ATLANTIS PRESS
Advances in Engineering Research, volume 163
7th International Conference on Energy, Environment and Sustainable Development (ICEESD 2018)

Anomaly detection for sleep EEG signal with Mahalanobis-Taguchi-Gram-Schmidt method

Xiaohong Peng[1],Rui Zheng[2],Jiufu Liu[2],*

Xiaobin Din[2]

1. Department of Information Engineering , Jiangsu Maritime Institute, Nanjing 211170, China

2. College of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Keyword: Anomaly detection, Electroencephalogram(EEG) signal,Mahalanobis-Taguchi-Gram-Schmidt, sleep stages, Signal to Noise ratio

Abstract：Considering the tedious steps, the poor accuracy and over-subjectivity of human sleep quality judgment artificially, this paper presents an automatic detection algorithm of sleep quality based on Mahalanobis-Taguchi system method. Based on the modeling and analysis of the human brain dual channel EEG signals, the normalized vector of each channel is obtained under different sleep stages. At the same time, the linear independent vector group is subjected to Gram-Schmidt orthogonalization, and the mean value of the signal-to-noise ratio of each sleep stage is calculated by using the Mahalanobis-Taguchi-Gram-Schmidt method. Through analyzing the waveform of the mean signal-to-noise ratio on different sleep stages, the normal and the abnormal sleep quality can be identified.

Introduction

Effectively distinguishing the level of sleep quality is very beneficial for the treatment of sleep apnea, insomnia and narcolepsy. Nowadays, people's sleep quality testing basically rely on the subjective feelings of the people themselves. Besides, medical experts usually determined the patient's sleep only through the patient's oral symptoms and other complications. This method is lack of objectivity, as well as the process is too cumbersome and inaccurate.

Polysomnography is a technique used in the diagnosis and treatment of sleep disorders. Electroencephalogram(EEG) signal is an electrical signal representing the activity of cerebral neurons. It can be used for noninvasive measurement. Therefore, EEG signals have been used to study brain activity of different sleep stages[1]. The traditional classification of sleep stages is the classification criteria developed by experts according to Rechtschaffen and Kales (R & K)[2]. This study used the six-state sleep stages of the R & K standard: Awake (Awa), Stage 1 (S1), Stage 2 (S2), Stage 3 (S3), Stage 4 (S4) and REM[3]. The five-state staging phase combines S3 and S4 into one state. The four-state staging phase combines S1 and S2 on a five-state basis, and sometimes stages S1, S2, S3 and S4 are represented as NREM[4]. The three-state staging phase includes Awa, NREM and REM[4],[5],[6]. In this paper, EEG data of six-state sleep stages, which has been marked by artificial sleep staging experts, are analyzed.

Mahalanobis-Taguchi method through building the Mahalanobis reference space, you can effectively distinguish between normal samples and abnormal samples. Through orthogonal table and signal to noise ratio to optimize each variable, we can select the optimal variables to better the classification and to predict of the study[7]. Based on the data of EEG signals classified by artificial experts on sleep stages, this study detects the normal and abnormal sleep quality of human beings using the Mahalanobis-Taguchi-Gram-Schmidt process(MTGS) method. This method is different from calculating the inverse matrix of the correlation matrix, but calculates the Mahalanobis distance(MD) through the Gram-Schmidt orthogonalization process(GSP), which can effectively deal with the advantages of the multicollinearity problem without calculating the signal-to-noise ratio by the orthogonal table and reduce the computational complexity of the algorithm. In this paper, the feasibility and superiority of this method in the detection of human sleep quality are verified through experiments.

Mahalanobis-Taguchi-Gram-Schmidt method

Mahalanobis-Taguchi system (MTS)

MTGS method has more advantages than MTS. However, the MTGS method also has some disadvantages in its original form. These shortcomings can be solved by calculating the Mahalanobis distance using the Gram-Schmidt Orthogonalization process and using an orthogonal array to evaluate the signal to noise ratio(S/N)[7]. To highlight the differences between these approaches, we will first describe the MTS and then discuss the MTGS and its variants for improvement.

