IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

Predictive Analysis of Mental Fog Using Machine Learning

Urang Awajionyi Samuel[1] and Onuodu, Friday Eleonu[2]

Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Nigeria[1]

Department of Computer Science, University of Port Harcourt, Nigeria[2]

Abstract: Mental fog, also known as confusion, is one of the main reasons for poor performance in the learning process or any type of daily task that involves and requires thinking. Detecting confusion in the human mind is a real time paradigm that appears to be more difficult and important tasks that can be applied to online education, driver fatigue detection, etc. The Random Forest model achieve a better performance compared to other machine learning approaches and shows a great robustness evaluated by cross validation. We can predict if a student is confused about 100% accuracy. In addition, we found that the most important characteristic for detecting brain confusion is the beta 2 and gamma 1 wave of the Electroencephalography (EEG) signal. Our results suggest that machine learning is a potentially powerful tool for modeling and understanding brain activity. This work could be beneficial to individuals, to Ministry of Health, patients with brain diseases and to any other organization that deals on human state of mind in terms of performance.

Keywords: EEG, Mental Fog, Random Forest Model, Electroencephalography (EEG)

I. INTRODUCTION

Brain fog is a constellation of symptoms that includes reduced mental acuity and cognition, inability to concentrate and perform multiple tasks, and short and long-term memory loss. It is well distributed not only in patients with brain diseases, but also in healthy people [1]. Confusion, which is one of the symptoms of brain fog, can reduce people's concentration and cognition. Detecting and preventing brain confusion is very important and has many benefits. When a driver is confused, his knowledge is reduced. This is very dangerous and can have serious consequences. Another example is the Open Online Course (MOOC), which is an online course for unlimited participation and open access through the web. Although there are several MOOC websites, the format is still lacking compared to traditional classes. Valerie et al. [2] have shown that the lack of feedback is one of the main problems in distance teacher-student communication. Students may feel confused about the conference as long as the teacher does not realize and the conference continues. If there is a practical approach to immediately detect student confusion, it will help teachers better understand student status and respond accordingly.

Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity in the brain. In clinical settings, EEG refers to the recording of the spontaneous electrical activity of the brain during a given period, recorded from several electrodes placed in the scalp. It measures the voltage fluctuations resulting from ionic currents in brain neurons. The EEG is most often used to diagnose epilepsy, which causes abnormalities in the EEG readings. It is also used to diagnose sleep disorders, coma, encephalopathy and brain death. Our motivation to choose EEG signals as data to detect confusion in people's brains is that the EEG signal is continuous and contains some patterns of state transitions. Our hypothesis is that when people are confused, their EEG signal will be different than normal. It is possible to create a model to analyze continuous data and predict whether the subject is confused or not.

II. RELATED WORK

In general, it is recognized that visual inspection of EEG waveforms can reliably identify driver fatigue or drowsiness. Many researchers apply machine learning methods to EEG data to perform various tasks, such as detecting driver fatigue. Yeo et al. [3] used Support Vector Machines (SVM) to detect driver drowsiness. Their results showed that the extraction of the characteristics of four EEG frequency bands produced an accuracy of 99.3%. In addition to sleepiness, Subashi et al. [4] applied SVM classifiers to predict whether EEG signals represented epileptic seizures and reached 100% accuracy. Wang et al. [5] have demonstrated the possibility of using EEG data to detect student confusion when watching MOOC videos. They analyzed the EEG data using Gaussian Naive Bayes classifiers. Naive Gauss Bayes classifiers achieved a classification accuracy of 57%. This article explores ways to improve the result of this confusing classification in the same data set.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

In-depth learning has recently demonstrated its effectiveness in many classification-related tasks compared to traditional machine learning methods. Boureau et al. [6] proposed a Deep Belief Network (DBN) that can learn a high-level function based on raw data and can capture higher order dependencies among the observed variables. Hajinoroozi et al. [7] applied DBN to EEG signals to predict the cognitive states of drivers. The classifiers that use the functionality learned by the DBNs outperformed those that used the Main Component Analysis (PCA) features. Lee et al. [8] introduced convolutive DBNs to learn better feature representations and machine learning approaches overcome through the use of raw features. Since the EEG signal is a time series, it can be difficult to detect events in EEG signals using fixed length characteristics.

