A NEW CHAOTIC FEATURE FOR EEG CLASSIFICATION BASED SEIZURE DIAGNOSIS

Su Yang1,2, Anqin Zhang1, Jiulong Zhang2, Weishan Zhang3

1Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai, China
2School of Computer Science, Xi’an University of Technology, Xi’an, China
3Department of Software Engineering, China University of Petroleum, Qingdao, China

ABSTRACT

Seeking effective measures to characterize the chaotic patterns of EEG signals for seizure diagnosis is a long-term endeavor in the literature. We propose to count the number of zero-crossing (ZC) points on Poincaré surface as a feature when the time series of interest is embedded into the reconstructed state space. The experiments show that Poincaré surface can act as a platform to observe the chaotic patterns of EEG signals and the ZC feature on Poincaré surface is a promising pattern descriptor to discriminate different categories of EEG signals. When used alone for EEG classification, the ZC feature achieves 100%, 99.27%, and 94.68% accuracy in 2-class, 3-class, and 5-class classification on a widely used benchmark.

Index Terms—Nonlinear time series analysis, feature extraction, pattern recognition, EEG signal classification

1. INTRODUCTION

EEG signal classification plays an important role in seizure detection and diagnosis. Feature extractor and classifier are the two essential components for an EEG classification system and have received much attention so far. To date, a variety of features as well as classifiers have been developed in the literature [17]. The widely applied features include nonlinear measures, such as approximate entropy [2][3], Lyapunov exponents [4][5], and fractal measurements [35][36], the spectrum analysis based features [3][6][7][8], the wavelet transformation based features [4][9][10][11][12] [39], the time-frequency analysis based features [13][38] [24], sparse representation [1], the high order statistics [37], and zero-crossing intervals in time series [22]. The classifiers can be sorted into the following 5 categories: neural networks [5][10][13][14][15], support vector machines [3][4][8][16][17], neuro-fuzzy inferences systems [2][11], Gaussian mixture models [22][17], and ensemble methods [7][12].

Owing to the many features available in the literature, better classification performance can be promised by using the so-called feature selection technique to select the best discriminative features [20]. For this reason, the effort of seeking effective features for EEG classification has never been stopped and new methodologies are proposed increasingly. In [16], the spectral-envelope-based speech recognition features are introduced into the EEG classification literature. In [35][36], nonlinear analysis is regarded as an emerging methodology for EEG classification. In fact, nonlinear feature is always one of the major concerns in terms of EEG signal classification [18] due to the experimentally verified chaotic nature of EEG signals [19]. However, we are suffering from the lack of methodologies to measure the chaotic signatures of EEG signals.

The aforementioned efforts have not resulted in a sound nonlinear feature that can be applied alone to achieve reasonably high classification accuracy. In fact, most nonlinear features function as auxiliary features to augment the other features and are not used alone for EEG classification. On account of that, it is meaningful to seek feature extraction means from other domains, for example, the speech features applied to EEG signal classification in [16]. It is known that acoustic signals are also chaotic and a couple of nonlinear features in reconstructed state space have been proposed for acoustic signal classification, including zero-crossing (ZC) statistics [21] on Poincaré Surface. In this study, we further explore the possibility of applying the ZC feature to EEG classification. Some encouraging results have been obtained in the experiments: When used alone, it achieves 100%, 99.27%, and 94.68% accuracy on the widely used benchmark [19] for 2-class, 3-class, and 5-class classification, respectively. As for the other features based on state space reconstruction such as the locally linear embedding (LLE) feature [23] and Lyapunov exponents [32], the performance is quite poor. It is known that Lyapunov exponents characterize the dynamical behaviors while the LLE feature captures the geometric profiles of chaotic signals. In contrast, the ZC feature fuses geometric and dynamic information in one descriptor. This accounts for why it performs better than the other chaotic features. Moreover, the ZC feature promises better or comparable performance in comparison with the state-of-

Index Terms—Nonlinear time series analysis, feature extraction, pattern recognition, EEG signal classification

1. INTRODUCTION

EEG signal classification plays an important role in seizure detection and diagnosis. Feature extractor and classifier are the two essential components for an EEG classification system and have received much attention so far. To date, a variety of features as well as classifiers have been developed in the literature [17]. The widely applied features include nonlinear measures, such as approximate entropy [2][3], Lyapunov exponents [4][5], and fractal measurements [35][36], the spectrum analysis based features [3][6][7][8], the wavelet transformation based features [4][9][10][11][12] [39], the time-frequency analysis based features [13][38] [24], sparse representation [1], the high order statistics [37], and zero-crossing intervals in time series [22]. The classifiers can be sorted into the following 5 categories: neural networks [5][10][13][14][15], support vector machines [3][4][8][16][17], neuro-fuzzy inferences systems [2][11], Gaussian mixture models [22][17], and ensemble methods [7][12].

