Analysis of Sleep Staging Based on Multivariate Symbolic Transfer Entropy

Meng GAO¹,a, Min WU²,b, Xiao LI³,c and Jun WANG¹,d,*

¹School of Biomedical Engineering, Nanjing Univ. of Posts & Telecomm, Nanjing, China
²School of Computer Science and Technology, Nanjing Univ. of Posts & Telecomm Nanjing, China
³School of Communication Engineering, Nanjing Univ. of Posts & Telecomm Nanjing, China

aB14090611@njupt.edu.cn, bB14090510@njupt.edu.cn, cB14090507@njupt.edu.cn,
dwangj@njupt.edu.cn

*Corresponding author

Keywords: symbolization, transfer entropy, multivariate, sleep staging

Abstract. Physiological electrical signals such as ECG, EEG and EMG during sleep contain a lot of physiological information. The analysis of these signals can provide effective advice for the diagnosis of sleep staging or sleep disorders. The symbol transfer entropy algorithm is applied to study the interaction information of physiological signals. However, because the traditional symbol transfer entropy algorithm only focuses on the relationship between one or two variables, this paper used the multivariate symbolic transfer entropy based on the traditional symbol transfer entropy to consider the coupling relationship between multiple variables. When dividing the time series, we used two methods respectively: static partitioning and dynamic adaptive. By comparing the multivariate symbolic transfer entropy values in the awake and sleep periods of different subjects, we can get that the multivariate symbol shift entropy of the awake period is significantly higher than that of the sleep subjects, although the entropy values are different from those of the subjects. And there is a significant difference between the two by using T test, which is consistent with the theory that the degree of brain disorder decreases and the entropy decreases when the sleep is deeper. The multivariate symbolic transfer entropy algorithm is effective in distinguishing human awake period from sleep period and can provide an effective way for other physiological signals research.

Introduction

Sleep is a very important physiological process, so the study of physiological signals during sleep can provide advice on sleep staging, sleep disorders [1] and so on. The analysis of sleep EEG [2] signal processing can extract the effective feature quantity, which can reflect the different functional states of the brain. In addition, the change of ECG [3] and EMG [4] signal in sleep time will also show some rules. In sleep research, R&K [5] principles based on EEG and other indicators such as ECG, EMG, EOG and so on, stipulates that the sleep is divided into awake period, non-rapid eye (NREM), rapid eye movement (REM). And NREM is divided into I, II, III, IV periods. The new AASM [6] principle, with the overall inheritance of the R & K division criteria, made a partial adjustment, in which the NREM III period was merged into the deep sleep (SWS).

At present, the research on physiological and electrical signals has developed rapidly. Methods commonly used are time-frequency domain analysis [7], neural network [8] and so on. Among them, the way of reflecting the physiological signal in the human body by calculating the information entropy such as symbolic transfer entropy algorithm STE [9] is quite effective in the degree of confusion. The traditional symbolic transfer entropy algorithm only quantifies one or two non-linear systems. However, the human body is a complex physiological process, which must include multiple nonlinear systems, so the STE algorithm has limitations. In this paper, a multivariate symbolic...
transfer entropy is proposed based on the traditional symbol transfer entropy to study the physiological signals of sleep staging.

In this paper, we used multivariate symbolic transfer entropy algorithm to further couple the quantification of sleep physiological signals. Then we can get the multivariate symbolic transfer entropy at different times. The experimental results show that the multivariate symbolic transfer entropy algorithm can better reflect the degree of chaos of physiological signals in different periods. The algorithm is potential to study sleep stage, sleep disease or other physiological signals. The traditional symbolic dynamics uses static division method. In this paper, we use the dynamic adaptive [10] method to divide the time series then use multivariate symbolic transfer entropy algorithm to obtain the consistent conclusion.

**Multivariate Symbolic Transfer Entropy Theory**

**Symbolic Transfer Entropy**

The transfer entropy [11] is used to describe the Interactive information between the two nonlinear systems, including the dynamic characteristics and directional information, it can reflect the degree of chaos and coupling between the two systems. However, due to the fact that its computational process is more sensitive to noise [2], the coordination between the parameters needs higher requirements. To improve that, symbolic transfer entropy STE [9] is proposed. The nature of symbolization is that the continuous signal is divided into several intervals according to some certain rules, so that many successive possible values are transformed into symbolic sequences with only a few different discrete values [18-19]. The symbol transfer entropy algorithm first implements the symbolization of the time series, and then reconstructs the phase space [9,12]. EEG, ECG, EMG and other electrical signals are continuous time series. Frequently we symbol these continuous time series

\[ X = \{ x_1, x_2, \ldots, x_N \} \]

whose length is N into the tractable symbolic sequence

\[ S = \{ s_1, s_2, \ldots, s_N \} \]

\[ s_j \in A, (A = 0,1,2,3) \]