The Mahalanobis-distance calculation in MTS can be found in the following formula[7]:

错误!未找到引用源。 (1)

错误!未找到引用源。 (2)

Where 错误!未找到引用源。 denotes the total number of variables, n denotes the total number of samples, 错误!未找到引用源。denotes the number of variables 错误!未找到引用源。, 错误!未找到引用源。denotes the number of samples 错误!未找到引用源。.错误!未找到引用源。 is the standardized vector of normalized variables 错误!未找到引用源。. 错误!未找到引用源。 denotes the ith variable value of the jth sample, 错误!未找到引用源。 denotes the average of the ith variable in the healthy group, 错误!未找到引用源。 denotes the standard deviation of the ith variable in the healthy group, 错误!未找到引用源。 Is the transpose of the vector, 错误!未找到引用源。 is the association matrix of the health group.

The normal group's correlation matrix is a square symmetric matrix of 错误!未找到引用源。 orders, with all diagonal terms being one[8]. 错误!未找到引用源。represent the correlation value between the 错误!未找到引用源。th and qth variables, and 错误!未找到引用源。 represent the correlation values between the qth and pth variables. Since the correlation between the pth and qth variables is the same as the correlation between the qth and pth variables, the correlation matrix is symmetric. All diagonal terms of the normal group's correlation matrix are consistent.

After calculating the Mahalanobis distance, an orthogonal array was used to calculate the S/N ratio for each experiment. In an orthogonal array, two levels of variables are considered, indicating their presence or absence. Based on the variables used to construct the MD, the type of signal to noise ratio is determined. For manufacturing tests with unknown true level anomaly, the bigger the S/N ratio, the better[8]. The signal to noise ratio can be calculated in the following way:

错误!未找到引用源。 (3)

Here 错误!未找到引用源。 represents the number of observations.

In Eq. (1), the correlation of variables and the distribution of variables are taken into account when calculating MD, which is the main difference between the calculation of MD and the calculation of Euclidean distance. Euclidean distance is the normal distance measure in algebra. For the case of no correlation between these variables, the MD is the same as the Euclidean distance. From Eq. (1), for cases where the correlation between the variables under consideration is very high, the correlation matrix becomes singular and the inverse of the correlation matrix 错误!未找到引用源。 is incorrect. This phenomenon of strong correlation between variables is called multicollinearity and may not be accurate according to Eq. (1) and Eq. (2). In addition, good outliers may also be observed in some cases. The MTS does not recognize the direction of the anomaly, and the MTGS method is preferred for the multicollinearity and recognition of the anomalous direction.

Mahalanobis-Taguchi-Gram-Schmidt method

MTGS is an improvement on MTS. MD in MTGS method can be calculated by Gram-Schmidt Orthogonalization Process.

Gram-Schmidt orthogonalization converts linear independent vectors into orthogonal vectors. The normalized vector of the normal variable 错误!未找到引用源。 obtained by the Eq. (2) is regarded as a linear independent vector to calculate the orthogonal set of vectors. The equations for the Gram-Schmidt orthogonalization process are as follows[9]:

错误!未找到引用源。 (4)

错误!未找到引用源。 (5)

错误!未找到引用源。 (6)

Where 错误!未找到引用源。is kth group of the standard vector according to the Eq. (2), and 错误!未找到引用源。 is kth group of perpendicular to each other with the same linear span.

The Mahalanobis distance corresponding to the jth observation is given by[10]:

错误!未找到引用源。 (7)

Where 错误!未找到引用源。 is the element of the orthogonal vector 错误!未找到引用源。, 错误!未找到引用源。 is the standard deviation of 错误!未找到引用源。, and 错误!未找到引用源。 is the total number of variables.

In the MTGS method, the S/N ratio of the variables can be directly calculated. The larger the better the S/N ratio for the ith variable, can be given as follows[10]:

错误!未找到引用源。 (8)

Where t indicates the number of observations. However, S/N ratio is calculated according to the Eq. (8). It is effective only if the partial correlation between variables under consideration is not significant. In addition, the results depend on the order in which the variables are taken into account by using the Eq. (8) in the MTGS process. Therefore, calculating the MD by the MTGS method and adopting the orthogonal array and evaluating the S/N ratio according to the Eq. (3) eliminate all the shortcomings of MTS and MTGS. Here the improved MTGS method is used.

In order to determine the direction of the anomaly, we first classify all the variables as the larger the better type and the smaller the better type category[10-12]. If the jth is abnormal and MD is above the user-defined threshold, good anomaly can be identified based on the following rules.

错误!未找到引用源。, (If 错误!未找到引用源。 is larger the better)

错误!未找到引用源。, (If 错误!未找到引用源。 is smaller the better)

Calculating the amount of calculation of MD by GSP and using an orthogonal array (modified MTGS) is not a problem as the whole process takes less time.

