Computer Methods and Programs in Biomedicine 152 (2017) 131–139

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Binary classification of multichannel-EEG records based on the ϵ-complexity of continuous vector functions

Alexandra Piryatinska[a,*], Boris Darkhovsky[b], Alexander Kaplan[c]

[a] Department of Mathematics, San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94070, United States
[b] Institute for Systems Analysis, FRC CSC RAS, Higher School of Economics, Moscow, Russia
[c] Department of Human Physiology, M.V. Lomonosov Moscow State University, Pirogov Russian National Research Medical University, Moscow, Russia

ARTICLE INFO

Article history:
Received 17 October 2016
Revised 2 August 2017
Accepted 4 September 2017

Keywords:
EEG analysis
Binary classification
ϵ-complexity
Schizophrenia

ABSTRACT

Background and objective: A crucial step in a classification of electroencephalogram (EEG) records is the feature selection. The feature selection problem is difficult because of the complex structure of EEG signals. To classify the EEG signals with good accuracy, most of the recently published studies have used high-dimensional feature spaces. Our objective is to create a low-dimensional feature space that enables binary classification of EEG records.
Methods: The proposed approach is based on our theory of the ϵ-complexity of continuous functions, which is extended here (see Appendix) to the case of vector functions. This extension permits us to handle multichannel-EEG records. The method consists of two steps. Firstly, we estimate the ϵ-complexity coefficients of the original signal and its finite differences. Secondly, we utilize the random forest (RF) or support vector machine (SVM) classifier.
Results: We demonstrated the performance of our method on simulated data. We also applied it to the problem of classification of multichannel-EEG records related to a group of healthy adolescents (39 subjects) and a group of adolescents with schizophrenia (45 subjects). We found that the random forest classifier provides a superior result. In particular, out-of-bag accuracy in the case of RF was 85.3%. Using 10-fold cross-validation (CV), RF gave an average accuracy of 84.5% on a test set, whereas SVM gave an accuracy of 81.07%. We note that the highest accuracy on CV was 89.3%. To compare our method with the classical approach, we performed classification using the spectral features. In this case, the best performance was achieved using seven-dimensional feature space, with an average accuracy of 83.6%.
Conclusions: We developed a model-free method for binary classification of EEG records. The feature space was reduced to four dimensions. The results obtained indicate the effectiveness of the proposed method.

1. Introduction

An electroencephalogram (EEG) provides a direct measure of the electrical activities of the brain along the scalp. It is a rich source of information about the brain for healthy individuals and patients with neurological diseases. It permits quantitative evaluation of cognitive functioning and mental states [1]. EEG records are used in the brain–computer interface [2–4] for decoding intentions and their translation into commands and in diagnosing mental illnesses such as schizophrenia [1,5,6]. For a review of classification algorithms for the EEG-based brain–computer interface, see [7].

To obtain useful information from EEG data, feature extraction is necessary. In the literature, a whole set of quantitative estimates of the spectral and temporal features of the EEG signal have been utilized (see, e.g., [8]). In particular, there is an extensive literature on attempts to use these characteristics for schizophrenia diagnosis (for a review and meta-analysis of such papers, see, e.g., [9]). However, such an approach requires assumptions about the data generating mechanism, and there are no generally accepted models on the data generation mechanism for EEG data.

Another approach for feature selection is related to the fundamental concepts of the modern theory of nonlinear dynamical systems, such as entropy, correlation dimension D_2, and the Lyapunov exponent (see, e.g., [10–13]). These features can genuinely reflect the complexity of a mechanism for generating a signal, but only under the assumptions of stationarity and ergodicity of the signal. Due to the nature of EEG signals, these assumptions are not fully justifiable (see, e.g., [14,15]). We also note that the entropy (and other "nonlinear") measures of complexity are characteristics of the whole ensemble of trajectories and not individual sample paths of a stationary and ergodic stochastic process. Therefore, such mea-

* Corresponding author.
E-mail address: alpiryat@sfsu.edu (A. Piryatinska).

http://dx.doi.org/10.1016/j.cmpb.2017.09.001
0169-2607/

sures cannot measure the "complexity" of the individual trajectory. Classification methods based on fractal dimension (Hausdorff dimension, box-counting, and the Higuchi dimension) employ the idea of counting the numbers of simple sets covering graphs of functions (in the case of the Higuchi dimension, the trajectory's length needs to be estimated as well). Those approaches require a considerable quantity of data and thus are much less economical than an approximation of the function by a set of known functions as is done in our approach via the concept of the ϵ-complexity. In addition, those methods are usually used for scalar functions (the Higuchi method cannot be extended to scalar functions of the vector argument), whereas our approach can be applied to vector functions of the vector argument.

The majority of papers that report a high accuracy for the classification of EEG signals use high-dimensional feature spaces and face the problem of overfitting (see, e.g., [16]). It is recommended (see, e.g., [17]) to use at least five to ten times as many training samples per class as the dimensionality of the feature space.

In [18–20], low-dimensional feature spaces were used for classification of schizophrenic vs. control subjects based on EEG records, but they considered various types of experiments such as reactions to the stimuli; in our case, all the data were collected in the resting state.

In this paper, we propose a model-free approach for the feature selection problem, one that assures good accuracy in low-dimensional feature space. In particular, we propose to apply the notion of the ϵ-complexity of continuous functions to the classification problem of multichannel-EEG records.

