

Application of Chaos Theory and Data Mining to Seizure Detection of Epilepsy

Chien-Liang Chen^{1,2}, Jenn-Long Liu³⁺ and Jhih-Yu Syu³

¹ Dept. of Information Management, Fortune Institute Technology, Kaohsiung 831, Taiwan

² Dept. of Information Engineering, I-Shou University, Kaohsiung 840, Taiwan

³ Dept. of Information Management, I-Shou University, Kaohsiung 840, Taiwan

Abstract. This study applies Chaos Theory to investigate the seizure detection of epilepsy with three groups of data (Groups H, S and E), showing the electroencephalography (EEG) changes of patients. The three groups, Group H (normal state), Group S (during seizures) and Group E (after seizures), contain 100 series of EEG signals each. The detected data are processed by chaotic theory to transform the signals to four parameters with Delay Time (τ), Embedding Dimension (dm), Correlation Dimension (CD) value, and Largest Lyapunov Exponent (LLE). Furthermore, well-known classification software for data mining, termed See5 based on entropy theory, is used to find out the classification rules for the EEG signals of epilepsy patients.

Keywords: Chaos Theory, Epilepsy, Decision Tree, Classification Rule

1. Introduction

Epilepsy is a physical condition in which a patient has recurrent seizures. According to a survey done by Professor Jing-Jane Tsai of National Chen Kung University, Taiwan, there is an average of 8.9 epilepsy patients per thousand people. Currently, seizure has been classified into three types: (1) partial seizures, (2) generalised seizures, and (3) unclassified seizures.

The main purpose of this research is to discuss the feasibility of applying chaos theory on the detection of seizure. We use the decision tree rules to find out the meanings of the data and create effective classification rules accordingly, and then to examine their accuracy and the credibility of the rules.

2. Literature Review

2.1. Epilepsy

The causes of epilepsy have to do with heredity, brain lesions, endocrine changes, or some other unknown reasons. Generally, epilepsy can be classified into two types by the clinical causes: (1) Idiopathic Epilepsy: Idiopathic epilepsy refers to epilepsy with uncertain causes. It is also called cryptogenic epilepsy. Patients with idiopathic epilepsy has syndromes like muscle stiffness, disturbance of consciousness, and spasm. (2) Symptomatic Epilepsy: Symptomatic epilepsy may be caused by brain injury or metabolic disorder. It may be caused by stroke, brain abnormality, encephalitis or brain degeneration disease. [8].

2.2. Classification of Seizures

There are three kinds of seizure for the epilepsy as below:

⁺ Jenn-Long Liu . Tel.: +886-7-657-7711#6551; fax: +886-7-657-8491.
E-mail address: jlliu@isu.edu.tw

- (1) Partial seizures: This affects just one part of the brain. Partial seizures are also referred to as ‘focal’ ones because the seizure occurs in one area only.
- (2) Generalized seizures: Generalized seizures are the result of simultaneous abnormal activity in the whole brain.
- (3) Unclassified seizures: Unclassified seizures cannot be classified due to insufficient information [8].

2.3. Chaos

Chaos, developed by the U.S meteorologist Edward Lorenz in 1963, explains how a random, dynamical system operates behind a complex phenomenon we perceive. The theory is to find the hidden interior regular rules and resolve the variable in the nonlinear systems. Its main features are below. [6-7, 9]

- (1) Nonlinearity: Chaotic systems must be nonlinear dynamic systems, but they are not necessarily chaotic. Chaotic systems only happen in some special nonlinear dynamic systems.
- (2) Sensitivity to initial conditions: This is also called the butterfly effect. It is the sensitive dependence on initial conditions; where a small change at one place in a nonlinear system can result in drastic difference in its later state.
- (3) Strange attractor: This is an important predictable factor in affecting the operation of a chaotic system. There are one or more hidden rules or principles that dominate the evolution of the system in some special fields.
- (4) Non periodic time path: As chaos theory is non-linear, no matter how much time passes, there will not be two exactly same states in chase system. If there are two same states in a system, it will make a regular cycle in time sequence. Such orbit will be predictable, then, it is not a chaos system.

2.4. Data Mining

Data mining is a crucial step in the Knowledge Discovery in Database (KDD) Process that consists of applying data analysis and discovery algorithms and can produce a particular enumeration of patterns (or rules) over the data. Common data mining technologies are (1) Associate Rule, (2) Classification, (3) Clustering Analysis [1-2, 4-5, 10].

