International Journal of Frontiers in Engineering Technology
ISSN 2706-655X Vol.5, Issue 4: 19-23, DOI: 10.25236/IJFET.2023.050404

Study on EEG signal of epileptic patients based on NLSOMAP algorithm

Chong Li, Shiyu Hu, Xinyan Zhang, Jiale Chen

School of Communication Engineering, Tongda College of Nanjing University of Posts and Telecommunications, Yangzhou, China

Abstract: *Epilepsy is a chronic disease with sudden abnormal discharge of brain neurons, leading to transient brain dysfunction, which has a great impact on the physical and mental health of patients. In this paper, we use the NISOMAP manifold algorithm to reduce the dimension of high-dimensional data randomly selected from the public epilepsy data set, aiming at the disadvantage that the traditional supervised model is nonlinear in feature extraction. The experimental results show that, compared with other dimension reduction algorithms, NISOMAP has a better dimension reduction effect. The visualized shape is oval, with certain regularity, and it is convenient to observe the attack point. It is superior to other algorithms in different sample data, and provides great help for the diagnosis of epilepsy patients.*

Keywords: *Manifold learning; Visualization; Dimension reduction; Epilepsy; EEG signal*

1. Introduction

Epilepsy, commonly known as "epilepsy" or "epilepsy", is a chronic disease with sudden abnormal discharge of brain neurons, which leads to transient brain dysfunction. Epilepsy, as a chronic disease, has little impact on patients in the short term, but long-term seizures may cause serious impact on patients' physical and mental health. At present, the diagnosis of epilepsy mainly depends on the direct observation of EEG by experienced neuroscientists. However, the EEG collected by most epileptic patients during epileptic seizures does not seem to be different from that of healthy people. Misdiagnosis often occurs depending on artificial diagnosis.

In the feature extraction of EEG signals, Gotman et al. first decomposed EEG signals into "half waves", and then extracted EEG features from them, including average amplitude, duration, coefficient of variation and other features relative to the background, and set thresholds according to expert experience to compare these feature parameters with the threshold to determine whether they are epileptic signals. However, feature extraction methods require prior knowledge of expert experience and frequency bands.

According to previous studies, we found that PCA, K-means, LDA and NMF models are better in the application of feature extraction, but these models are linear and cannot show the local features of data. In addition, LDA is supervised learning, and label information needs to be given during training, but there are few labeled data in the medical field, so it is not ideal for EEG analysis. Recently, most of the research began to favor the unsupervised learning of EEG. Researchers found that the data points in the high-dimensional space are approximately located on a submanifold of the embedded low dimensional space. Therefore, the manifold based model can solve the problems encountered very well[1-2].

2. Research ideas

At present, the mainstream EEG signal analysis methods include time domain analysis, frequency domain analysis and nonlinear dynamics, as shown in Table 1.

Table 1: EEG signal analysis method

Time domain analysis	Frequency domain analysis	Time-frequency analysis	nonlinear dynamics	entropy
Period amplitude analysis	Average periodogram method	Fourier transform	correlation dimension	Shannon entropy
Fujimori method	Welch method	wavelet transform	Lyapunov exponent	Wavelet entropy

ISSN 2706-655X Vol.5, Issue 4: 19-23, DOI: 10.25236/IJFET.2023.050404

In this paper, we use different manifolds and dimension reduction algorithms to apply LLE, LE and NISOMAP algorithms to the epileptic EEG data set disclosed in medical treatment. Under unsupervised conditions, we reduce and cluster EEG signals, and make analysis and comparison. Find out the algorithm with better clustering performance under different sample conditions.

3. Dimension Reduction Algorithm Based on NISOMAP Manifold

3.1 Dimension reduction algorithm based on NISOMAP manifold

3.1.1 Principle of dimension reduction algorithm of NISOMAP manifold

ISOMAP (Isometric Feature Mapping) is a nonlinear dimensionality reduction method. Based on metric MDS, ISOMAP attempts to retain the geometric structure contained by geodesic distance inherent in data. The principle is basically the same as that of MDS. The only difference is that the distance between two points in MDS high-dimensional space is:

$$d_{ij}^{2}=\left\|z_i-z_j\right\|^2=\left\|z_i\right\|^2+\left\|z_j\right\|^2-2z_i^{T}z_j \tag{1}$$

