An Algorithm for Extracting Entropy Features from EEG Signals Based on T-test and KPCA and Its Application on Driving Fatigue State Recognition

Shuli Zou, Peifan Huang, Pengpeng Shangguan, Zhiqiang Lin, Beige Ye and Taorong Qiu*

Department of Computer, Nanchang University, Nanchang 330029, China

*Corresponding author email: qiutaorong@ncu.edu.cn

Abstract. In consideration of the nonlinear characteristics of electroencephalography (EEG) signals collected in the research on driving fatigue state recognition, the recognition accuracy and the time performance of the driving fatigue state recognition method based on EEG is still not ideal, we construct a driving fatigue state recognition model and corresponding recognition method by combining t-test with kernel principal component analysis based on EEG entropy features. By applying this method to 30-electrode EEG data, testing it with 7 kinds of classifiers and comparing the results with the results without t-test, we find that the proposed method not only improve time performance, but also has the ideal accuracy. Through selecting the best classifier, the recognition accuracy and time performance are improved.

1. Introduction

Driving fatigue refers to the physiological and psychological dysfunction caused by the driver’s continuous driving for a long time, resulting in blurred vision, unresponsiveness, dull movement and decreased driving performance [1]. Although China’s car parc accounts for only 2-3% of global car parc, the number of people involved in traffic accidents accounts for about 20% of the global total. Among these traffic accidents, driving fatigue accounts for a considerable proportion. Therefore, it is of great significance and application value to study the scientific prevention of driving fatigue. A series of research on driving fatigue state recognition have been carried out currently both here and abroad. Khalaf et al. [2] used multi-scale analysis and common spatial pattern algorithm to extract EEG and fTCD features, and they proposed probabilistic fusion of EEG and fTCD evidences. Wang et al. [3] proposed a non-contact portable driving fatigue detection technology based on the physiological signals of driver, in which the no-fatigue and fatigue states of driver can be clearly distinguished by obtaining the driver’s myoelectric signal and ECG signal through the sensor. Chen et al. [4] put forward a fatigue detection method on the basis of brain network characteristics, which have important sense to the development of driver fatigue detection system. Huang et al. [5] designed a driving fatigue detection algorithm on the basis of face multi-features, which realized the extraction and fusion of mouth yawning, blinking frequency and eye closure under different illumination conditions. Chai and Naik et al. [6] put forward a two-class EEG-based classification to classify the driver fatigue state, which can be effectively used as a confrontation device for driver fatigue recognition and other adverse event applications.

The driving fatigue recognition algorithm on the basis of sample entropy and kernel principal component analysis (KPCA) proposed in previous research improved the recognition accuracy but increased the recognition time. Therefore, we explore the issue of time performance. In this paper, the
t-test is added based on the previous research, and we proposed an algorithm of entropy and t-test combined with KPCA. Two sets of data with significant differences on each electrode were obtained by t-test to achieve the effect of EEG data reduction, result in the improvement of time performance. In order to choose the most suitable classifier, we also test our algorithm under 7 kinds of classifiers, thereby to achieve the better experimental results. The results demonstrated that this method not only increases the recognition accuracy of driving fatigue state recognition under the appropriate classifier, but also improves the time performance of the experiment.

2. Materials and Methods

2.1. EEG Data and Preprocessing
The platform environment consists of a static simulator which comprises three 24-inch monitors and a software teaching system for driving simulations. A EEG collecting cap with 32-electrode was used. 25 normal subjects were tested for the present fatigue degree during the training process, such as the quality of sleep the night before, the eating habits during the day, etc., and then recorded 2 sets of experiment data for each subject, namely fatigue state and non-fatigue state. Every subject was demanded to drive for 40 mins without rest, and then asked them to conduct a questionnaire to check current states [7]. The EEG data is a 600s time series with 32-electrode and a sampling rate of 1000 Hz, which includes 300s of non-fatigue state and 300s of fatigue state. The EEG data is filtered and processed after collected.

