Two Approaches to Machine Learning Classification of Time Series
Based on Recurrence Plots

Lyudmila Kirichenko
Department of Applied Mathematics
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
lyudmila.kirichenko@nure.ua

Tamara Radivilova
V.V. Popovsky dept. of Infocommunication Engineering
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
tamara.radivilova@gmail.com

Vitalii Bulakh
Department of Applied Mathematics
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
bulakhvitalii@gmail.com

Petro Zinchenko
Department of Applied Mathematics
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
petro.zinchenko@nure.ua

Abed Saif Alghawli
Department of Computer Science
College of Sciences and Humanities
Prince Sattam Bin Abdulaziz University
Aflaj, Kingdom of Saudi Arabia
alghawly@yahoo.com

Abstract—The article considers the task of classifying fractal time series based on the construction of their recurrence plots. Short realizations of EEG signals were used as input data. Two classification methods were considered: the first case, quantitative fractal and recurrent characteristics of the time series were classification features, in the second case, image recognition of recurrence plots was carried out. The results showed fairly high classification quality for both methods, and relative advantage of the image classification method.

Keywords—fractal time series, classification of time series, recurrence plot, machine learning classification

I. INTRODUCTION

Time series analysis is found in many practical applications ranging from human activity detection to cybersecurity. In many cases, the problem of time series classification occurs. Moreover, any classification task that uses ordered data can be considered as a time series classification task.

Over the past decades, time series classification methods based on machine learning have been developed [1-5]. One of the most popular approaches to classification is to extract from time series a set of some features, which are input data for the classifier [6-9].

Nowadays methods of nonlinear dynamics, including the method of recurrence plots, are widely used for time series analysis. Recurrence analysis is based on repeatability of time series states and allows presenting a time series as a geometric structure. The topology of such geometric structure allows to reveal and analyze the characteristic features of time series dynamics of different nature [10, 11].

With the development of machine learning methods and approaches to the selection of features for classification, a number of researchers began to use quantitative recurrence characteristics of time series in the tasks of classification. Thus, for example, in [12] the classification of time series is considered by an example of video compression algorithms with the use of distance measure of cross recurrence plots; in [13], using the support vector machine, the heart rate variability is classified based on recurrence features; in [14] support vector machine algorithm and features of recurrence plots are used.

The other approach to time series classification is the visualization of recurrence plots and image recognition using deep neural networks. In [15], this approach is proposed to recognize and monitoring user behavior; in [16] and [17], the recognition of recurrence plots using convolutional neural networks was applied to different types of time series.

The classification of time series with fractal properties is of particular interest. Numerous studies have shown that time series, changing the character of dynamics, also change the fractal properties. This concerns attacked traffic [18], biomedical signals [19], financial series [20], etc. In [21–23], it was shown that change in the fractal properties of the time series leads to change in the recurrence characteristics. Thus it is possible to use recurrence quantitative characteristics as features for classification of fractal time series [23-25] or to use visualization of recurrence plots for classification [26].

The purpose of the present work is comparative analysis of two approaches to classifying fractal time series: first use fractal and recurrence quantitative characteristics as features for a classifier and second base on the classification of recurrence plot images

II. RECURRENCE PLOTS

A recurrence plot is a matrix of distances between points of a time series in pseudophase space. It is assumed that the analyzed time series \( x(t) \), \( t = 1, \ldots, N \) is the trajectory of some dissipative system. Using the Packard-Takens method [27], one can reconstruct attractor of the system in the m-dimensional phase space from this trajectory, where the phase variables are time-delay values of \( x(t) \).
\[ F = (x(t), x(t + \tau), \ldots, x(t + (m-1)\tau)) \]

where \( \tau \) is the time lag.

In this case, the recurrence plot is a square matrix \( RP_{i,j} \), \( i, j = 1, \ldots, N \). An element of this matrix \( RP_{i,j} \) is equal to 1 if the distance between the points \( x(t_i) \) and \( x(t_j) \) in the \( m \)-dimensional phase space \( F \) does not exceed some pre-assigned value \( \varepsilon \). Otherwise, \( RP_{i,j} \) equal to 0. [28]

If put the recurrence plot elements \( RP_{i,j} \) according black and white, a some topological structure of the time series is got. Qualitative and quantitative topological analysis of the recurrence structures allows to classify time series based on their recurrence 1 properties. Quantitative measures of recurrence structures were discussed in detail in [29,30]. Consider the main ones.