The distance between two points in ISOMAP is the shortest path of two points in the figure. The ISOMAP algorithm is based on the premise that the low dimensional manifold where the data is located is equidistant from a subset of the Euclidean space as a whole. Its core idea is that the distance from the nearest point to the present is replaced by the Euclidean distance, and the distance from the far point is approximated by the shortest path, which is nonlinear. The so-called Euclidean distance is the absolute distance between two points in the multi-dimensional space, which is expressed as:

$$d(x_i,x_j)=\left[\sum_{k=1}^{p}(x_{ik}-x_{jk})^2\right]^{1/2} \tag{2}$$

NISOMAP is an improvement on ISOMAP, which only calculates the geodesic distance from each sample point to Landmark point to generate a dimensional matrix. This algorithm is suitable for learning low dimensional manifolds with relatively flat interior. In the public epilepsy dataset used in this paper, the data clustering effect is obvious after dimension reduction by NISOMAP algorithm[3-6].

3.1.2 Dimension reduction algorithm flow of NISOMAP manifold

Visualize the flow chart of NISOMAP algorithm, as shown in Fig 1.

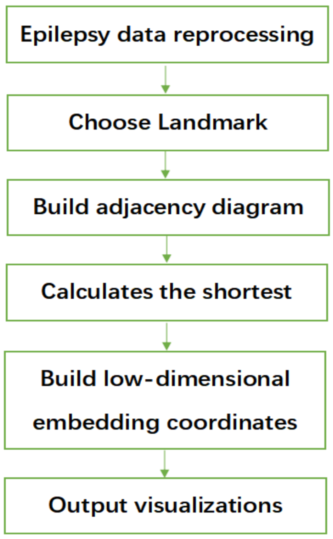

Figure 1: NISOMAP algorithm flow chart

According to the visualization flow chart, the specific steps of NISOMAP algorithm are as follows:

(1) First, the collected epileptic Enbo data set is preprocessed. We selected the data at the time of seizure and the data collected in the epileptic region as sample data segments. Each data segment

ISSN 2706-655X Vol.5, Issue 4: 19-23, DOI: 10.25236/IJFET.2023.050404

contains 4096 sampling points. We set N * 4096 dimensional data, k nearest neighbor value, low dimensional spatial target dimension and other parameters.

(2) Select n Landmark points from N samples: calculate the Euclidean distance between N all sample points and the selected n points to obtain the matrix d . d_{ij} represents the Euclidean distance between sample point x_i and Landmark point x_j .

(3) Build adjacency graph. Based on the Euclidean distance $d_x(i,j)$ between adjacent point pairs i, j on manifold G in the input space x, select k points closest to each sample point or select a constant radius at the sample point ε All the points in the circle of are the nearest neighbors of the sample point, and these adjacent points are connected by edges. The manifold G is constructed as a weighted flow graph G reflecting the proximity relationship. The weight of each edge in the adjacency graph G is $d_x(i,j)$.

(4) Calculate the shortest path. If the sample points x_i and x_j are connected in Fig. G, the initial value of the shortest path between them is equal to $d_G(i,j)$ and $d_x(i,j)$, otherwise $d_G(i,j)$ =∞. Let y=1,2,3 N. Where N is the total number of samples, there are:

$$d_G(i,j) = \min\{d_G(i,j), d_G(i,y) + d_G(y,j)\} \tag{3}$$

From this, the shortest path between any two points on the adjacency graph G is calculated, and the geodesic distance matrix $D_G = \{d_G(i,j)\}$ on the approximation manifold is calculated. The shortest path is mainly implemented by Floyd or Dijkstra algorithm.

(5) Build low dimensional embedded coordinates. According to the graph distance matrix $D_G = \{d_G(i,j)\}$, use the classic MDS algorithm to construct the embedded coordinate representation of data in N-dimensional space Y:

$$E = \left\|\tau(D_G) - \tau(D_Y)\right\|_{L^2} \tag{4}$$

Wherein $\tau(D) = -hsh/2, S_{ij} = D_{ij}^2, H = 1 - \frac{1}{m}ee^T$, and I represent the identity matrix, $e = (1,1,\ldots,1)^T$.

By minimizing the objective function, the maximum eigenvector is obtained, and the optimal embedding coordinate Y is obtained.