3. Methods

This work analyzes some electroencephalography (EEG) data obtained from a German epilepsy research center by applying chaos theory to yield Delay Time (τ), Embedding Dimension (dm), Correlation Dimension (CD) Value, and Largest Lyapunov Exponent (LLE). Thereafter, we then check the values of CD and LLE to study whether EEG signals are chaotic or not, and then proceed the processes of generating the decision tree rules for analyzing the four variables and finding out their order

3.1. Chaos Theory and the Formula

In a defined chaotic condition, it’s important that there should be a time series. The signals of EEG can be seen as time series. We are to find out whether some regularity exists in the signals by applying Chaos Theory [3, 7]. To do so, we have to reconstruct a phase space. Since the signal varies with Delay Time, we use it and Minimum Embedding Dimension to set up the phase space. First, we have to determine the best Delay Time and Minimum Embedding Dimension. We then obtain the best Delay Time using Mutual Information Formula shown below as (1). The smaller the $I(X, Y)$ value, the less collaborative signal of two time sequences share.

$$I(X, Y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} P_{XY}(i, j) \log_2 \left\{ \frac{P_{XY}(i, j)}{P_X(i)P_Y(j)} \right\} \quad (1)$$

where $P_{XY}(i, j)$ represents a probability when $X_i = Y_j$, and $P_X(i)$ and $P_Y(j)$ are the probability of X_i and Y_j , respectively. Thereafter, we apply (2), (3) and (4), which were developed by Cao in 1997. These equations are used to evaluate the minimum space dimension containing all attracters. We then find an Embedding Dimension (dm) that makes E (d) value very close to 1, which makes Minimum Embedding dimension.

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i, d) \quad (2)$$

$$a_i(d) = \frac{\|Y_i(d+1) - Y_{n(i,d)}(d+1)\|}{\|Y_i(d) - Y_{n(i,d)}(d)\|} \quad (3)$$

where $i = 1, 2, \dots, N - m_0d$ and $Y_{n(i,d)}(d)$ is near $Y_i(d)$ while dimension equals d_0 . The location of $Y_{n(i,d)}(d+1)$ is near $Y_i(d+1)$ while dimension equals d_0 .

$$E1(d) = \frac{E(d+1)}{E(d)} \quad (4)$$

After we gain the best Delay Time (τ) and Minimum Embedding Dimension, we are able to set up the reconstructive phase space in (5).

$$X_i(t) = \{x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (dm - 1)\tau)\} \quad (5)$$

where $i = 1, \dots, N - m\tau$ and $X_i(t)$ is Reconstruct Data, and $X_i(t)$ is Original Data. To find out the characteristic of the attracters, we then calculate Correlation Dimension (CD) value which represents the complexity of a system. By calculating the distance between every two different points, the CD value is achieved. Below is the definitive formula of CD value as below.

$$v = \left[\frac{2}{N_c(N_c - 1)} \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} \log \left(\frac{|Y_i(d_M) - Y_j(d_M)|}{\varepsilon} \right) \right]^{-1} \quad (6)$$

The value of $|Y_i(d_M) - Y_j(d_M)|$ is calculate distance between two points, while $i \neq j$; ε is a radius. In a chaotic condition, CD value converges with the increasing dm value, as Fig.1 shown below. In a phase space, the distance change between two points from time 0 to time t is shown in Fig.2 and formulated (7). The symbol $d(x)$ represents the distance between the two points at time x. The formula of calculate distance variation is:

$$d(t) = d(0)e^{t\lambda_{\max}} \quad (7)$$

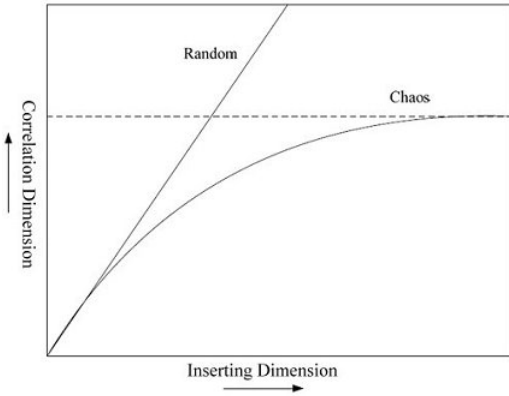


Fig. 1: Correlation curve of inserting dimension and correlation dimension

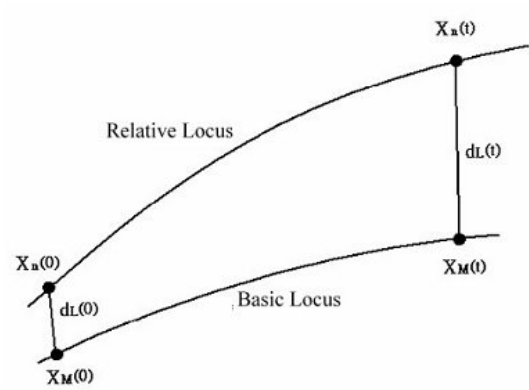


Fig.2 The variational locus of the distance between two points at time x

Table. 1: Characteristics of common movement types

System	Attractor	LLE
Stable	Stable Attractor	≤ 0
Regular	Regular Attractor	
Chaotic condition	Strange Attractor	> 0

By using (7), we can obtain the value of λ_{\max} , which is also called Largest Lyapunov Exponent (LLE). In this paper, the method developed by Wolf et al. [11] in 1985 is used to obtain the LLE. As shown in Table 1,

when LLE is larger than zero, the system is in a chaotic condition. On the contrary, when LLE equals or is smaller than zero, the system is in a stable and regular status.

3.2. Data Mining

Of the many data mining methods, we apply the decision tree rules in the clustering analysis using the four variables: τ , dm, CD Value, and LLE. In this work, we use the decision software indicated in See5 to conduct the clustering analysis. After that, we'll employ some tests to the data obtained from the transformation using chaos theory.

4. Results & Discussion

In this research, we divide the electroencephalography (EEG) data from German epilepsy research center into three groups: Group H (normal state), Group S (during seizures), and Group E (after seizures). The three groups of brainwave signal, O001 (from Group H), S001 (from Group S) and F001 (from Group E) are displayed in Fig. 3. The Fig. 3 shows clearly that the brainwave is in a state of much unrest during the seizure. After that, the brainwave becomes more stable than normal.

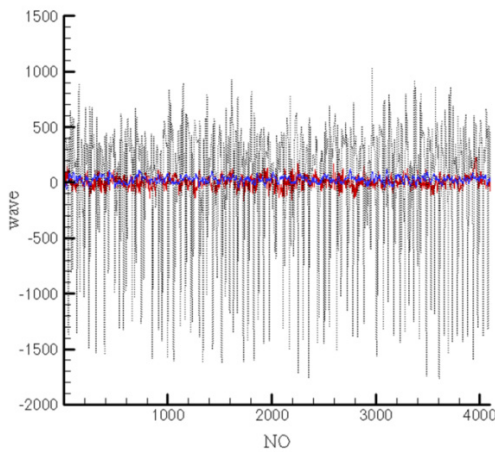


Fig.3: Brainwave signal chart in the normal state(Group H), during seizures(Group S), and after seizures(Group E)

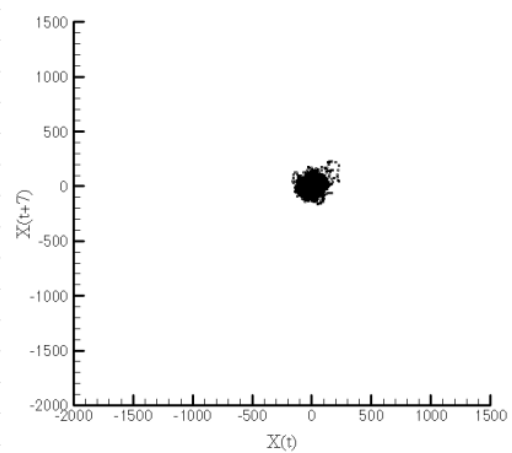


Fig. 4: Reconstructed chart in two-dimensional space for normal state (O-001) at $\tau=7$

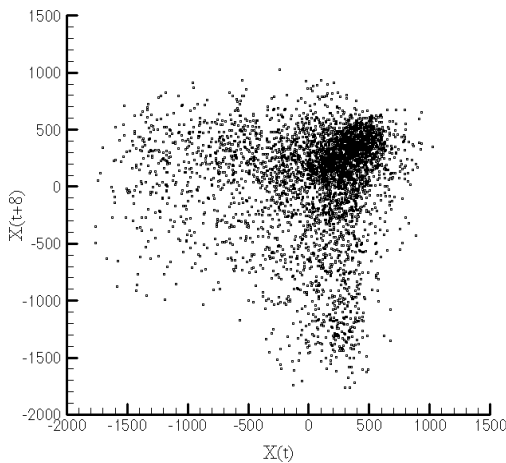


Fig. 5: Reconstructed chart in two-dimensional space for during seizures (S-001) at $\tau=8$