2.2. Entropy Feature
Pincus et al. [8] proposed the concept of approximate entropy (AE) [9] in 1991. It reflects the probability of generating the incidence of new information in the time series. Sample entropy (SE) is a measurement method of time series complexity which first proposed by Richman et al. [10]. Fuzzy entropy (FE) was first put forward by Chen et al. [11]. Spectral entropy (SPE) uses the power spectrum of the signal to evaluate the regularity of the time series, the amplitude component is used for the probability of entropy calculation. Combined Entropy (CE) [7] was proposed by Mu et al. that four kinds of entropy (FE, SE, AE and SPE) were used to extract features. Wavelet Entropy (WE) [12] uses the norm sequence of wavelet coefficient vectors to measure the proximity of signals at various scales.

2.3. Significant Difference
Significant difference is a kind of quantitative evaluation of probabilistic. If there is a significant difference between two sets of data, it can be inferred that the two sets of data are from two distinct populations with differences. The test algorithm used in this paper is the test2 test of the t test. The value of p will be p<=0.05 in the subsequent experiments.

2.4. Kernel Principal Component Analysis
Kernel principal component analysis (KPCA) [13,14] transforms input space to feature space through nonlinear mapping $\Phi$, and then it performs a linear PCA on the mapped data, so it has a strong nonlinear processing capability. Radial basis function (RBF) will be the kernel function in the following experiments.

2.5. Driving Fatigue State Recognition Algorithm Based on Entropy and T-test Combined with KPCA
Given $n$ electrodes, suppose the total sampling time (in seconds) of each electrode are $m$, denoted the entropy feature corresponding to the $i$ th second of the $j$ th electrode as $x_{ij}$, the driving fatigue state recognition detection algorithm on the basis of entropy combined with PCA (ENTROPY_PCA) is described as follows:
1. Set the parameters which related to entropy feature calculation.
2. Calculate entropy feature based on entropy feature calculation formula, Construct an entropy feature matrix for all EEG data as $X_{mea}$. 

2
3. Perform t test on the entropy feature matrix of EEG data based on t-test algorithm, calculate the difference between the fatigue data and the resting data on each electrode by the ttest2 algorithm in the t-test. Then using feature vectors which p<= 0.05 to construct a new EEG feature matrix $Z_{mxk}$.

4. Choose the kernel function , perform kernel principal component analysis on the new EEG feature matrix $Z_{mxk}$, constitute a new EEG feature matrix $Z_{m\times k}$.

5. Randomly divide all EEG feature matrices $Z_{m\times k}$ into two parts: the training part $Z_{m1\times k}$ and the test part $Z_{m2\times k}$, where $m1:m2 = 7:3$.

6. Construct a classifier, then using the training data $Z_{m1\times k}$ to train the classifier.

7. Test the performance of the trained classifier on test data $Z_{m2\times k}$.

8. Obtain the test result through ten-fold cross-validation method.

9. End the algorithm.

3. Results and Discussion

3.1. Test Data for Experiments

Two groups of experiments were conducted. The first group of data was obtained from 10 people, 60s for each person (the first 30s in non-fatigue state and the remaining in fatigue state), which constituted a 600 * 30 data matrix, as shown in the left matrix of figure 1. The other group of data was obtained from 15 people, and 60s of each person, which constituted a 900 * 30 data matrix, as shown in the right matrix of figure 1.

![Figure 1. Entropy data matrix for 10 people (left) and 15 people (right)](image-url)

3.2. Test Based on Entropy and T-test Combined with KPCA

Preprocess and classify the collected EEG signal data, perform the t-test after using different entropy feature extraction algorithms to calculate the p value of two sets of data, use feature vectors which p<= 0.05 to construct a new EEG feature matrix, then perform KPCA to the feature matrix. Randomly select 70% of the data as the training set, and the rest 30% as the test set. SVM classifier [15] was used to test the accuracy of driving fatigue recognition through ten-fold cross-validation algorithm, the test results are shown in table 1 and table 2. “SE_T_KPCA” represents sample entropy and t-test combined with KPCA, the rest may be deduced by analogy. Only the best performing algorithms and their parameters are given in the tables.

The experiment above only uses SVM classifier, in order to compare experiment test results under different kinds of classifiers, six kinds of classifiers are added into this paper to study driving fatigue state recognition under the SE_T_KPCA algorithm: K-nearest neighbour algorithm(KNN) [16], Naive bayes (NB) [17], Random forest(RF) [18], (Linear discriminant(LDA) [19], Decision tree(DT) [20], Artificial neural network(ANN) [21]. The results show that the LDA is the best classifiers with the accuracy of 99.27% and the time of 5.78s in group one, the accuracy of 97.07% and the time of 38.84s in group two.