During symbolic process, although some details are lost, the dynamic characteristics of the time series are preserved, and the computation speed is greatly improved. The symbolic methods used in this paper are as follows:

\[
s_j(x_i) =
\begin{cases}
0: & u_i < x_i \leq (1+a)u_i \quad \text{or} \quad (1-a)u_i \leq x_i < u_i \\
1: & (1+a)u_i < x_i < \infty \quad \text{or} \quad -\infty < x_i < (1+a)u_i \\
2: & (1-a)u_i < x_i \leq u_i \quad \text{or} \quad u_i \leq x_i < (1-a)u_i \\
3: & (1-a)u_i \leq x_i \leq (1-a)u_i
\end{cases}
\]

Here \( i = 1,2,\ldots,N \), \( u_i \) is the mean of \( x_i \) in \( X \) sequences which is greater than zero, \( u_2 \) is the mean of \( x_i \) in \( X \) sequences which is smaller than zero, \( a \) is constant parameters. During the process of symbolization, if the value of \( a \) is not correctly chosen, original time series will lose some detailed information, here we choose \( a = 0.05 \) [13].

**Multivariate symbolic transfer entropy:**

The traditional symbolic transfer entropy algorithm only focuses on one or two variables, ignoring the interaction of multiple variables. If a stationary multivariate discrete-time stochastic process \( X \) is given, its sub process can be expressed as \( X, Y, Z, \ldots \), then its multivariate transfer entropy [14] is defined as:

\[
I_{X \rightarrow Y}^{STE} = I_{X \rightarrow Y}^{DTE} = \sum_{i=1}^{N} I(X_{t+i} ; Y_{t} | S_{t}, X_{t})
\]

Based on the definition of the multivariate transfer entropy, we can deduce the definition of the multivariate symbolic transfer entropy, i.e. sequence \( X \) into symbolic sequence as follows:
Sequence $Y$ into symbolic sequence $J = \{ j_1, j_2, \ldots, j_n \}$, $j_i \in \{ A = 0, 1, 2, 3 \}$. Sequence $Z$ into symbolic sequence $K = \{ k_1, k_2, \ldots, k_n \}$, $k_i \in \{ A = 0, 1, 2, 3 \}$.

The definition of entropy of multi-variable symbol transfer is [15]:

$$I_{S \rightarrow J}^{TE} = I_{S \rightarrow K}^{TE} = \log \frac{p(s_t, j_{t-1}, j_{t-2}, \ldots, j_{t-n}, k_{t-1}, k_{t-2}, \ldots, k_{t-m})}{p(s_t, j_{t-1}, j_{t-2}, \ldots, j_{t-n}, k_{t-1}, k_{t-2}, \ldots, k_{t-m})} = \log \frac{p(s_t, j_{t-1}, j_{t-2}, \ldots, j_{t-n}, k_{t-1}, k_{t-2}, \ldots, k_{t-m})}{p(s_t, j_{t-1}, j_{t-2}, \ldots, j_{t-n}, k_{t-1}, k_{t-2}, \ldots, k_{t-m})}$$

### Dynamic Adaptive Formula

Compared with the traditional time series symbolization, the dynamic adaptive time series symbolization can capture the features contained in the time series more accurately, especially the dynamic characteristics.

The dynamic adaptive time series symbolization uses the adaptive template segmentation method, i.e. after the data collected on the whole partition, then solve the problem on the templates of each small interval. Dynamic adaptive partitioning method [10, 16] as follows: Set a length of $N$ points for the time series: $u = \{ u(i) : 1 \leq i \leq N \}$. For time sequence $u(i)$, we embed the $m$-dimensional phase space in time series to reconstruct the space:

$$X(i) = [u(i), u(i+L), \ldots, u(i+(m-1)L)]$$

Where $m$ is embedding dimension, and $L$ is time delay. When the time delay $L$ is selected as 1, the number of $m$-dimensional vector is $N-m+1$. For any $m$-dimensional vector, the BS (basic scale) [10] is calculated by the root mean square of the difference between the adjacent points of the $m$-dimensional vector:

$$BS(i) = \sqrt{\frac{1}{m-1} \sum_{j=1}^{m-1} (x(i+j) - x(i+j-1))^2}$$

Based on the basic scale, classification standards can be seen as $\alpha \times BS(i)$. Since the basic scale $BS(i)$ of each $m$-dimensional vector is different from the calculation result of vector value, the division criteria of the time series symbol interval are also dynamically changing. It can be adjusted adaptively according to the ever-changing data values in the sequence. We convert any $m$-dimensional vector $X(i)$ into a symbolic sequence:

$$S(X(i)) = \{ s(i), s(i+1), \ldots, s(i+m-1) \} \text{ } s \in A \text{ } A \in \{ 0, 1, 2, 3 \}$$

