considered a lower estimate of D. Numerical details for the calculation of D_F are provided in Section 2.

Central to our analyses are the log–log plots such as the one shown in Fig. 1d. For successful application of the method, it is not necessary that the data exhibit the actual properties of fractals. Fractals produce a straight line in the log–log plot (self-similarity) while plots for the data might have curvatures (changes in the slope). Curvatures provide important information about alterations in data dimensionality across different scales, indicating that the data are not simple fractals but could be instead described as multifractals, which can in turn lead to the discovery of interesting data properties (e.g., Feder, 1988, pp. 185–186). One important application is the comparison between the log–log plots for the original data and those for samples recreated by a model of the data. This allows one to test, in a novel way, how well the model accounts for the original data (for the present analyses only GLM models are tested, Fig. 1e). An example application to real data is made for local field potentials (LFP), simultaneously recorded with 16 electrodes from cat visual cortex.

2. Materials and methods

2.1. Experimental procedures

Intracranial LFP recordings were performed on an adult cat under anesthesia induced with ketamine and maintained with halothane and a mixture of N_2O (70%) and O_2 (30%). The cats were paralyzed with intravenously applied pancuronium bromide (Pancuronium, Organon, 0.15 mg kg^{-1} h^{-1}). LFP activity was recorded from area 17 with 16-channel silicon probes (organized in a 4 × 4 spatial matrix) which were supplied by the Center for Neural Communication Technology at the University of Michigan (Michigan probes). The inter-contact distances were 200 μm (0.3–0.5 MΩ impedance at 1000 Hz). Signals were amplified 1000× and filtered 1–100 Hz to extract local field potentials (LFP) (1 kHz sampling rate). To evoke visual responses drifting sinusoidal gratings were presented on a 21 in. computer screen (100 Hz refresh rate) using ActiveSTIM software for visual stimulation (ActiveSTIM, high precision stimulation tool, http://www.ActiveSTIM.com). One stimulus condition is presented in total 20 times. More details on methods for data acquisition can be found in (Biederlack et al., 2006).

2.2. Artificially generated data

The artificial datasets shown in Fig. 2a–c (2000 points each) were generated by a help of a Mersenne Twister pseudo random-number generator (Matsumoto and Nishimura, 1998). In Fig. 2a

Download English Version:

https://daneshyari.com/en/article/4336030

Download Persian Version:

https://daneshyari.com/article/4336030

Daneshyari.com