Comparison of the driving fatigue state recognition time and accuracy between SE_T_KPCA algorithm and SE_KPCA algorithm is shown in table 3, SE_T_KPCA not only had increased the classification recognition rate compared with SE_KPCA, but also improved the time performance.
Table 1. Test result for various entropy and t test combined with KPCA algorithms (10 people).

| Contribution Rate | Parameters | 0.1 | 0.2 | 0.7 | 1 |
|-------------------|------------|-----|-----|-----|---|
| Accuracy [%]      | SE_T_KPCA  | 98.00 | 91.33 | 82.83 | 82.17 |
| Time [s]          | FE_T_KPCA  | 53.24 | 7.78  | 4.37  | 3.62 |

| Contribution Rate | Parameters | 0.95 | 0.95 | 0.99 | 0.99 |
|-------------------|------------|-----|-----|-----|-----|
| Accuracy [%]      | FE_T_KPCA  | 91.67 | 91.33 | 85.33 | 82.83 |
| Time [s]          | SE_T_KPCA  | 51.58 | 8.95  | 6.36  | 4.00 |

| Contribution Rate | Parameters | 0.99 | 0.99 | 0.99 | 0.99 |
|-------------------|------------|-----|-----|-----|-----|
| Accuracy [%]      | SE_T_KPCA  | 98.00 | 91.33 | 85.67 | 82.83 |
| Time [s]          | FE_T_KPCA  | 82.62 | 10.11 | 11.40 | 7.05 |

Table 2. Test result for various entropy and t test combined with KPCA algorithms (15 people).

| Contribution Rate | Parameters | 0.1 | 0.2 | 0.7 | 1 |
|-------------------|------------|-----|-----|-----|---|
| Accuracy [%]      | FE_T_KPCA  | 96.11 | 94.89 | 69.00 | 65.56 |
| Time [s]          | SE_T_KPCA  | 115.18 | 48.71 | 24.38 | 22.91 |

| Contribution Rate | Parameters | 0.95 | 0.95 | 0.99 | 0.99 |
|-------------------|------------|-----|-----|-----|-----|
| Accuracy [%]      | FE_T_KPCA  | 96.11 | 95.22 | 73.00 | 69.11 |
| Time [s]          | SE_T_KPCA  | 158.56 | 81.33 | 37.55 | 36.25 |

| Contribution Rate | Parameters | 0.99 | 0.99 | 0.99 | 0.99 |
|-------------------|------------|-----|-----|-----|-----|
| Accuracy [%]      | FE_T_KPCA  | 96.11 | 95.22 | 75.22 | 74.00 |
| Time [s]          | SE_T_KPCA  | 249.39 | 167.81 | 78.93 | 55.43 |

Table 3. Comparison of the driving fatigue state recognition time and accuracy under the best classifiers (LDA) between SE_T_KPCA and SE_KPCA.

| Groups           | Parameters | SE_T_KPCA | SE_KPCA |
|------------------|------------|-----------|---------|
| Group One        | Accuracy [%] | 99.27     | 98.67   |
|                  | Time [s]    | 5.78      | 23.96   |
| Group Two        | Accuracy [%] | 97.07     | 96.29   |
|                  | Time [s]    | 38.84     | 84.11   |

4. Conclusion
In this paper, firstly we preprocess the EEG signal, and then extract six kinds of entropy features from EEG data, finally perform driving fatigue state recognition based on ENTROPY_KPCA algorithm, which combines entropy features with KPCA algorithm. The experiment results show that KPCA algorithm has significant effect on dimensionality reduction, and the ENTROPY_KPCA algorithm has different experimental effects in different cases of kernel function selection, the algorithm gets highest recognition accuracy but increased required time when RBF is the kernel function of KPCA. To reduce the time required, we add t-test based on the algorithm proposed above, construct driving fatigue state recognition algorithm combined with t-test and KPCA, i.e. the ENTROPY_T_KPCA algorithm. The experiment results show that t-test algorithm has significant effect on the time reducing, after add KPCA algorithm, the ENTROPY_T_KPCA algorithm not only improve the time performing but also improve the recognition accuracy when it chooses the appropriate classifier.