Laurent et al. [2] proposed an approach based on a hidden Markov model for the detection of mental status in EEG signals. Petrosiana et al. [9] have shown that recurrent neural networks can identify the first signs of Alzheimer's disease in long-term EEG records. Few studies have focused on the detection of confusion of the EEG signal using Deep Neural Networks (DNN). Since LSTM recurrent neural networks can easily analyze time series data, the current study applies them to detect confusion in the EEG signal.

We use batch standardization, which has been shown to accelerate the formation of DNN. Ioffe et al. [10] proposed a batch normalization layer, which uses mini-batch statistics to normalize the characteristics of deep neural networks, providing the same accuracy in much less time. Laurent et al. [11] have shown that the application of batch normalization to recurrent neural networks leads to a faster convergence of training.

III. DATA AND METHODOLOGY

From the EEG data of 10 students, our task is to predict their confusion using machine learning methods. The data comes from the "EEG Brain Confusion" dataset, EEG data is downloaded from Kaggle challenge. 10 students were assigned to watch 20 videos, 10 of which were previously labeled "easy" and 10 "difficult". Each video lasted approximately 2 minutes. For "difficult" videos, the two-minute clip was carefully taken in the middle of a theme to make the videos more confusing.

The students used a single-channel wireless Mindset EEG device that measured activity in the frontal lobe. The mentality measures the voltage between an electrode resting on the forehead and two electrodes (one ground and one reference) each in contact with an ear. After each session, the student assessed their level of confusion on a scale of 1 to 7, one at least confused and seven among the most confused correspondents. These labels were quantified in two classes that indicated whether the students were confused or not. The two-class label serves as the objective for our prediction task.

Since the confusion label is true or false, our problem is a classification problem of two kinds. In theory, many machine learning approaches can be applied to this task. To take advantage of the properties of EEG data, we propose a confusion detection framework using Random Forest model and R software for the data analysis.

The Proposed System

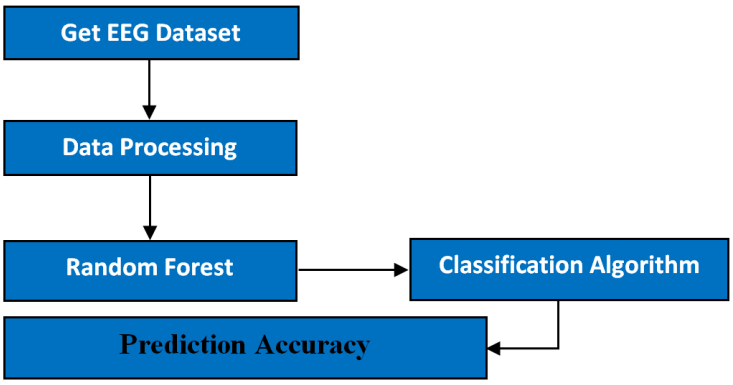


Figure 1.0 Proposed System block diagram

The working of this system is described in a step by step:
1. EEG Dataset collection.
2. Data Processing: Variables selection process selects the useful attributes for the prediction, clean and range in desired form.
3. Random Forest: Applying random forest classification technique to EEG dataset to predict the student mental fog state and also get the accuracy.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

Random Forest

Random Forest is also a supervised machine learning algorithm. This technique can be used for regression and classification tasks, but it generally works best for classification tasks. As the name implies, the Random Forest technique takes into account several decision trees before giving a result. Therefore, it is essentially a set of decision trees. This technique is based on the belief that more trees would converge on the right decision. For classification, use a voting system and then decide on the class, while in the regression take the average of all the results of each of the decision trees. It works well with large data sets of high dimension.