4. Experimental results and analysis

The data set is based on the data published by the University of Bonn, Germany. The whole EEG data packet consists of 5 sets, each of which includes 100 single channel EEG data segments. Each data segment contains 4096 sampling points, and the time length is 23.6s. It is worth noting that the artifacts caused by hand or eye movements have been manually deleted by the data creator, so there is no need to preprocess the data. Data sets A and B are extracranial EEG data of five normal volunteers when their eyes are open and closed, respectively. Data sets C and D are intracranial data during the interictal period, while data sets E are data at the time of seizure. The data segment of set D is from the epileptic region, while set C is from the hippocampus region of the opposite cerebral hemisphere, far away from the focus. The sampling frequency of all EEG signals is 173.61Hz, and the frequency band is 0.5~85Hz.

In order to compare the effect of dimensionality reduction visualization, we use three dimensionality reduction algorithms, LLE (Locally Linear Embedding), LE (Laplacian Eigenmaps), and NISOMAP (New Isometric Feature Mapping), to reduce the dimensions of N * 4096 dimensional EEG data, obtain two-dimensional embedded coordinates respectively, and visualize them for easy observation. In this paper, E and D datasets are selected, and 100 groups of data in each dataset are randomly scrambled. A

ISSN 2706-655X Vol.5, Issue 4: 19-23, DOI: 10.25236/IJFET.2023.050404

total of 2 * 4096 dimensional datasets are selected as input data. To facilitate observation, we divide the selected data into four groups for visualization, as shown in Fig 2.

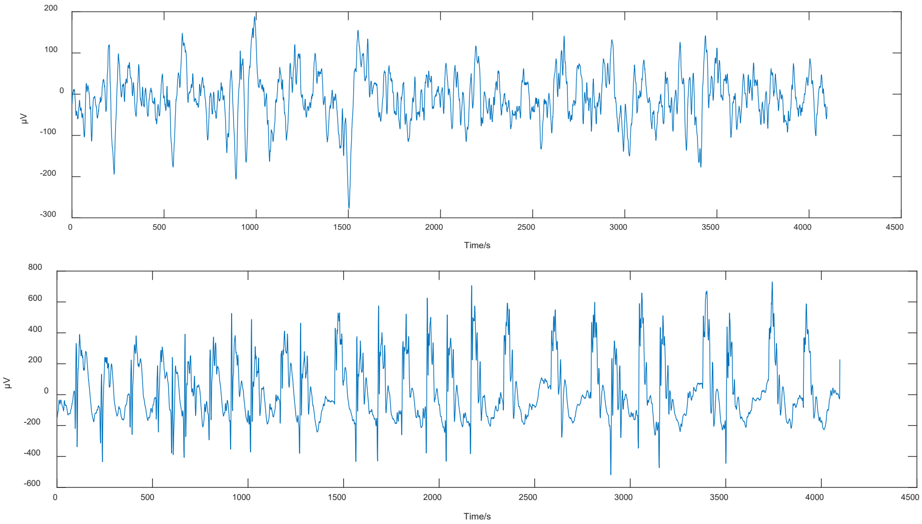

Figure 2: EEG Data Visualization Graph

Apply the selected 2 * 4096 dimensional EEG data to the three algorithms to obtain the reduced dimensional visualization graph, as shown in Fig 3.