Recurrence rate (RR) shows the density of recurrence points, i.e., the probability of recurrence \( x(t_i) \) in the \( m \)-dimensional phase space:

\[
RR = \frac{1}{N^2} \sum_{i,j=1}^{N} RP_{i,j}.
\]

Important in the topology of recurrence structures are diagonal structures, which show the time interval when one time series section comes close enough to another. Deterministic time series, for example, periodic ones, have long diagonal lines, while stochastic series are either very short or do not have them at all.

To calculate quantitative characteristics based on diagonal structures, we consider the frequency distribution of the lengths \( l \) of diagonal lines in a recurrence plot:

\[
P(l) = \{l_i;i=1,\ldots,N_l\},
\]

where \( l_i \) is length of \( i \)-th diagonal line, \( N_l \) is number of diagonal lines. The average length of the diagonal lines is calculated as follows:

\[
L = \frac{\sum_{i=1}^{N_l} l_i \cdot P(l_i)}{\sum_{i,j=1}^{N} P(l_i)}.
\]

The determinism (DET) or system predictability measure is called the following relation of recurrence points:

\[
DET = \frac{\sum_{i}^{N_l} l_i \cdot P(l_i)}{\sum_{i,j=1}^{N} RP_{i,j}}.
\]

Denote the frequency distribution for the set of lengths of vertical lines as

\[
P(v) = \{v_k;k=1,2,\ldots,K\},
\]

where \( v_k \) is length of \( k \)-th vertical line, \( K \) is number of vertical lines.

The laminarity measure (LAM) characterizes the presence of system states when the system motion along the phase trajectory practically stops:

\[
LAM = \frac{\sum_{i,j=1}^{N} v P(v)}{\sum_{i,j=1}^{N} RP_{i,j}}.
\]

Quantitative measures of recurrence can be used as features in the classification using machine learning [23-25].

III. FRACTAL CHARACTERISTICS OF TIME SERIES

The fractality of stochastic processes is to preserve distribution laws when changing the time scale. A stochastic process \( X(t) \) is self-similar with a parameter \( H \) if the process \( a^{-H}X(at) \) is described by the same distribution laws as \( X(t) \). The parameter \( H, 0 < H < 1 \), is called the Hurst exponent and it is the self-similarity measure. The moments of the self-similar stochastic process satisfy the following scaling relation

\[
E\left[X(t)^q\right] \propto t^{qH}.
\]

Multifractal stochastic processes are inhomogeneous fractal ones. Scaling relation for their moment characteristics is described by

\[
E\left[X(t)^q\right] \propto t^{h(q)},
\]

where \( h(q) \) is generalized Hurst exponent. The Hurst exponent \( H \) of multifractal processes is equal to the value \( h(q) \) at \( q = 2 \). [31]

There are many. One of the most practical approaches to estimate the fractal characteristics by time series is the method of multifractal detrended fluctuation analysis (MFDFA) [32]. MFDFA allows to calculate the fluctuation function \( F(\tau) \) which has scaling relation \( F_q(\tau) \propto \tau^{h(q)} \).

IV. CLASSIFICATION METHODS

As a result of a number of preliminary experiments, the simplest type of feedforward neural network (perceptron) was selected as a classifier based on quantitative features. In this work the perceptron with seven layers was used.

The ReLU (rectified linear unit) semi-line function was used as an activation function. The main advantage of this function is the multifold increase in the rate of the gradient descent convergence in comparison with traditional activation functions, which is due to the linear nature of the function [33].

After each fully connected layer of the neural network, the regularization layer was connected. The method of batch normalization was used in layer regularization to improve the efficiency of neural network learning and to solve the problem of overfitting [34].

Batch normalization reduces the covariance shift, i.e. the divergence of distribution parameters of features values in the
learning and test samples (mathematical expectation, dispersion, etc.). That is, when using normalization batches, the input data are normalized in such a way as to obtain a zero math expectation and a unit dispersion. Normalization is performed before entering each layer.