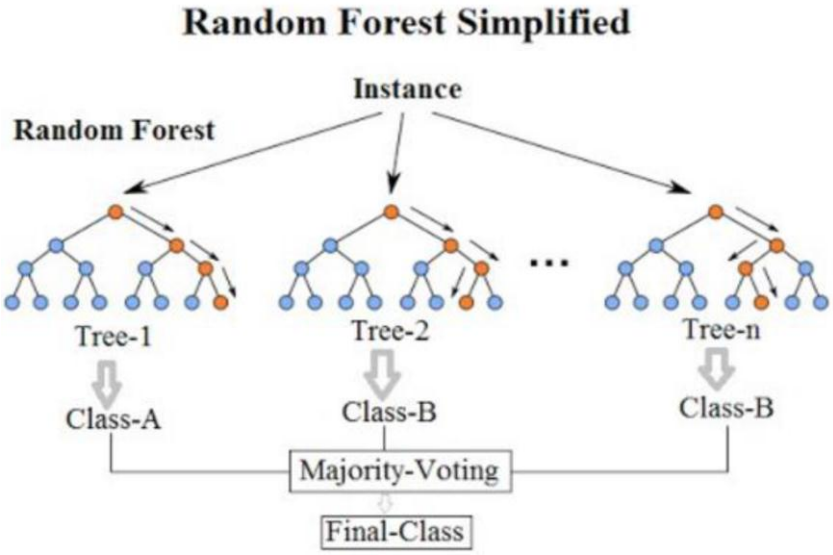


Figure 2: Random Forest (Ramalingam et al, 2018)

IV. RESULTS AND DISCUSSION

Table 1.0. The data structure of dataset.
'data.frame': 12811 obs. of 15 variables:
$ SubjectID : num 0 0 0 0 0 0 0 0 0 0 ...
$ VideoID : num 0 0 0 0 0 0 0 0 0 0 ...
$ Attention : num 56 40 47 47 44 44 43 40 43 47 ...
$ Mediation : num 43 35 48 57 53 66 69 61 69 69 ...
$ Raw : num 278 -50 101 -5 -8 73 130 -2 17 -59 ...
$ Delta : num 301963 73787 758353 2012240 1005145 ...
$ Theta : num 90612 28083 383745 129350 354328 ...
$ Alpha1 : num 33735 1439 201999 61236 37102 ...
$ Alpha2 : num 23991 2240 62107 17084 88881 ...
$ Beta1 : num 27946 2746 36293 11488 45307 ...
$ Beta2 : num 45097 3687 130536 62462 99603 ...
$ Gamma1 : num 33228 5293 57243 49960 44790 ...
$ Gamma2 : num 8293 2740 25354 33932 29749 ...
$ predefinedlabel : num 0 0 0 0 0 0 0 0 0 0 ...
$ user.definedlabeln: num 0 0 0 0 0 0 0 0 0 0 ...

From table1.0, it is clear that we have 12811 observations of 15 variables and the data class is a data frame, also all the variables are numeric in nature, range from subject id to user definedlabeln. Hence we set the predefinedlabel as a factor variables for our classification.

Table 2.0. The factor variables		
0	1	Total
6662	6149	12811

The table 2.0 shows the factor variables having numbers ranging between 0 and 1. The 0 means not confused and 1 confused.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

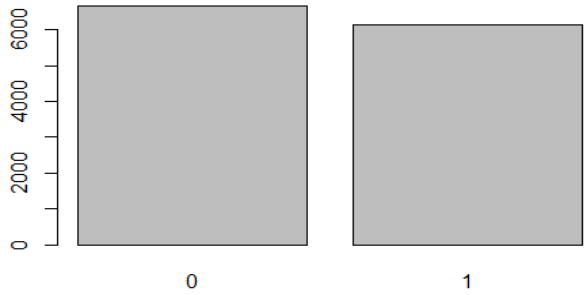


Figure 2.0 Histogram of the factor variable.

The figure 2.0 gives us a graphical view of the factor variable. Also from table 2, it can be seen that 0 which is not confused greater than one which is confused.