Symbolic conversion process is:

$$S_{j(i)} = \begin{cases} 0: & \bar{x} < x_{i+k} \leq \bar{x} + \alpha \times BS(i) \\ 1: & x_{i+k} > \bar{x} + \alpha \times BS(i) \\ 2: & \bar{x} - \alpha \times BS(i) < x_{i+k} \leq \bar{x} \\ 3: & x_{i+k} \leq \bar{x} - \alpha \times BS(i) \end{cases}$$

Among them, $i = 1, 2, 3, \ldots, N-m+1, k = 0, 1, 2, \ldots, m-1$. $\bar{x}$ is the mean of the $m$-dimensional vector $X(i)$, and $BS(i)$ is the basic dimension of the i-th m-dimensional vector. The symbol 0,1,2,3 is used to
mark all the regions, so it is meaningless to divide this value. $\alpha$ is a special constant parameter, so it is very important to choose the right $\alpha$ for the interval and the division of the symbol. If the value of $\alpha$ is too large, the original time series will be converted into a symbol sequence, and the detailed information will be lost. If the value of $\alpha$ is too small, the time series will be affected by noise. In this paper, we use the method used in the Wessel correlation test analysis to select the value of $\alpha$ [17].

Schematic diagram shown in Fig. 1:

![Schematic Diagram](image)

**Figure 1. Schematic representation of sequence.**

**Multivariate Symbolic Transfer Entropy is Applied to Sleep Signals**

The paper uses sleep data from the MIT-BIH Polysomnographic Database from PhysioBank data. The signal data whose sampling frequency is 250Hz, and the data contain first lead ECG signal, first lead EEG signal, first lead EMG signal and other multi-parameter signal. The data used herein are ECG signals, EEG signals, and EMG signals measured by subjects during both awake and sleep phases.

We extracted the ECG signal, EEG signal and EMG signal from three subjects slp41, 45, 481 when they are at the awake period and the NREM-I. Because the sampling frequency of test data is 250Hz, every 30s contains 7500 test points. There is too much data, so we use the method that every 80 points takes a sampling point to reduce the calculation, and extract the source data for the test points.

Since the length of the selected time series has some influence on the size of the entropy, this paper finds the optimal length by changing the length of the time series. And find the optimal time series length when the difference of entropy between the awake period and the NREM-I is largest. The length of the original time series of EEG, ECG and EMG during sleep and wakefulness were taken as $L=40,40\times2,40\times3,40\times4,40\times5$, respectively, we get that $L=40\times4$ is the best sequence length after calculation.

The average of the multivariate symbolic transfer entropy calculated in each group of the subjects is taken as the final multivariate symbolic transfer entropy result of the individual. The specific results of entropy of different subjects in two periods are shown in Table 1:

| Subjects | Awake period | NREM-I period |
|----------|--------------|---------------|
| slp41    | 0.4183       | -0.0512       |
| slp451   | 0.3234       | 0.0214        |
| slp481   | 0.3129       | 0.0167        |
| mean     | 0.3515       | -0.0044       |
| std. deviation | 0.0581   | 0.0406        |

As we can see the multivariate symbolic transfer entropy of the human body in awake period from the above table is obviously larger than that of NREM-I, and the entropy of the awake period falls
within \([0.3515 \pm 0.0581]\), while the NREM-I entropy is in the range \([-0.0044 \pm 0.0406]\), as the Fig. 2a shows. The distribution of entropy in the two periods is obviously different. After using T test with SPSS, we can get \(P = 0.025\). It reflects that there is a significant difference in the multivariate symbolic transfer entropy of the two periods.

When the optimal length sequence is \(L=40\times 4\), we use the previous section of the multivariate symbol entropy calculation method after the dynamic adaptive time series division, and calculate the results as follows:

Table 2. Multivariate symbolic transfer entropy in two periods when using improved algorithm.

| Subjects   | Awake period | NREM-I period |
|------------|--------------|---------------|
| slp41      | 11.201       | 6.7137        |
| slp451     | 9.8182       | 6.2917        |
| slp481     | 12.301       | 6.2917        |
| mean       | 11.1067      | 6.4324        |
| std. deviation | 1.2441     | 0.2436        |

We find that the entropy of the awake period falls within \([11.1067 \pm 1.2441]\), and the NREM-I entropy is in the range \([6.4324 \pm 0.2436]\), as the Fig. 2b shows. The entropy of the sleep period is obviously less than that of the awake period, and there is an obvious Interval distribution. When using dynamic adaptive method, we can get \(P = 0.023\) by using T test with SPSS. Similarly there are also significant differences.

Figure 2. The ranges of multivariate STE with static division about two stages (mean±std).