Acknowledgments
This research was financially supported by the National Science Foundation of China [grant numbers 61762045, 61841201, 81460769].
References

[1] Zhang, G.Q, Discussion on driving fatigue and preventive measures. Journal of Technology Wind 2018, 365, 255.

[2] Khalaf, A.; Sejdic, E.; Akcakaya, M, Common spatial pattern and wavelet decomposition for motor imagery EEG- fTCD brain-computer interface. Journal of Neuroscience Methods 2019, 320, 98 – 106.

[3] Wang, L.; Zhang, C.; Yin, X.W.; Fu, R.R.; Wang, H, Non-contact driving fatigue detection technology based on driver physiological signal. Journal of Automotive Engineering 2018, 40, 333–341.

[4] Chen, J.C.; Hong, W.; Wang, Q.X.; Hua, C.C.; Liu, C, Fatigue detection based on cascade convolutional neural network. Journal of Automotive Engineering 2018, 40, 515–520.

[5] Huang, B, Driver Fatigue Detection Algorithm and Implementation Based on Face Features. PhD thesis, Nanjing University of Aeronautics and Astronautics, 2017.

[6] Chai, R.; Naik, G.; Nguyen, T.N.; Ling, S.; Tran, Y.; Craig, A.; Nguyen, H, Driver Fatigue Classification with Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-based System. IEEE Journal of Biomedical & Health Informatics 2017, 21, 715–724.

[7] Mu, Z.; Hu, J.; Min, J, Driver Fatigue Detection System Using Electroencephalography Signals Based on Combined Entropy Features. Applied Sciences 2017, 7.

[8] Pincus.; M., S, Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America 1991, 88, 2297–2301.

[9] Li, Z.; Li, S.; Li, R.; Cheng, B.; Shi, J, Driver fatigue detection using approximate entropic of steering wheel angle from real driving data. International Journal of Robotics and Automation 2017, 32.

[10] Zhang, Y.; Luo, M.W.; Luo, Y, Wavelet transform and sample entropy feature extraction methods for EEG signals. CAAI Transactions on Intelligent Systems 2012, 7, 339–344.

[11] Kumar, Y.; Dewal, M.L.; Anand, R.S, Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. Neurocomputing 2014, 133, 271–279.

[12] Zhang, N.N.; Wang, H.; Fu, R.R, Feature extraction of driving fatigue EEG based on wavelet entropy. Journal of Automotive Engineering 2013, 35, 1139–1142.

[13] Gao, X.W, Kernel PCA Feature Extraction Method and Its Application. PhD thesis, Nanjing University of Aeronautics and Astronautics, 2009.

[14] Zeng, F.; Huang, W.L.; Xia, W.P.; Feng, H, Acoustic Target Recognition Based on Wavelet Packet Feature Energy Operator and Multi-Core Function Combination KPCA. Journal of Electronics Optics & Control 2017, pp. 5–7.

[15] Liu, C.; Zhao, H.B.; Li, C.S.; Wang, H, Motion imagined EEG signal classification based on CSP and SVM algorithms. Journal of Northeastern University 2010, 31, 1098–1101.

[16] Khondoker, M.; Dobson, R.; Skirrow, C.; Simmons, A.; Stahl, D, A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. Statistical Methods in Medical Research 2016, 25, 1804–1823.

[17] Abdulhay, E.; Elamaran, V.; Chandrasekar, M.; Balaji, V.; Narasimhan, K, Automated diagnosis of Epilepsy from EEG signals using Ensemble Learning approach. Pattern Recognition Letters 2017, p. S0167865517301691.

[18] Cutler, A.; Cutler, D.; Stevens, J, Random forests. Machine Learning 2004, 45, 157–176.

[19] Tahernezhadjavazm, F.; Azimirad, V.; Shoaran, M, A review and experimental study on application of classifiers and evolutionary algorithms in EEG based brain-machine interface systems. Journal of Neural Engineering 2017, 15.

[20] Wang, G.; Deng, Z.; Choi, K.S, Detection of epilepsy with Electroencephalogram using rule-based classifiers. Neurocomputing 2016, 228, S0925231216312607.

[21] Rout, N, Analysis and Classification Technique Based On ANN for EEG Signals. International Journal of Computer Science and Information Technologies 2014, 5.