In our approach, the complexity of an individual continuous function given by a discrete set of values is measured by the number of function values on a uniform grid that is necessary to reconstruct the function with a given error by a given set of methods. The notion of ϵ-complexity is in line with Kolmogorov's general idea of the "complexity" of an object. Such an approach was first tested for evaluation of functional states of the brain using EEG recordings in [21,22].

In 2012–2014 (see [23–26]), the theory of the ϵ-complexity of continuous functions defined on a compact set in a finite-dimensional space was developed. This theory enabled us to develop a *novel approach* to the problems of segmentation and classification of time series of an arbitrary nature. In this paper, we extend the theory of the ϵ-complexity of continuous functions to the case of continuous vector functions (see Appendix). This extension enables us to apply such an approach to the binary classification problem of EEG records.

The paper is organized as follows. In Section 2, we describe our method. In particular, in Section 2.1 we give a description of the notion of the ϵ-complexity of a vector function on a semantic level and provide a characterization of the ϵ-complexity for vector functions given by a finite set of values. In Section 2.2, we provide an algorithm for estimation of the ϵ-complexity coefficients for multichannel EEG records. In Section 2.3, we describe our classification procedure. In Section 3, we provide results of the simulations and apply our method to the classification of the EEG records of adolescents with schizophrenia and of healthy subjects. We also performed classification of the EEG data using the spectral features. In Section 4, we provide conclusions and discuss our results. The Appendix provides the precise definition of ϵ-complexity and the theorem characterizing the complexity of Hölder vector functions.

2. Method

In this section, we give a description of the proposed method for classification of multichannel-EEG records. We will treat a multichannel-EEG record as a d-dimensional vector function $x(t) = (x_1(t), \ldots, x_d(t))$, where d is the number of channels, which is given on some fixed time interval $t \in [0, T]$.

Since modern recording equipment is digital, instead of a continuous vector function $x(t)$ the researcher receives discrete samples $(x(0), x(T/n), x(2T/n), \ldots, x(T))$, i.e., sequences of n d-dimensional vectors. Here, $n = fT$, where f is a sampling frequency (if the frequency is measured in hertz, i.e., times per second). For example, if $T = 60$ seconds and $f = 128$ Hz, the procedure produces 7680 d-dimensional vectors. Without loss of generality, we can assume that $\max_{0 \le k \le n} |x_i(kT/n)| = R_i > 0,\ i = 1, 2, \ldots, d$.

2.1. Description of the ϵ-complexity of a continuous vector function

Let us describe our notion of ϵ-complexity on the semantic level. The precise definition and formulation of the relevant theorem are given in the Appendix.

Firstly, we choose a number $0 < S < 1$. From each component of a vector function $\{x_i(kT/n)\}_{k=0}^{k=n}$, $i = 1, \ldots, d$, we discard $[(1-S)n]$ values(henceforth, the symbol $[a]$ denotes the integer part of a number a) in such a way that the remaining values are approximately uniformly distributed. For example, if $S = 0.5$, then even or odd values in each component of the function are discarded (for details, see the Appendix).

Assume we have some fixed collection $\mathcal{F}$ of approximation methods that can be used for the reconstruction of a continuous function by its values on some uniform grid. Employing the collection of methods $\mathcal{F}$, we reconstruct the values of the ith component ($i = 1, \ldots, d$) of the vector function in discarded points using the retained values of the function component. For each component, we find the method that reconstructs it with the minimum relative (in relation to $R_i, i = 1, \ldots, d$) error. The error can be measured in any norm because we are dealing with a finite set of values. Denote the value of the minimum relative reconstruction error in the ith component by $\epsilon_i(S)$ and find the value $\epsilon(S) = \sum_{i=1}^{d} \epsilon_i(S)$.

We *define the* $(\epsilon, \mathcal{F})$*-complexity of continuous vector function* $x(t)$, *which is given by its values on the uniform grid, as* $(-\log S)$. (Hereafter, we write the ϵ-complexity.) In other words, the ϵ-complexity of a vector function is the negative logarithm of the relative fraction of their values required for its reconstruction by methods from the family $\mathcal{F}$, with the relative error no greater than ϵ. In particular, it is "the shortest" description of the vector function (see details in Appendix).

Let us consider the class of vector functions satisfying the Hölder condition. This means that for any $(t, s) \in [0, T] \times [0, T]$, for some constants $L > 0$ and $p > 0$ the following inequality holds:

$$\sum_{i=1}^{d} |x_i(t) - x_i(s)| \le L|t - s|^p. \quad (1)$$

This class of vector functions is very wide. In effect, it includes all vector functions that can be found in applications.

The main idea of our classification method is as follows. For "almost all" vector functions satisfying the Hölder condition, in the case of a sufficiently rich family $\mathcal{F}$ of approximation methods and a sufficiently large sample size n, there exist numbers $0 < \alpha(n) < 1, 0 < \beta(n) < 1$, and $\alpha(n) \le S \le \beta(n)$ (which depend on the vector function) such that the following equality holds:

$$\log \epsilon \approx A + B \log S. \quad (2)$$

The precise meaning of the expression "almost any" and the symbol $\approx$ is given in the Appendix. The parameters A, B are called the ϵ*-complexity coefficients*. These ϵ-complexity coefficients will be utilized as features for the classification of multichannel-EEG records.

These features are independent from the data generating mechanism, and therefore our approach is *model-free*. In the scalar