Sleep analysis and experimental data processing

Experimental data

The experimental data used in this study comes from the Sleep-EDF database, which is part of the Physionet database[12]. This article uses two sets of data records from sc4001e0 and st7022j0. The first set of data is the sleep data of migrant healthy volunteers recorded in 1989 and the second set of data is the sleep data of subjects with mild sleep disturbances recorded in 1994. The data from each subject is saved in the EDF file and each record includes one horizontal EOG and two EEG channels (Fpz-Cz and Pz-Oz)[15]. The three channel signals are sampled at 100 Hz. In this study, EEG signals from both Pz-Oz and Fpz-Cz channels are selected to analyze and identify the sleep stage. Since experts have generated sleep-phase sequence diagrams based on R&E recommendations every 30 seconds of EEG data, the interval for each phase in this study is defined as 30 s and contained 3,000 data points.

The original sleep stages for these segments are labeled with six categories: S1, S2, S3, S4, REM and Awa. Only the records sc4001e0 and st7022j0 have MVT data in the original EDF file. This study only involved AWA, S1 to S4 and REM sleep stages. Each sleep phase of each channel is sampled separately. Two channels of six stages are used to respectively collect 20 groups of EEG signals. The EEG signal diagram is shown in Fig.1.

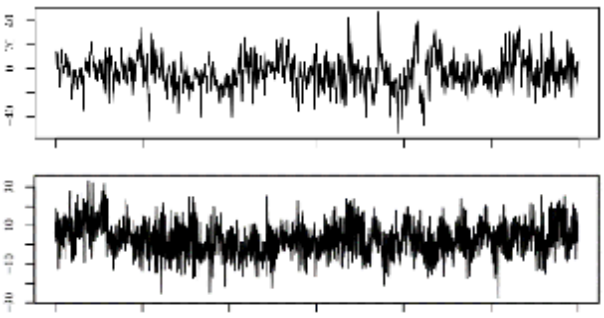


Fig.1 EEG signal

Data processing

This study used the aforementioned GSP method to process the sleep data. Fig.1 shows the whole process of data processing method. This method is used for both normal group and abnormal group.

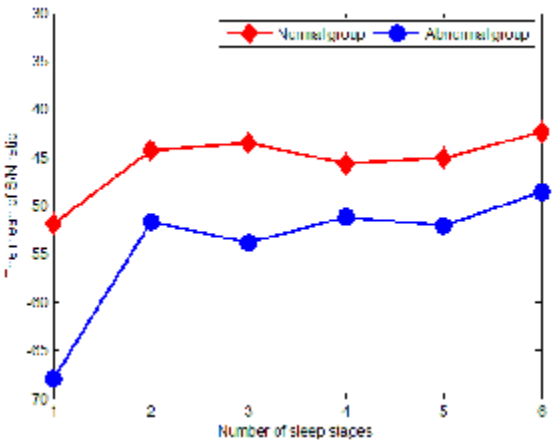
Tab.1The mean 错误!未找到引用源。 and the standard deviation 错误!未找到引用源。 of normal groups

Stages	Standard deviation 错误!未找到引用源。		Mean 错误!未找到引用源。	
	Fpz-Cz	Pz-Oz	Fpz-Cz	Pz-Oz
Awa	32.086	21.393	2.589	-0.881
1	13.862	6.2789	-1.243	1.611
2	16.274	8.2748	-2.465	-2.577
3	27.914	17.531	5.801	-5.152
4	23.470	18.986	-14.634	4.977
R	9.828	7.700	-4.304	-1.013

As shown in Tab.1，using the sample data of six-stages normal group, the mean 错误!未找到

引用源。 (i = 1,2) and standard deviation 错误!未找到引用源。 (i = 1,2) of each channel are calculated. The original vector is then normalized according to Eq.(2) to get a linear independent vector. Next, we need to orthogonalize the linear independent vectors. Lastly, we can use Eq. (8) to calculate the signal to noise ratio for each sample.

Fig.2 The S/N ratio mean waveform in healthy and abnormal groups



Results and discussion

The average S/N ratios of six sleep stages in normal group and abnormal group are calculated respectively We draw the waveform of the mean value distribution of signal-to-noise ratio of normal group and abnormal group, as shown in Fig.2.

1. As a whole, from Awa stage to S4 stage, and then to REM stage, the average signal-to-noise ratio of the normal group in the sleep period is above that of the abnormal. From this, it can be clearly distinguished the normal sleep group and the abnormal group.