Owing to the many features available in the literature, better classification performance can be promised by using the so-called feature selection technique to select the best discriminative features [20]. For this reason, the effort of seeking effective features for EEG classification has never been stopped and new methodologies are proposed increasingly. In [16], the spectral-envelope-based speech recognition features are introduced into the EEG classification literature. In [35][36], nonlinear analysis is regarded as an emerging methodology for EEG classification. In fact, nonlinear feature is always one of the major concerns in terms of EEG signal classification [18] due to the experimentally verified chaotic nature of EEG signals [19]. However, we are suffering from the lack of methodologies to measure the chaotic signatures of EEG signals.

The aforementioned efforts have not resulted in a sound nonlinear feature that can be applied alone to achieve reasonably high classification accuracy. In fact, most nonlinear features function as auxiliary features to augment the other features and are not used alone for EEG classification. On account of that, it is meaningful to seek feature extraction means from other domains, for example, the speech features applied to EEG signal classification in [16]. It is known that acoustic signals are also chaotic and a couple of nonlinear features in reconstructed state space have been proposed for acoustic signal classification, including zero-crossing (ZC) statistics [21] on Poincaré Surface. In this study, we further explore the possibility of applying the ZC feature to EEG classification. Some encouraging results have been obtained in the experiments: When used alone, it achieves 100%, 99.27%, and 94.68% accuracy on the widely used benchmark [19] for 2-class, 3-class, and 5-class classification, respectively. As for the other features based on state space reconstruction such as the locally linear embedding (LLE) feature [23] and Lyapunov exponents [32], the performance is quite poor. It is known that Lyapunov exponents characterize the dynamical behaviors while the LLE feature captures the geometric profiles of chaotic signals. In contrast, the ZC feature fuses geometric and dynamic information in one descriptor. This accounts for why it performs better than the other chaotic features. Moreover, the ZC feature promises better or comparable performance in comparison with the state-of-
the-art methods according to the performance data reported in the literature.

2. FEATURE EXTRACTION ON POINCARE SURFACES IN RECONSTRUCTED STATE SPACE

In [21], a feature referred to as geometric statistical feature in state space is proposed for acoustic signal classification, which is the zero-crossing (ZC) statistics on Poincaré surface. Since EEG signals are confirmed to be nonlinear signals with chaotic and fractal properties [35][36], in this study, we explore whether the chaotic measure proposed in [21] is useful for EEG classification. The feature is computed through the following 3 steps:

2.1. State Space Reconstruction

The state space reconstruction is a widely used technique for nonlinear time series analysis [25][26]. By means of it, the time series [S][i=1,2,...,N_T] can be embedded to a high-dimensional space in the form of a N-dimensional trajectory [X[j][j]=1,2,...,M] by letting X[j]=(S_j,S_{j+1},S_{j+2},...,S_{j+(N-1)})^T, where the delay time J and the embedding dimension N are two parameters to be determined at first. In computing J, we refer to [27]. Let A(k) represent the autocorrelation function of the time series, where k denotes the discrete-time step. Once A(k) drops below A(0)e, where e=exp(1), let J=k. The determination of N is application-dependent [28], the optimal value of which is usually obtained by trial and error.

Following state space reconstruction, we can obtain a high-dimensional trajectory reconstructed from the time series of interest. Such high-dimensional vector sequence can be viewed as a point cloud, which is also referred to as manifold. The spatial configuration of the manifold and the dynamic evolution of the trajectory are the key factors to characterize different chaotic attractors arising from various dynamic systems. Here, we make use of the zero-crossing statistics on Poincaré surface of section to capture the geometric and dynamic characteristics at the same time and fuse the geometric and dynamic information in one descriptor, which results in better classification performance than using each descriptor alone.