Also batch normalization has a number of advantages, including: faster convergence of models is achieved, each layer of the network is trained more independently of the others; batch normalization is a mechanism of regularization, because it brings some noise to the outputs of nodes hidden layers, etc.

The Adam (adaptive moment estimation) algorithm was used as the optimization method for the neural network [35]. In contrast to the classic stochastic gradient descent, which supports a single learning rate, Adam calculates individual adaptive learning rates for each parameter calculated on the basis of the first and second gradient moments.

The Adam algorithm has many advantages, among which are the following: it is simple to implement, computationally efficient, has small memory requirements, and is good for tasks with a large number of parameters.

The most frequently used for image classification are convolution neural networks, which are deep neural networks. Deeper neural networks should theoretically work better than smaller models. But the big problem is optimization in deep models, as they are more difficult to optimize because gradients spread from upper to lower layers. [36]

To solve this problem, a deep residual network architecture was developed, based on a residual block, which is a number of bundled layers with activations, with a shortcut connection (fast connections). The basic idea of the residual network is that instead of the usual serial connection of layers of neural network, a fast access connection is used, which allows the signal to pass almost without changes. [37,38]

In this case, if on some layer the network has already sufficiently well approximated the original function that generates the data then on further layers the optimizer in Res blocks can make weights close to zero, and the data will pass through the shortcut connection almost unchanged, which leads to a much faster convergence and better learning.

In this work, a residual neural network containing 131 layers with weights was constructed for image classification. These layers were divided into 11 separate layers and 120 combined into blocks with a quick access connection. The Adam algorithm was also used as the optimization method for this residual neural network.

V. INPUT DATA

To conduct a comparative analysis, the well-known dataset was used [39]. It contains records of electroencephalograms (EEG) for various human conditions. It is widely known that EEG realizations have fractal properties, which vary depending on the state of a person [19, 40]. It allows us to use fractal and recurrent characteristics as features for classifier.

This data set, which contains 5 classes of EEG records, is traditionally used for binary classification: epileptic seizure data (one class) and all other cases (four classes). Each class consists of 100 files, where each file corresponds to one object. Every file contains 23 records with length of 178 values, which corresponds to 1 second. Thus, the data is divided into 2 classes, where the first contains 2300 time series, and the second - 11500.

Fig. 1 represents two EEG realizations from the first class (above) and two realizations from the second class (below).

Fig. 2 shows the recurrence plots constructed by the above time series. At the top of Fig. 2, plots of the first class are represent, and at the bottom - of the second class. These recurrence plots were the basis of inputs to neural networks. In the first case, the recurrence plot was the matrix for calculation of quantitative recurrence characteristics, and in the second case, recurrence plot was image for recognition.

VI. EXPERIMENT DESCRIPTION

When classifying the entire data set was used (11,500 time series), where 8,500 time series were selected for training (6,800 cases with epilepsy and 1,700 without it) and 3,000 for testing (2,400 cases with epilepsy and 600 without it).

Consider the case when the classification was carried out according to quantitative characteristics. Fig. 3 shows the time realizations and the corresponding functions of the generalized Hurst exponent $h(q)$, evaluated by the MFDFA method.
The left part of Fig. 3 represents a typical realization and function $h(q)$ for the 1st class of EEG records, and on the right part the 2nd class realization with the corresponding function $h(q)$ is shown. Preliminary studies have demonstrated that realizations of different classes have significant differences in the range of generalized Hurst exponent $h(q)$ and Hurst exponent $H$ values. It should be noted that the short length of the time series (178 values) causes large errors in the direct estimation of generalized Hurst exponent [5, 8].

Fig. 4 shows the recurrence plots for realization presented in Fig. 3.

For each recurrence plot, recurrence characteristics, such as determinism, average length of the diagonal lines, laminarity and others were obtained. Table 1 presents some quantitative recurrence characteristics corresponding to the plots shown in Fig. 3. We can note a significant difference between the characteristics of different classes realizations. Studies have shown that such differences are typical.

