Application of Ensemble Learning in EEG Signal Analysis of Fatigue Driving

Songhui Rao, Kairui Li, Jun Wu and Zhendong Mu*
Jiangxi University of Technology, Nanchang, China

*Corresponding author e-mail: zdmu123@jxut.edu.cn

Abstract. Fatigue detection has now become an important direction of fatigue driving research. As a reliable physiological signal, EEG signals can be used to detect fatigue driving. This passage uses the idea of ensemble learning to establish an ensemble learning classification model of Bagged Tree, RSM Discrimination and RUSBoosted Tree, and uses the fatigue driving experiment data as the object to divide the fatigue driving object into normal state and fatigue state, using Pearson Correlation Coefficient method build a functional brain network. Six methods of Betweenness Centrality, Edge Betweenness Centrality, Clustering Coefficient, Degree, Local Efficiency, and Rich Club Coefficient are used to process and analyze six kinds of EEG signal characteristics, and then bring them into different in the ensemble learning model, the EEG signal characteristics and the ensemble learning model are matched and selected to improve the accuracy and stability of the detection model. By simulating 24 subjects, the results show that the accuracy and robustness of RSM Discrimination ensemble learning model and Degree method are the best, with an average accuracy rate of 90.35% and a standard deviation of 0.081.

Keywords: Fatigue Detection, EEG Signal, Ensemble Learning, Functional Brain Network

1. Introduction
Fatigue driving has become one of the main causes of traffic accidents, and the harm it brings cannot be ignored. At present, fatigue detection is an important research direction of driving fatigue. The EEG signal is the main method of fatigue detection, but also because of its weak signal, small amplitude, large noise, etc. The accuracy of the measurement results is not high and the stability is poor. With the development of science and technology, EEG signals introduce the idea of ensemble learning, which effectively improves the accuracy and stability of EEG signal detection.

The current research on driving fatigue detection is mainly in the following aspects: (1) Detecting the physiological signals of the driver [1]. (2) Detecting the external behavior of the driver [2]. (3) Detecting the driving process vehicle behavior characteristics [3]. The literature [4] believes that EEG signals are the best way to reflect the physiological activities of the brain, so it is a very good way to judge whether fatigue driving. Tran et al. [5] used the second-order difference structure and sampling entropy to perform nonlinear processing and analysis of EEG signals. King L M et al. [6] collected...
and extracted the EEG signals of different drivers in different bands and used neural networks to process EEG signals. The accuracy of the above methods and the ability to adapt to the environment are relatively poor, and ensemble learning has the characteristics of improving accuracy and robustness, so this passage combines the idea of ensemble learning to analyze EEG signals.

Ensemble learning can efficiently solve practical application problems, attracting many research workers to explore and optimize ensemble learning. The RUSBoost algorithm is a hybrid algorithm that combines the under-sampling method and Boosting algorithm [7]. Breiman [8] proposed the Bagging algorithm, which combines base learners from another angle. The random RSM method proposed by Tin Kam Ho [9] reduces the correlation between each classifier by using random partial features instead of all features to train each classifier. The above models have their own advantages. This passage combines their advantages and EEG signal characteristics for combined analysis to improve the accuracy and robustness of the model.

This passage establishes three ensemble learning models through the application of ensemble learning ideas, selects seven methods to process and analyze EEG signals, forms six EEG signal characteristics, and substitutes the six EEG signal features into three ensemble learning models for complete training and integration, check the accuracy and robustness of its different EEG signal characteristics in different models, and screen out the EEG signal features with high accuracy and good robustness and the ensemble learning model. Use the best combination to apply to the research of fatigue driving detection, make the detection accuracy rate higher, and have better stability, and accelerate the pace of putting fatigue driving detection into practical application.

2. Functional Brain Network

2.1 Source of Experimental Data
This experiment was carried out in the EEG signal driving laboratory of Jiangxi University of Technology, using Neuroscan’s 30-lead EEG signal acquisition equipment, and scan4.3, which is matched with the acquisition equipment, as EEG signal acquisition software and preprocessing software. The experiment has been approved by the Academic Committee of Jiangxi University of Technology.

2.2 Feature Acquisition and Processing of Brain Network
This passage uses the Pearson Correlation Coefficient method to obtain the features of the functional brain network. The specific calculation method is as follows:

$$P_{ij} = \frac{\sum_{k=1}^{N}(x_i(k)-\bar{x}_i)(x_j(k)-\bar{x}_j)}{\sqrt{\sum_{k=1}^{N}(x_i(k)-\bar{x}_i)^2} \sqrt{\sum_{k=1}^{N}(x_j(k)-\bar{x}_j)^2}}$$

(1)

Where N is the length of the sampling period, N = 1000, $x_i$ and $x_j$ represent the sample time series of the i-th electrode and the j-th electrode respectively, and k is the time component of the EEG signal sample. Finally, for each sample with 30 electrodes, a symmetric matrix of 30×30 is obtained.

