SEP 786 Course Project

EEG Analysis of Confusion Levels in College Students While Watching Educational Videos Analysis

Henis Nakrani (400547270)

Hardik Navadiya (400547048)

Darshan Parabadiya (400544636)

Department of Systems and Technology

McMaster University

Course Number : Artificial Intelligence and Machine Learning Fundamentals

Submitted to: Dr. Jeff Fortuna

McMaster University

Abstract

The purpose of the study was to investigate the relationship between frontal lobe EEG signals and college students' self-reported levels of perplexity during instructional video segments. Ten college students participated in the experiment and were shown twenty separate two-minute video clips that were divided into areas that were either potentially difficult (such as quantum mechanics or stem cell research) or non-confusing (such as fundamental algebra and geometry). Electrodes on the forehead and ears were used to measure frontal lobe activity in the EEG data, which were obtained using a single-channel wireless MindSet. On a standardized scale of 1 to 7, self-rated confusion levels were used to signify confusion or non-confusion.

Introduction

Enhancing learning experiences requires an understanding of the cognitive states that arise during the consumption of educational content. This study explores the connection between college students' self-reported perplexity, EEG signals, and comprehension of video information. Participants were shown carefully chosen educational videos that were both potentially perplexing and non-confusing, all the while EEG activity was being monitored and subjective confusion ratings were being collected..

Data Inputs (Kaggle Link)

10 college students participated. 20 videos were used: 10 non-confusing, 10 potentially confusing, each 2 minutes long. EEG data captured via a single-channel MindSet measuring voltage between a forehead electrode and two ear electrodes. Self-rated confusion levels were obtained on a 1-7 scale.

We have 13 features in our dataset. SubjectID, VideoID, Attention, Mediation, Raw, Delta, Theta, Alpha1, Alpha2 ,Beta1, Beta2, Gamma1, Gamma2. First We did data preprocessing and then we applied feature selection and PCA to our data and performed model LDA(linear discriminant and KNN) and conducted 4 experiments.

Experiments

1. Forward search Feature selection with linear discriminant analysis (LDA)
2. Forward search Feature selection with K Nearest Neighbor (KNN)
3. PCA (Principal component analysis) with linear discriminant analysis (LDA)
4. PCA (Principal component analysis) with K Nearest Neighbor (KNN)

Results and Discussion

	Accuracy	Training time	Testing time
Forward search with LDA	99%	0.0034	0.0004
Forward search with KNN	98%	0.0049	0.14
PCA with LDA	99%	0.0030	0.004
PCA with KNN	96%	0.0040	0.1

By looking at the results of forward search, that output label is more dependent on these 4 SubjectId, VideoID, Meditation and Raw.

Accuracy : Forward search gives very high accuracy results from starting and then it reduces gradually (99% to 97%) when we increase the features in forward search While in PCA in starting it gives very less accuracy and over the time it increases (59% to 99%) the accuracy when we add the components.

Training and Testing time: LDA takes very less time to train and test comparatively then KNN in principal component analysis and Forward search both.

Confusion Matrix: By looking at the confusion matrix in forward search, non-diagonal values in starting were zero but when we increase the features then the non diagonal values increase. Whereas, In PCA, the non-diagonal matrix was very high in starting, but when we increase the components the non-diagonal matrix becomes very less.

Conclusion

The study shows that picking certain features for the model makes it really accurate at first but more important. Meanwhile, another method gradually gets better at accuracy and keeps things simpler, pointing out how important basic brain signals and who the person is are for understanding how confused students get during videos.

This research highlights how the way we choose what information to use in the model affects how well it works and how complicated it gets. One way is super accurate quickly but gets complex, while the other way gets better slowly and stays simpler. It shows that basic brain data and personal details matter a lot in figuring out how confused students get while watching educational videos.