2.2. Orientation and Position Normalization

Prior to feature extraction, the point cloud should be normalized in terms of orientation and position. The reason is: the point clouds undergoing classification can be viewed as N-dimensional data objects. However, such data objects could possess different poses following state space reconstruction, which prevent us from comparing the data objects in terms of geometric similarity. Therefore, the data objects should be normalized at first to promise that the subsequent pattern matching is not affected by the rotation, scaling, and transition of the data objects. Note that the trajectories reconstructed from the same class of signals should be similar in shape but the pose of every reconstructed trajectory may vary, so orientation normalization is performed at first. Here, the orientation normalization is obtained through principal component analysis (PCA) [29]. Let \( a_1 \geq a_2 \geq \cdots \geq a_N \) represent the eigenvalues of \( XX^T \), where \( X=[X_1,X_2,...,X_n] \) is the matrix being composed of the previously defined trajectory data. Let \( U_1,U_2,...,U_N \) represent the corresponding eigenvectors. Then, perform the coordinate transformation with regard to every trajectory point as follows:

\[
Y_j = (U_1^T X_j, U_2^T X_j,..., U_N^T X_j)^T
\]

The direction determined by \( U_1 \) corresponds with the primary axis on which the projections of the trajectory points have the biggest variance in terms of statistics. Then, \{U_1,U_2,...,U_N\} forms an orthogonal base.

Position normalization is performed as follows: Let \( Y=(y_1,y_2,...,y_N)^T \) represent the coordinate values of point \( Y_j \). Let \( \bar{Y}=(\bar{y}_1,\bar{y}_2,...,\bar{y}_N)^T \) represent the center of the entire trajectory, \( \{Y_j\}_{j=1,2,...,M} \). The center is defined as

\[
\bar{y}_i = m_{1,0}/m_{0,0},...,\bar{y}_N = m_{0,1}/m_{0,0}
\]

where

\[
m_{j,...,p_N} = \sum_{j=1}^{M} \prod_{i=1}^{N} y_{ji}^{j,...,p_N}
\]

is the regular moment of all trajectory points with order \((P_1,P_2,...,P_N)\) [30]. The position normalization is achieved via the following translation:

\[
Y_j \leftarrow Y_j - \bar{Y}
\]

After such position normalization, the center of the trajectory is positioned at the original.

2.3. Zero-crossing Statistics on Poincaré Surface of Section

Following the aforementioned normalization, we can then observe the characteristics of the EEG signal of interest on the so-called Poincaré surface of section, which is a hyperplane to intersect with the high-dimensional trajectory of the signal in the reconstructed state space. Poincaré surface of section is a widely used means for studying chaotic attractors. Before one hundred years, it was observed by Poincaré that the trajectory points on Poincaré surfaces exhibit regular structures for some chaotic attractors. In Poincaré surface of section, we can see where and when the trajectory passes through such hyperplane. The points at which a trajectory passes through a Poincaré surface are referred to as points of section, which reflect not only the spatial configuration of the manifold but also the evolution of the trajectory due to the time stamps on the points of section. Intuitively, such information should be useful for signal classification.

4652
Supposing that a Poincaré surface intersects perpendicularly with the i-th axis at position \( y_0 \), the trajectory points intersecting with this Poincaré surface is the union of two sets, denoted as \( P=\mathcal{A}\cup\mathcal{B} \):

\[
\mathcal{A} = \{ Y(y_0) = (y_1, y_2, \ldots, y_M)^T | y_0 \in [1,M] \} \quad (5)
\]

\[
\mathcal{B} = \{ Y' = I(Y(y_j), Y(y_{j+1}))(y_j - y_0)(y_{j+1} - y_0) \leq 0; j \in [1,M-1] \} \quad (6)
\]

where \( I(Y(y_j), Y(y_{j+1})) \) is the linear interpolation between \( Y_j \) and \( Y_{j+1} \), which is defined as

\[
I(Y_j, Y_{j+1}) = Y_j + \frac{y_{j+1} - y_j}{y_{j+1} - y_j}(Y_{j+1} - Y_j) \quad (7)
\]

Note that the i-th coordinate value of all the points on the Poincaré surface is a constant \( y_0 \). A constant attribute is meaningless in the sense of classification, so we eliminate the i-th coordinate value of all the points contained in \( P \). Then, we rewrite the point set on the Poincaré surface as \( P = \{ Z = (z_1, z_2, \ldots, z_M)^T | i = 1, 2, \ldots, K \} \). It is obvious that \( K \leq M \). Note that \( L = N - 1 \) is the data dimension after removing the constant coordinate value \( y_0 \).