This is consistent with the conclusion we get of the traditional time series division. The multivariate symbolic transfer entropy algorithm under two methods both can distinguish between sleep stage and awake period. Multivariate transfer entropy value reflects the degree of chaos of time series, which can be explained that the body with the degree of sleep deepened, the amount of information received by the brain will reduce. The uncertainty of the information and physical activity are decreased, which is consistent with the actual situation.

Summary

In this paper, we study the theory of nonlinear dynamics in the field of physiological signal analysis, and proposes a multivariate symbolic transfer entropy algorithm based on the traditional symbol transfer entropy algorithm. Without limiting to one or two nonlinear systems, multivariate coupling better reflects the complexity of the human body. At the same time, when the time series is symbolized, the traditional method and the dynamic adaptive method are used respectively. The
multivariate symbol transfer algorithm under the two methods can distinguish between the awake period and the NREM-I.

In the future study of sleep, the MSTE algorithm can be further extended to other stages of sleep staging, such as NREM-II, NREM-III and so on. Experiments show that the multivariate symbolic transfer entropy algorithm is expected to provide effective advice on the study of human sleep disorders, sleep staging or other physiological aspects of the human body.

Acknowledgement

Project supported by the project is supported by the National Natural Science Foundation of China (Grant Nos. 61271082, 31671006), Jiangsu Provincial Key R & D Program (Social Development) (Grant No.BE2015700, BE2016773), the Natural Science Foundation of Jiangsu Province (Grant No. BK20141432), Natural Science Research Major Program in Universities of Jiangsu Province (Grant No.16KJA310002), Environmental Protection Research Project of Jiangsu Province (Grant No.2015045), and Students’ Innovation and Entrepreneurship Training Program(Grant No.SZDG2016029).

References

[1] Lane J, Liang J, Vlasac I, et al. Genome-wide association of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric metabolic traits [J]. 2017, 49(2): 274-281.
[2] Musselman M, Djurdjanovic D. Time-frequency distributions in the classification of epilepsy from EEG signals [J]. Expert Systems with Applications, 2012, 39(39): 11413–11422.
[3] Fenske S, Auer F, Hassan S, et al. Comprehansive multilevel in vivo and in vitro analysis of heart rate fluctuations in mice by ECG telemetry and electrophysiology [J]. Nature Protocols, 2016, 11(1): 61-86.
[4] Deluca C. Physiology and mathematics of myoelectric signals [J]. IEEE Transactions on Biomedicine Engineering, 1979, 26: 313-325.
[5] Rechtschaffen A, Kales A, eds. A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects. Los Angeles: Brain Information Service/Brain Research Institute, University of California, 1968.
[6] Iber C, Ancoli-Israel S, Chesson A, Quan SF, eds. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification, 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.
[7] Petrantonakis P C, Hadjileontiadis L J 2012 IEEE T. on Signal Proc. 60 2604
[8] Musselman M, Djurdjanovic D 2012 Expert Syst. Appl. 39 11413.
[9] Matthaus S, Klaus L. Symbolic transfer entropy [J]. Phys. Rev. Lett. 2008, 100(15): 158101.
[10] Li J, Ning X B. Dynamical complexity detection in short-term physiological series using base-scale entropy. Physical Review E, 2006, 73: 052902-1-052902-4.
[11] Schreiber T. Measuring information transfer [J]. Phys. Rev. Lett, 2000, 85(2): 461-464.
[12] Jing X R, Hu Y T, Wang J. Symbolic transfer entropy analysis of sleep stage classification [J]. Beijing Biomedical Engineering, 2012, 31(4): 373-376.
[13] Wessel N, Ziehmann C, Kurths J, et al. Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates [J]. Phys. Rev. E, 2000, 61(1): 733-739.
[14] Jakob R, Jobst H, Vladimir P, et al. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy [J]. Phys. Rev. Lett., 2012, 108(25): 258701.

[15] Liu Q Q, Dai J F, Li J, Wang J, Hou F Z. Epileptic EEG Based on Improved Multivariate Symbolic Transfer Entropy [J]. Journal of Data Acquisition and Processing 2016, 31(5):983-988.

[16] Wang J, Ma Q L, Ning X B. The base scale entropy analysis of electrocardiogram [J]. Journal of Nanjing University (Natural Sciences), 2008, 44(4):26-29.

[17] Wessel N, Ziehmann C, Kurths J, et al. Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates [J]. Physical Review E, 2000, 61: 733-739.

[18] Daw C, Finney C, Tracy E, A review of symbolic analysis of experimental data [J]. Review of Scientific Instruments, 2003, 74(2): 916-930.

[19] Brida J G, Punzo L F, Symbolic time series analysis and dynamic regimes [J]. Structural Change and Economic Dynamics, 2003, 14(2): 159-183.