2. In Fig.2, compared the average signal-to-noise ratio of normal group on Awa stage with those of abnormal group, the difference is very large. Compared with those with normal group, abnormal group have more volatility in the mean signal to noise ratios on six sleep stages. In addition, the curves of the two signal-to-noise mean values all show an upward trend, in which the abnormal group increases significantly faster than normal group.

3. As we can see from the figure above, on Awa stage, normal and abnormal sleep normalized signal-to-noise ratio are -51.84 and -67.94, the difference between the two reached 16.1, a wide gap. On S1 stage, the average signal-to-noise ratio of normal group is -44.22, and the average signal-to-noise ratio of abnormal group is -51.62, with 7.4 lower than normal subjects. On S2 stage, the average signal-to-noise ratio of the normal is -43.47, and the average signal-to-noise ratio of the abnormal is -53.84. The abnormal is 10.37 lower than the normal. Compared with S1 stage, the gap has been widened futher. The average signal-to-noise ratio of the normal is -45.64 on the next stage, and the mean signal-to-noise ratio of subjects with the abnormal is -51.13, with 5.49 lower than the normal value. The gap is narrowed, but the gap is still greater than the gap of S1 stage. On S4 stage, the average signal-to-noise ratio of the normal is -45.05, the average signal-to-noise ratio of the abnormal is -52.05, with 7.0 lower than the normal, a slight increase compared with the previous stage. On REM stage, the average signal-to-noise ratios of the normal and abnormal are -42.32 and -48.53, respectively. The difference between the normal and the abnormal reached 6.21, a slight decrease from the previous stage.

Conclusion

EEG signal is processed by the MTGS algorithm. The mean of signal-to-noise on six sleep stages are found. The signal-to-noise ratios of all the persons with normal sleep status are above the ones with abnormal sleep status. The use of MTGS method can effectively distinguish between normal sleep group and abnormal sleep group. The algorithm is applied to the determination of human sleep quality and provides a new idea for artificial intelligence to automatically detect sleep quality instead of the traditional manual discrimination.

References

[1] S. A. Imtiaz and E. Rodriguez-Villegas, Automatic sleep staging using state machine-controlled decision trees, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 378–381, 2015.

[2] A. R. Hassan and M. I. H. Bhuiyan, Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating, Biomedical Signal Processing and Control, 24, pp. 1–10, 2016.

[3] B. Boashash, S. Ouelha, Automatic signal abnormality detection using time-fre quency features and machine learning: a newborn EEG seizure case study, Knowledge-Based Systems，106, pp. 38–50, 2016.

[4] M. Diykh, Y. Li, P. Wen, EEG Sleep Stages Classification Based on Time Domain Features and Structural Graph Similarity, IEEE Transactions on Neural Systems and Rehabilitation Engineering，24, pp. 1159–1168, 2016.

[5] M. Adnane, Z. Jiang, and Z. Yan, Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram, Expert Systems with Applications, 39, pp. 1401–1413, 2012.

[6]A. R. Hassan and M. I. H. Bhuiyan, A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features, Journal of Neuroscience Methods, 271, pp. 107–118, 2016.

[7] P. Shakya, A.K. Darpe, M.S. Kulkarni, Use of Mahalanobis Taguchi system as data fus ion approach for monitoring health of rolling element bearing, in:P. Sas, D. Moens, S. J onckheere (Eds.), ISMA International Conference on Noise and Vibration Engineering, L euven, pp. 3389–3402, 2012,

[8] D. Mohan, C. Saygin, J. Sarangapani, Real-time detection of grip length deviation during pull-type fastening: a Mahalanobis–Taguchi System (MTS)-based approach, International Journal of Advanced Manufacturing Technology, 39, pp. 995–1008, 2008.

[9] E.A. Cudney, D. Drain, K. Paryani, S. Naresh, A comparison of the Mahalanobis– Taguchi system to a standard statistical method for defect detection, International Journal of Advanced Manufacturing Technology, 2, pp. 250–258, 2009.

[10] Chand RR, Kim IS, Lee JH, Kim JS Prediction welding quality in multi-pass welding processsing Mahalanobis distance method. Applied Mechanics and Materials pp. 2688–2693, 2013

[11] S. B. Kim, K. L. Tsui, T. Sukchotrat, and V. C. P. Chen,Acomparison study and discussion of the Mahalanobis-Taguchi system, International Journal of Industrial and Systems Engineering, 4, no. 6, pp. 631–644, Jan. 2009.

[12] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Computers in Biology and Medicine, 41, pp. 380–389, 2011.