The number of the points contained in \( P \), say zero-crossing rate on Poincaré surface of section, is subject to the shape of the trajectory, so it can act as a measure to distinguish the shape signatures of different EEG signals. Note that a Poincaré surface can intersect perpendicularly an axis at any position. In this study, we let every Poincaré surface intersect the corresponding axis at the coordinate origin.

### 3. Experiments

We use the data developed in [19] to evaluate the performance of the proposed feature for EEG classification. This data set is composed of 5 classes of EEG signals denoted as Z, O, N, F, S, each of which contains 100 samples of 23.6 second duration with 173.61Hz sampling rate. The data length of every signal is 4097. A Bandpass filter of 0.53-80Hz is applied prior to data sampling. The description of the 5 classes is provided in Table 1.

For every class, 50% samples are selected randomly to train the classifier while the other samples are used for testing. Test as such is repeated 10 times with randomly selected training and testing samples with regard to each run. The performance data reported in the following is the average over the 10 tests. We use the classification accuracy as the performance index, which is the ratio of the number of the correctly classified samples to that of the total samples for each class, referred to as sensitivity in [13]. We use support vector machine (SVM) as the classifier, where we apply 10-fold cross validation on the training data to obtain the best parameters for SVM. We extract the ZC feature with 3 embedding dimensions, 10, 15, and 20, respectively. For every axis, a Poincaré surface is applied to intersect perpendicularly with it at the origin. Since we obtain one ZC-based variable per Poincaré surface, the 10, 15, and 20 embedding dimensions lead to 45 variables/features in total, that is, the dimension of the feature space is 45.

The ZC feature captures both manifold and dynamic information [21]. Here, we compare it with two chaotic descriptors focused on either manifold or dynamics. One is principal component analysis on locally linear embedding (LLE) [31] of reconstructed state space, which is a pure manifold based feature for EEG signal classification [23]. The other is the largest Lyapunov exponents [32], which is a pure dynamics based feature.

In computing the ZC and LLE feature, we let the embedding dimension be 10, 15, and 20, respectively. As for the LLE feature, we let the number of nearest neighbors for data fitting be 10, 20, and 30, respectively. Regarding the Lyapunov feature, we take into account 50 evolutionary steps. Accordingly, the dimension of the aforementioned features should be 45, 180, and 150, respectively.

#### Table I: Description of the Data Set

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>5 healthy persons with eye open.</td>
</tr>
<tr>
<td>O</td>
<td>5 healthy persons with eye close.</td>
</tr>
<tr>
<td>N</td>
<td>5 patients in seizure-free intervals with signals from epileptogenic zone</td>
</tr>
<tr>
<td>F</td>
<td>5 patients in seizure-free intervals with signals from the opposite zone</td>
</tr>
<tr>
<td>S</td>
<td>5 patients during seizure-active period</td>
</tr>
</tbody>
</table>

#### Table II: Comparison of the Chaotic Features on 3-Class Classification (%)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Z &amp; O</th>
<th>N &amp; F</th>
<th>S</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZC</td>
<td>99.9</td>
<td>99.5</td>
<td>98.4</td>
<td>99.27</td>
</tr>
<tr>
<td>LLE</td>
<td>97.5</td>
<td>96.2</td>
<td>90.6</td>
<td>94.77</td>
</tr>
<tr>
<td>Lyapunov</td>
<td>93.1</td>
<td>91</td>
<td>76.4</td>
<td>86.83</td>
</tr>
</tbody>
</table>

#### Table III: ZC Against LLE on 5-Class Classification (%)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Z</th>
<th>O</th>
<th>N</th>
<th>F</th>
<th>S</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZC</td>
<td>92.8</td>
<td>89.4</td>
<td>94.6</td>
<td>98.8</td>
<td>97.8</td>
<td>94.68</td>
</tr>
<tr>
<td>LLE</td>
<td>83.2</td>
<td>83.4</td>
<td>75</td>
<td>63.4</td>
<td>91.2</td>
<td>79.24</td>
</tr>
</tbody>
</table>