<table>
<thead>
<tr>
<th>Class</th>
<th>RR</th>
<th>Det</th>
<th>$L$</th>
<th>LAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.027</td>
<td>0.147</td>
<td>4.58</td>
<td>0.293</td>
</tr>
<tr>
<td>2</td>
<td>0.035</td>
<td>0.031</td>
<td>3.12</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Thus, in the case of classification based on quantitative characteristics, the following features were received at the input of a fully connected perceptron:

- statistical characteristics (median, coefficient of variation, etc.) calculated from the values of the normalized EEG time series;
- fractal characteristics (the value of the Hurst exponent, the range of values of the generalized Hurst exponent and some its values, etc.) obtained from sample generalized Hurst exponent which was estimated by the MFDFA method;
- recurrent characteristics, such as $DE', LAM, L$, etc., obtained from the recurrence plot of EEG time series recurrence plot.

In the second case (classification based on recurrence plots), black-and-white images of recurrence plots of EEG time series were provided to the input of a residual neural network.

VII. RESULTS AND DISCUSSION

As a result of binary classification, in both cases confusion matrix with True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) values was obtained. The following metrics were selected as the classification results.

**Accuracy** is proportion of correct algorithm answers:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}.$$  

This metric is poorly informative in tasks with unbalanced classes. Since in our task the distribution of EEG records by classes is uneven, it is also necessary to consider other metrics.

**Precision** shows the proportion of objects that are called positive by the classifier and are truly positive:

$$\text{Precision} = \frac{TP}{TP + FP}.$$  

**Recall** shows what proportion of positive class objects from all objects of a positive class the algorithm found:

$$\text{Recall} = \frac{TP}{TP + FP}.$$  

**Recall** demonstrates the ability of the algorithm to detect a given class in general, and **Precision** demonstrates the ability to distinguish this class from other classes. **Precision** and **Recall** do not depend, unlike **Accuracy**, on the correlation of
classes and are therefore applicable in conditions of unbalanced samples.

One of several ways to combine Precision and Recall metrics into one aggregated criterion is to calculate the Fmeasure, in this case it is harmonic mean of Precision and Recall:

\[
F \text{ measure} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Table 2 shows the classification metrics obtained after the experiment by both methods.

<table>
<thead>
<tr>
<th>TABLE II. CLASSIFICATION METRICS</th>
<th>Classification by quantitative characteristics</th>
<th>Classification by recurrence plots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.97</td>
<td>0.984</td>
</tr>
<tr>
<td>Precision</td>
<td>0.93</td>
<td>0.9792</td>
</tr>
<tr>
<td>Recall</td>
<td>0.926</td>
<td>0.94</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.928</td>
<td>0.9592</td>
</tr>
</tbody>
</table>

Both methods showed a fairly high classification quality. The obvious advantage of image-based classification allows us to make the assumption that the most important features for recognition are recurrence ones. In this case, the recurrence plots themselves, presented in the form of black and white images, contain more information about the time series than the quantitative characteristics calculated from them. It can also be assumed that fractal characteristics calculated over short time series do not carry enough information to create the advantage of classification according to the quantitative characteristics of the time series.

It is worth noting that in [41], the classification results of the considered dataset using the machine learning algorithms Artificial Neural Networks, Naive Bayesian, k-Nearest Neighbor, Support Vector Machines and k-Means are presented. These results do not exceed the results presented in this paper.

VIII. CONCLUSION

A comparative analysis of the machine learning classification methods of fractal time series based on the use of recurrence plots have been carried out. The data for the binary classification were the EEG records. One class was represented by EEG realizations with epileptic seizure, another - without a seizure.

Two different classification methods have been used. In the first, classification on the basis of quantitative characteristics of the time series, such as fractal and recurrent was carried out. In the second, images of recurrence plots were classified.

In both approaches, neural networks were used as classifiers. In the case of classification based on quantitative characteristics, fully connected perceptron with regularization layers was chosen. In the case of classification based on recurrence plots images, the residual neural network was a classifier.

Both methods showed a fairly high classification quality. Higher accuracy was shown by the method of image recognition. The results can be used to classify biomedical signals with fractal properties.

Future research will be aimed at improving the network architecture for recognizing images of recurrence plots and development of methods for obtaining more “contrasting” recurrence plots in the sense of image classification.

REFERENCES