For the processing of functional brain network features, this passage adopts the methods of Betweenness Centrality, Edge Betweenness Centrality, Clustering Coefficient, Degree, Local Efficiency, and Rich Club Coefficient[10-12] for feature processing, and obtain six kinds of 30*600 EEG signal characteristics.

3. Ensemble Learning Model
This experiment uses RUSBoost algorithm and decision tree learner to form a RUSBoosted Tree ensemble learning model; uses Bagging algorithm, decision tree learner to form a Bagged Tree ensemble learning model; uses RSM algorithm to discriminate learner to form RSM Discrimination ensemble learning model.

3.1 RUSBoost Algorithm
RUSBoost algorithm is a combination of under-sampling method and Boosting algorithm. The idea of this algorithm in literature [13] is: divide the data set into \( N \) parts, and the random weight of the total number divided by \( N \) is the initial weight of the sub-data set. Then train the sub-data set, use regularization to update the weights, a sub-data set repeats the training classifier that meets the conditions, and finally selects the best model.

3.2 Bagging Algorithm

Literature [14] believes that the Bagging algorithm obtains multiple training subsets from the samples with replacement in the initial training set through resampling, and then uses different training subsets to train individual learners, and finally integrates them into an overall model.

3.3 Random RSM Algorithm

The random RSM method [15] randomly samples the feature attributes of the training samples to form a new multiple different training sample sets, and then trains different training sample sets to generate base classifiers, because each base classifier the difference in classification performance makes the classifier ensemble better, and then the base classifier in the ensemble is ensemble with linear weights to obtain the ensemble classifier.

3.4 Ensemble Strategy

The model built in this passage uses the majority voting method and the weighted voting method. In the majority voting method, when the number of votes obtained for a certain type of subject is greater than half of the individual learner or the highest number of votes is selected as the output of the prediction result. Weighted voting method [16]: the weight is inversely proportional to the estimated individual learner error, and then \( K \) individual learners are weighted to get the result. The expression is as follows:

\[
H(x) = \sum_{i=1}^{K} u_i \nu_i(x)
\]

Where \( u_i \) is the weight of the \( i \)-th individual learner, \( u_i > 0 \), \( \sum_{i=1}^{K} u_i > 0 \).

4. Results and Analysis

This passage uses the methods of Pearson Correlation Coefficient, Betweenness Centrality, Edge Betweenness Centrality, Clustering Coefficient, Degree, Local Efficiency, and Rich Club Coefficient to analyze and process the data obtained in the laboratory to form a 30*300 fatigue state EEG signals and 30*300 normal EEG signals, and then synthesize these two EEG signals to form six 30*600 EEG signal characteristics, which are substituted into the three integrated learning models of Bagged Tree, RSM Discrimination and RUSBoosted Tree, all models use 5-fold cross-validation, and PCA is disabled.

4.1 Classification Results of Subject 1

Using the six EEG signal characteristics formed by all samples of subject 1 to complete training and integration of three ensemble learning models, three ROC curves were obtained. The experimental results are shown in Figure 1:
Figure 1. ROC curves of the six features of the three ensemble models

It can be seen from the figure that the RSM Discriminant model is more friendly to the six EEG signal features, with better accuracy and stability; while the BUSBoosted Tree model has a lower accuracy and a larger feature fluctuation. Comparing the six features in the three models, it can be seen that Degree and Edge Betweenness Centrality have better accuracy and stability. Rich Club Coefficient cannot draw ROC diagram in the RSM Discriminant model and Degree cannot draw ROC diagram in the BUSBoosted Tree model, indicating that the accuracy and stability of different features are different in different models. For Subject 1, the combination of RSM Discriminant model and Degree is the most reasonable, with an accuracy rate of 98.8% and AUC=0.99.
4.2 Classification Results of All Subjects
The six EEG signal characteristics formed by all samples of 24 subjects were completely trained and ensemble for three ensemble learning models.

For the Bagged Tree ensemble learning model, the experimental results are shown in Figure 2:

![Figure 2. Distribution of accuracy rates of six features of 24 subjects in the Bagged Tree ensemble learning model](image)

The results show that the accuracy and stability of the rich club coefficient and the clustering coefficient in the Bagged Tree model are not very good. The accuracy and stability of Local Efficiency, Betweenness Centrality and Degree of the edge are relatively good, among which 24 people the average accuracy is the highest, reaching 87.2%, and the standard deviation is 0.076. Therefore, in this model, the effect of using one of the EEG signal characteristics of Local Efficiency, Edge Betweenness Centrality and Degree is better.

For the RSM Discriminant ensemble learning model, the experimental results are shown in Figure 3:

![Figure 3. Distribution of accuracy rates of the six features of 24 subjects in the RSM Discriminant ensemble learning model](image)

The results show that the accuracy and stability of the Rich Club Coefficient is very poor, and it is not suitable for the identification of EEG signal characteristics. The accuracy and stability of
Betweenness Centrality are not ideal. The Degree performs best, with an average accuracy rate of 90.35% and a standard deviation of 0.081. Therefore, this model has the best effect of using Degree as the identification of EEG signals.