Table 3.0. Applying Random Forest to train data
Call: randomForest(formula = predefinedlabel ~ ., data = train) Type of random forest: classification Number of trees: 500 No. of variables tried at each split: 3 OOB estimate of error rate: 0% Confusion matrix: 0 1 class.error 0 5312 0 0 1 0 4919 0

From table 3.0, after applying random forest algorithm to the train data, with default number of trees 500 and split $\mathbf{m^{try}}$ 3, we observed that out of bag error rate was 0% indicating that there is no misclassification of data from our data set.

Table 4.0 Confusion Matrix of Train data	Table 5.0 Confusion Matrix of Test data
Confusion Matrix and Statistics Reference Prediction 0 1 0 5312 0 1 0 4919 Accuracy : 1 95% CI : (0.9996, 1) No Information Rate : 0.5192 P-Value [Acc > NIR] : < 2.2e-16 Kappa : 1 Mcnemar's Test P-Value : NA Sensitivity : 1.0000 Specificity : 1.0000 Pos Pred Value : 1.0000 Neg Pred Value : 1.0000 Prevalence : 0.5192 Detection Rate : 0.5192 Detection Prevalence : 0.5192 Balanced Accuracy : 1.0000 'Positive' Class : 0	Confusion Matrix and Statistics Reference Prediction 0 1 0 1350 0 1 0 1230 Accuracy : 1 95% CI : (0.9986, 1) No Information Rate : 0.5233 P-Value [Acc > NIR] : < 2.2e-16 Kappa : 1 Mcnemar's Test P-Value : NA Sensitivity : 1.0000 Specificity : 1.0000 Pos Pred Value : 1.0000 Neg Pred Value : 1.0000 Prevalence : 0.5233 Detection Rate : 0.5233 Detection Prevalence : 0.5233 Balanced Accuracy : 1.0000 'Positive' Class : 0

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

From table 4.0, we observed that the accuracy of the prediction is 1, which is 100%, also the confidence interval is between (0.9996 – 1), which very good. Hence we now test which our test data set for validation purpose. Also, from the table 5.0, it is again observed the confidence interval lies between (0.9986 – 1) and the prediction accuracy is 100%. We need to further confirmed with the error rate of the model.

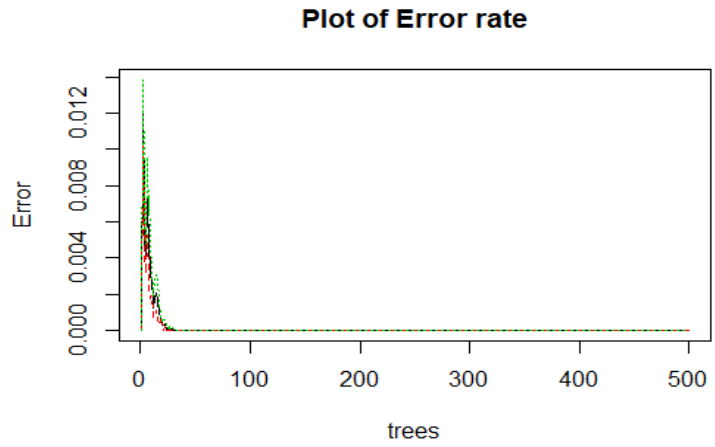
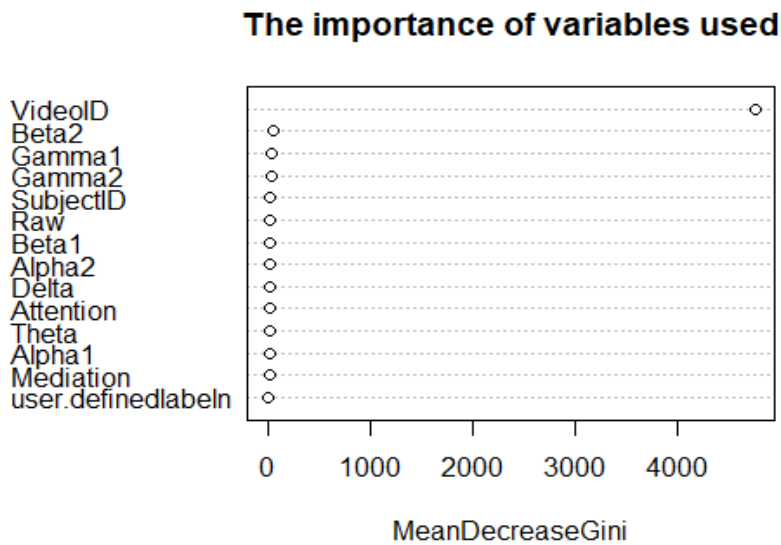


Figure 3.0. The out of bag error rate.