#### Table IV: Classification Accuracy Using ZC with Different Classifiers (%)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Z</th>
<th>O</th>
<th>N</th>
<th>F</th>
<th>S</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>92.8</td>
<td>89.4</td>
<td>94.6</td>
<td>98.8</td>
<td>97.8</td>
<td>94.68</td>
</tr>
<tr>
<td>1-NN</td>
<td>98.2</td>
<td>75.6</td>
<td>94.2</td>
<td>92.6</td>
<td>92.4</td>
<td>90.6</td>
</tr>
<tr>
<td>5-NN</td>
<td>99.8</td>
<td>72.2</td>
<td>90.8</td>
<td>95.8</td>
<td>89.2</td>
<td>89.56</td>
</tr>
<tr>
<td>Bayes</td>
<td>93.2</td>
<td>81.8</td>
<td>78.2</td>
<td>90.2</td>
<td>92.6</td>
<td>87.2</td>
</tr>
</tbody>
</table>

The performance comparison on 3-class classification is listed in Table 2, where the overall classification accuracy is sorted in descending order as 99.27%, 94.77%, and 86.83% for ZC, LLE, and Lyapunov exponents, respectively. As the performance of Lyapunov exponents is far from that of the ZC feature, we only compare LLE with ZC on the 5-class classification problem, the performance of which is shown in Table 3. We can see that the averaged classification accuracy of ZC drops a little from 99.27% to 94.68% but that of LLE drops drastically from 94.77% to only 79.24%. According to the comparison, we see that ZC is the best...
feature for EEG classification among such chaotic features. LLE performs better than the largest Lyapunov exponent but worse than the ZC feature. Besides, its discriminant power in classifying the 5 classes is limited. ZC incorporates both geometric and dynamic information since it is subject to not only trajectory evolution but also the spatial configuration of trajectory points. This account for why it outperforms remarkably the other chaotic features.

A comprehensive performance comparison is provided in Table 5–7 on the same benchmark and it can be seen that the proposed feature is comparable to the state-of-the-art solutions. As for the 2-class classification problem, expect for the proposed method, many other features perform also very well with 100%, for instance, STFT [38], time-frequency analysis [13], wavelets [39], high-order statistics [37], and Fuzzy approximate entropy [33]. The time-frequency feature applied in [13] [38] also achieves the best 3-class classification accuracy of 100%. However, the performance of time-frequency analysis on 5-classification is not satisfactory, which is 89% [13]. In terms of 2-class and 3-class classification, the overall performance of the proposed feature is comparable to that of the time-frequency feature, which is 100% and 99.73%, respectively. For the 5-class classification problem, however, the accuracy of the proposed feature is 94.68%, which is obviously better than the 89% accuracy obtained by using the time-frequency feature. The highest classification accuracy regarding 5-class classification is achieved by using the combination of wavelet features and Lyapunov exponents coupled with support vector machine classifier [4]. Also, the individual wavelet or spectral feature combined with fuzzy inference, neural network ensemble, or support vector machine classifier also result in high performance of over 98% classification accuracy [11] [7] [8]. As for the classical nonlinear features like entropy or Lyapunov exponent, however, the performance cannot be guaranteed to be satisfactory if used alone [2] [3] [5]. In contrast, the proposed feature is a more promising feature, which can be used independently to achieve over 94% accuracy in classifying the 5-class EEG signals.

To examine how much classifiers can affect the classification performance, we summarize the performance data resulting from different classifiers in Table 4. It is apparent that support vector machine performs much better than k Nearest Neighbor (kNN) classifier and Naïve Bayes classifier.

4. CONCLUDING REMARKS

This study reveals that Poincaré surface provides a useful means to capture the intrinsic chaotic patterns of EEG signals. The number of zero-crossing points on Poincaré surface is effective in distinguishing different categories of EEG signals. As far as we know, it is a unique nonlinear feature that can be used independently to achieve over 94% accuracy in the 5-class EEG classification problem.
REFERENCES