For the RUSBoosted Tree ensemble learning model, the experimental results are shown in Figure 4:

![Figure 4. Distribution of accuracy of six features of 24 subjects in the RUSBoosted Tree ensemble learning model](image)

The results show that the accuracy and stability of Betweenness Centrality are better than the other five features. Other features are not satisfactory in the RUSBoosted Tree ensemble learning model. The average accuracy of Betweenness Centrality is 79.43%, and the standard deviation is 0.0999.

In summary, the performance of the RUSBoosted Tree in the three ensemble learning models is not ideal, so try not to use this model to judge the characteristics of EEG signals. The accuracy and stability of the RSM Discrimination ensemble learning model are relatively good. The accuracy and stability of the model and the Bagged Tree model are relatively good. So the RSM Discriminant ensemble learning model and the Degree method are used to detect fatigue driving with the best accuracy and stability.

5. Conclusion
In this passage, seven methods including Pearson Correlation Coefficient are used to process and analyze EEG signals, and the processed data is brought into three ensemble learning models for training and integration. Finally, the accuracy and ROC curve of the three ensemble learning models with six characteristics of 24 subjects were obtained. Through experiments, it is found that for the same subject, the accuracy and stability of different features under different models may also be quite different. It can be found in 24 subjects that the RUSBoosted Tree ensemble learning model is not suitable for the detection of fatigue driving EEG signals. Ensemble learning model and EEG signal characteristics in the same case, the accuracy and stability of different subjects are different, indicating that the detection of EEG signals is also affected by different subjects. Finally, it is concluded that the RSM Discriminant ensemble learning model and the Degree method are more suitable for fatigue driving EEG signal detection.

Acknowledgements
This work was supported by Natural Science Foundation of China (No 61762045 and No 61841201), Jiangxi Education and Teaching Reform Project (No. JXJG-17-24-12), Natural Science Foundation of Jiangxi (No.20171BAB202031), Science and Technology Research Project of Jiangxi Education Department (No.GJJ161143)
References

[1] Mao Z, Yan X P, Wu C Z. Driving Fatigue Identification Method Based on Physiological Signals [J]. 2008.

[2] Zhi-Chun L I, Ren H E, Mou-You L, et al. Study Situation and Developing Trend on Detecting Technology of Motor Driver's Fatigue [J]. Journal of Agricultural Mechanization Research, 2006.

[3] Zuojin L, Shengbo L, Renjie L, et al. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions[J]. Sensors, 2017, 17(3):495.

[4] Wu S B, Gao L, Wang L A. Detecting Driving Fatigue Based on Electroencephalogram [J]. Beijing Ligong Daxue Xuebao/Transaction of Beijing Institute of Technology, 2009, 29(12):1072-1075.

[5] Tran Y, Wijesuryia N, Thuraisingham RA, et al. Increase in regularity and decrease in variability seen in electroencephalography (EEG) signals from alert to fatigue during a driving simulated task [J]. Conf Proc IEEE Eng Med Biol Soc, 2008, 2008:1096-1099.

[6] King L M, Nguyen H T, Lal S K L. Early Driver Fatigue Detection from Electroencephalography Signals using Artificial Neural Networks[ C] // International Conference of the IEEE Engineering in Medicine & Biology Society. Conf Proc IEEE Eng Med Biol Soc, 2006:2187-2190.

[7] Seiffert C, Khoshgoftaar T M, Van Hulse J, et al. RUSBoost: A Hybrid Approach to Alleviating Class Imbalance [J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 2010, 40(1):185-197.

[8] Breiman L. Bagged Predictors [J]. Machine Learning, 1996, 24(2):123-140.

[9] Ho T K. The random RSM method for constructing decision forests [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(8):832-844.

[10] Kas M, Wachs M, Carley K M, et al. Incremental algorithm for updating betweenness centrality in dynamically growing networks[C]// IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining. IEEE, 2013:33-40.

[11] Wanzeng, Kong, Weicheng, et al. Investigating Driver Fatigue versus Alertness Using the Granger Causality Network. [J]. Sensors, 2015.

[12] Ling L. Synchronization of spatiotemporal chaos in large scale rich-club network [J]. Acta Physica Sinica, 2010, 59(10):6864-6870.

[13] Muhammad Adil, Nadeem Javaid, Umar Qasim, et al. LSTM and Bat-Based RUSBoost Approach for Electricity Theft Detection. 2020, 10(12)

[14] Ji-Wei X U, Yu Y, Software S O, et al. A survey of ensemble learning approaches [J]. Journal of Yunnan University (Natural ences Edition), 2018.

[15] Zhong-Xia Z, Zheng-Qun W, Chun-Lin X U, et al. Discriminant analysis method ensemble based on RSM [J]. Computer Engineering and Design, 2013.

[16] WEI Chen, CHEN Lanlan, ZHANG Ao. Emotion Recognition of EEG Based on Ensemble Convolutional Neural Networks [J]. Journal of East China University of Science and Technology, 2019, 45(4): 614-622.