The figure 3.0 shows that the data is error free, meaning that there is no misclassification of data, in the data set.

Table 6.0. The importance and contribution of variables use



The table 6.0 shows us the importance and contribution of variables used. So from our 15 variables in the data set the most importance are Beta2, Gamma1, gamma2 and so on, the list is the user defined labeln.

V. CONCLUSION AND FUTURE WORK

We have proposed the random forest model to detect student confusion when watching online lesson videos. The accuracy obtained by our model is better compared to other machine learning methods, including a single layer RNN-LSTM, SVM, KNN model and provides the latest results. When analyzing the contribution of each characteristic to the model, we find that the characteristics beta2, gamma1 and gamma2 are the most important in this task. We plan to validate our model in a larger EEG data set.

Predicting similar tasks in the future, relating to the EEG, such as detecting driver drowsiness, stress level and others, the random forest model can be compared to the rest machine learning algorithms.

REFERENCES

[1]. Theoharides T.,C., Stewart J. M, and Erifili H., (2015). Brain "fog," inflammation and obesity: key aspects of neuropsychiatric disorders improved by luteolin. Frontiers in euroscience. 9:225.
[2]. Vézard L., Pierrick L., Marie C., Frédérique F. and Leonardo T.,(2015). EEG classification for the detection of mental states. *Applied Soft Computing*. 32,113–131.
[3]. VM Yeo M., Xiaoping L., Kaiquan S., and Wilder-Smith Einar PV.(2009). Can SVM be used for automatic EEG detection of drowsiness during car driving? Safety Science. 47(1):115–124.
[4]. Subasi A., and Ismail G. M., (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications. 37(12):8659–8666.
[5]. Wang H., Yiwei L., Xiaobo H., Yucong Y., Zhu M. and Kai-min C., (2013). Using EEG to improve massive open online courses feedback interaction. AIED Workshops. 2013
[6]. Boureau Y. & Cun Yann L,(2008). Sparse feature learning for deep belief networks. Advances in neural information processing systems. 1185–1192.
[7]. Hajinoroozi M, Tzyy-Ping J, Chin-Teng L, & Yufei H, (2015). Signal & Info Processing (ChinaSIP), IEEE China Summit & International Conf on. IEEE; 2015. Feature extraction with deep belief networks for driver's cognitive states prediction from EEG data, 812–815
[8]. Lee H., Roger G., Rajesh R., Ng Andrew Y., (2009). Proceedings of the 26th annual international conference on machine learning. ACM; 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, 609–616.
[9]. Petrosian A.A., Prokhorov D.V., Lajara-Nanson W., and Schiffer R. B., (2001). Recurrent neural network-based approach for early recognition of alzheimer's disease in EEG. Clinical Neurophysiology. 112(8), 1378–1387.
[10]. Ioffe S. & Christian S.,(2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167
[11]. Cooijmans T., Nicolas B., César L., Çağlar G. and Aaron C.,(2016). Recurrent batch normalization. arXiv preprint arXiv:1603.09025.
[12]. Wang H., (2016). EEG brain wave for confusion.2016 Online: https://www.kaggle.com/wanghaohan/eeg-brain-wave-for-confusion.
[13]. NeuroSky.,(2009).NeuroSky's eSense meters and detection of mental state.
[14]. Laurent C., Gabriel P., Philémon B., Ying Z., Bengio and Yoshua B. (2016). Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. IEEE; 2016. Batch normalized recurrent neural networks. 2657–2661.
[15]. Hochreiter S., and Jürgen S., (1997). Long short-term memory. Neural computation. 9(8) 1735–1780.