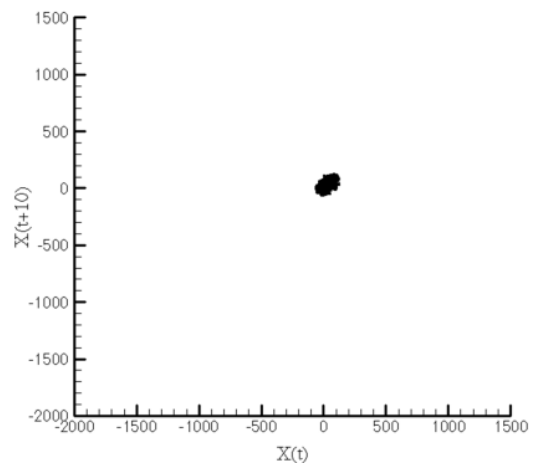


Fig. 6: Reconstructed chart in two-dimensional space for after seizures (F-001) at $\tau=10$

To determine whether the brainwave signals are in chaotic, we obtain the best Delay Time (τ) and Minimum Embedding Dimension (dm) to depict the reconstructive phase space graph. Figures 4, 5 and 6 show the reconstructive phase space graph of the signals based on the data of groups H (O001), E (F001), and S (S001), respectively. We can see that the points in the graphs demonstrate much regularity compared with their original disarray.

Figure 7 shows the results of clustering analysis for the 300 cases of signals depending on the four variables, τ , dm, CD value, and LLE, and the results were shown with decision tree rules. To evaluate the accuracy and credibility of using decision tree rules, the EEG with 300 cases are divided into 240 cases for training data sets and 60 cases for test data sets. A software on classification mining, termed See5, is used for the prediction of knowledge rules. The 60 cases are chosen randomly from these groups, 20 cases each from the three groups. Figure 8 displays the decision rules and accuracy of rules. As Fig. 9 shows the result of using See5 using the four variables (τ , dm, CD and LLE). The average accuracy of rules with 92.5% is achieved. The forecast accuracy of Groups H, S, and E are 70%, 65% and, 90%, respectively. The average forecast accuracy is 75% as shown in Table 2.

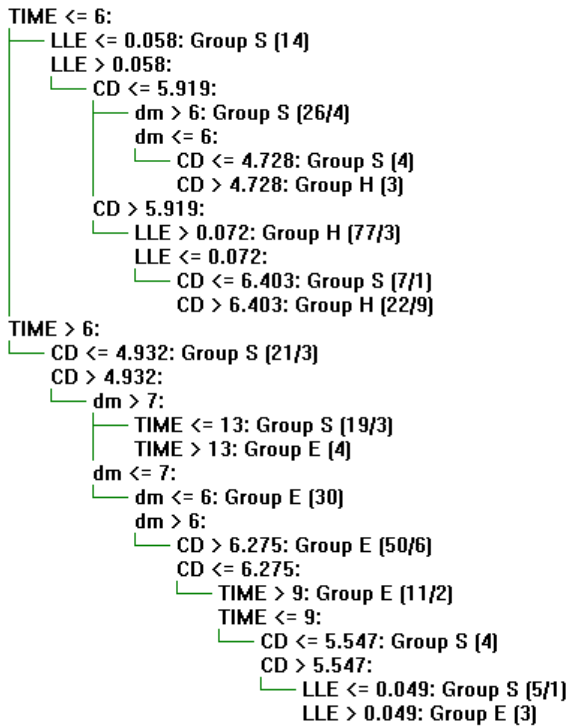


Fig. 7: The decision trees using See5 with 300 training cases.