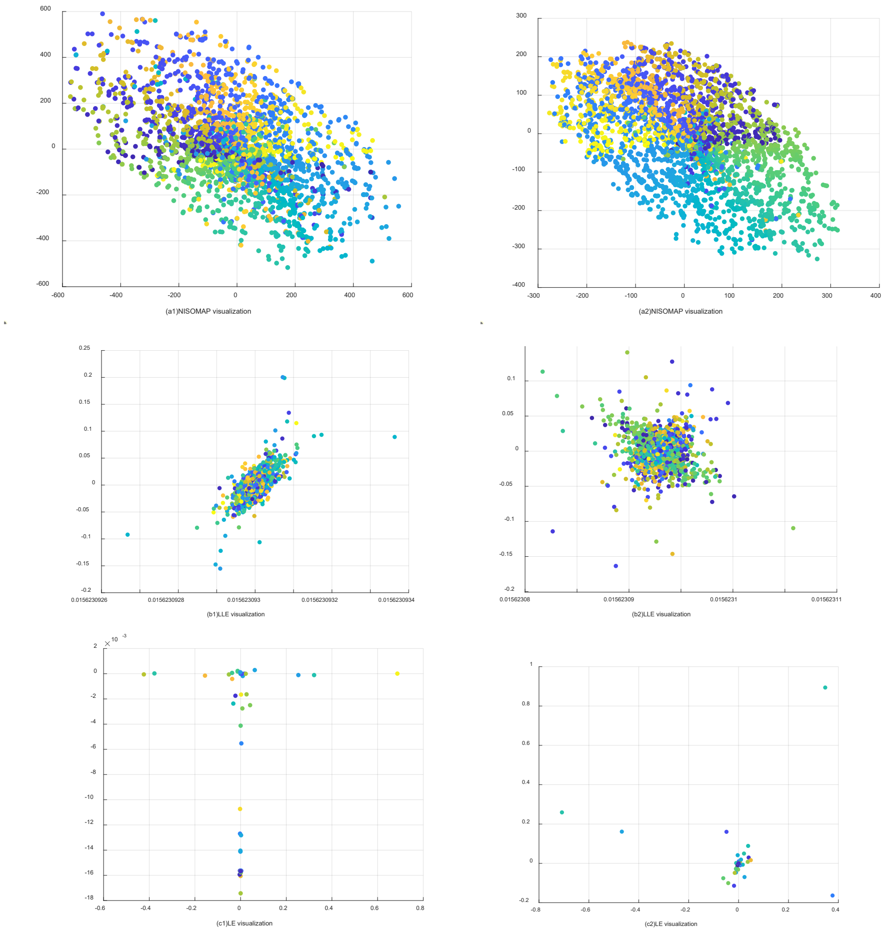

Figure 3: 2D visualization results of three dimension reduction algorithms in epilepsy dataset

ISSN 2706-655X Vol.5, Issue 4: 19-23, DOI: 10.25236/IJFET.2023.050404

4.1 Experimental analysis

This article uses the data of the public dataset, and reduces the dimensions through visualization, as shown in Figure 3. The position coordinates on each picture plane in Figure 3 are the embedded coordinates of 4096 dimensional data after different dimension reduction algorithms. It can be seen intuitively that the LE dimensionality reduction algorithm has serious distortion on the random 2 * 4096 dimensional data after dimensionality reduction, and it is unable to distinguish the attack interval and normal data; The LLE dimension reduction algorithm does not appear serious loss, but its scattered points are confused, overlapping, and there is no obvious clustering effect; After dimensionality reduction, the NISOMAP dimensionality reduction algorithm presents an elliptical distribution, which can clearly identify the attack point data on both sides of the ellipse, without loss or overlapping. Compared with the other two algorithms, the visualization effect after dimensionality reduction is better. Later, we used EEG signals from other locations for multiple tests, and the results were uneven. In addition, we also found some shortcomings. Due to the lack of test data, the test accuracy may be lacking, which is the point to be improved later[7].

5. Conclusion

In this paper, three different manifold algorithms are applied to the public epileptic EEG data set. After the dimensionality reduction visualization processing of randomly selected data, we found that the dimensionality reduction effect on the NISOMAP model is better. After visualization, the shape is oval, the seizure point has a clear boundary with the normal time data point, and the data does not overlap or lose. After dimensionality reduction, it has certain regularity, which is convenient for people to make a preliminary judgment on epileptic seizures.

References

[1] Dong Yingchao, Wang Bin, Ma Sasa, Liu Hui, Xiong Xin, Xue Jie. Research on dimension reduction method of brain network state observation matrix based on t-SNE [J]. Computer Engineering and Application 2018(01)

[2] Wang Ziqiang, Qian Xu, Kong Min. Overview of manifold learning algorithms [J]. Computer Engineering and Application 2008(35)

[3] Zhang Shanwen, Zhang Chuanlei. Gait recognition method based on supervised Isomap [J]. Computer Application Research 2012(11)

[4] Liu Jinling. Isomap based Chinese short message text clustering algorithm [J]. Computer Engineering and Application 2009(34)

[5] Yang Shuo, Wu Songhua, Liu Tongliang, Xu Min. Bridging the Gap between Few-Shot and Many-Shot Learning via Distribution Calibration. [J]. IEEE transactions on pattern analysis and machine intelligence. 2021

[6] Jung Hong-Gyu, Lee Seong-Whan. Few-Shot Learning with Geometric Constraints. [J]. IEEE transactions on neural networks and learning systems. 2020

[7] Lee Sangmin S., Lee Kiwon, Kang Guiyeom. EEG Artifact Removal by Bayesian Deep Learning & ICA [J]. 42ND Annual International Conferences of the Ieee Engineering In Medicine And Biology Society: Enabling Innovative Technologies For Global Healthcare Embc'20. 2020