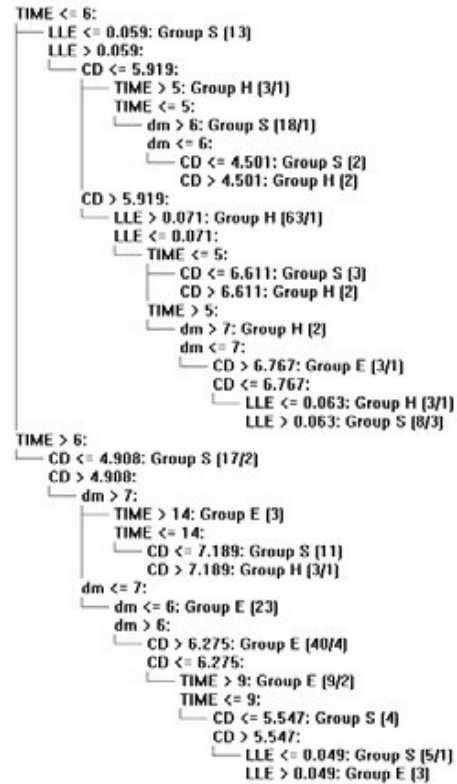


Fig. 8: The decision trees using See5 with 240 training cases and 60 testing cases

Evaluation on training data (240 cases):

Decision Tree			
Size	Errors		
22	18(7.5%)	<<	
(a)	(b)	(c)	<-classified as
74	3	3	(a): class Group H
2	74	4	(b): class Group S
2	4	74	(c): class Group E

Fig. 9: See5: Result accuracy of 240 training cases and 60 testing cases

Table. 2: The three forecast accuracy of 240 cases

Group	Accuracy of Rules	Forecast Accuracy	Average Accuracy of Rules	Average Forecast Accuracy
H	92.5%	70%	92.5%	75%
S	92.5%	65%		
E	92.5%	90%		

According to the predicted results of See5, the rules of decision tree showed an acceptable accuracy and reliability. We could use these rules to examine unknown groups by the four main variables (τ , dm, CD and LLE), and then could to examine which group it belongs to and determine whether they are seizures.

5. Conclusion

This work divided EEG signal data of epilepsy patients obtained from a German epilepsy research center into Groups H (normal state), S (during seizure), and E (after seizure), each of them containing 100 cases of signals, and then applied chaos theory to the data for achieving the parameters of Delay Time (τ), Embedding Dimension (dm), Correlation Dimension (CD) Value, and the Largest Lyapunov Exponent (LLE). Thereafter, this study used See5 to apply to the clustering analysis for getting classification rules from decision trees. Our results showed that the classification rules were generated with a high degree of accuracy and credibility. Accordingly, these classification rules can be applied to the interpretation of EEG signals to estimate the state of seizure or not of epilepsy patients based on the signal analysis.

6. Acknowledgements

This work was supported in part by grant NSC 100-2221-E-214-040 from the National Science Council of Republic of China.

7. References

- [1] R. Agrawal, T. Imielinski, A. Swami. Mining Association Rules between Sets of Items in Large Database, The 1993 ACM SIGMOD Conference. 1993, pp. 207-216.
- [2] R. Bellazzi. Predictive data mining in clinical medicine: Current Issues and Guidelines. *International Journal of Medical Informatics*. 2008, **77** (2): 81-97.
- [3] E. Conte, A. Federici, J. P. Zbilut. An Application of Chaos Theory for Estimation of Simultaneous Variability of RR-intervals in Heart and Systolic Blood Pressure in Humans, eprint arXiv:0810.4090, 2008.
- [4] U. Fayyad, P. S. Gregory, P. Smyth. From Data Mining Knowledge Discovery in Database, *AI Magazine*. 1996, **17**: 37-54.
- [5] J. Han, and M. Kamber. *Data Mining: Concepts and Techniques*, Morgan Kaufmann Pub. 2000.
- [6] N. K. Hayles. Chaos Bound: Orderly Disorder in Contemporary Literature and Science, *Cornell University Press*. 1990.
- [7] Y. Liu, Z. Hou, Y. Wu. Research on Forecasting the State of Driver Based on Chaos Theory. *Applied Mechanics and Materials*. 2011, **48-49**: 954-959.
- [8] B. Pohlmann-Eden, and M. Newton, First Seizure: EEG and Neuroimaging Following an Epileptic Seizure. *Epilepsia*. 2008, **49**: 19-25.
- [9] Y. Sun, V. Babovic, E. S. Chan. Multi-step-ahead Model Error Prediction Using Time-delay Neural Networks Combined with Chaos Theory. *Hydrology [J]*. 2010, **395** (1-2): 109-116.
- [10] P. N. Tan, M. Steninbach, V. Kumar. *Introduction to Data Mining*, Boston: Pearson Addison Wesley. 2006.
- [11] A. Wolf, J. B. Swift, H. L. Swinney, J. A. Vastano. Determining Lyapunov Exponents from a Time Series. *Physic*. 1985, **16D** (3): 285-317.