PREDICTING CONFUSION USING K NEAREST NEIGHBOR ALGORITHM FOR EEG DATA

PALAK VYAS
M.TECH STUDENT
LNCT BHOPAL

MRS.KIRTI JAIN
ASSIS.PROFESSOR
LNCT BHOPAL

ABSTRACT

Stress is major concern these days. According to WHO, stress is a mental health problem affecting the life of one in four citizens. According to neuroscience, human brain is the main target of mental stress, because only human brain can determine a situation which is threatening and stressful. Using wearable sensors and bio signal processing, several technologies are developed for human stress detection. Human bio signals such as electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG), galvanic skin response (GSR), blood volume pulses (BVP), blood pressure (BP), skin temperature (ST) and respiration are used to detect stress. In proposed method, electroencephalography (EEG) is used to detect confusion among students. We collected EEG signal data from 10 college students while they watched MOOC video clips. We extracted online education videos that are assumed not to be confusing for college students, such as videos of the introduction of basic algebra or geometry. We also prepare videos that are expected to confuse a typical college student if a student is not familiar with the video topics like Quantum Mechanics, and Stem Cell Research. In this study we use the Python environment for processing the EEG signals. KNN algorithm is applied for classifying whether the student is least confused or most confused. Obtained classification report, shows 83% accuracy performance for this data using KNN.

KEYWORDS: Mental stress, electroencephalography (EEG), Machine learning, K Nearest Neighbor (KNN).

I.INTRODUCTION

Stress at basic level is defined as a condition wherein a person is expected to response more than the sheer pressure and where he/she can barely content with the desired expectation. Stress is major concern these days. According to WHO, stress is a mental health problem affecting the life of one in four citizens. Well being is categorized in two state-Physical, Psychological wellness including stress, depression, fatigue, anger. Basically, psychosocial stress exist in everyone day today life, resulting into degrading overall quality of life by adversely affecting person's emotion performance, health. Hence, it is become a social concern since it is a reason behind functional disability during routine work. Psychosocial stress is a leading cause of several psychosocial disorders like depression, stroke and cardiac arrest. According to neuroscience, human brain is the main target of mental stress, because only human brain can determine a situation which is threatening and stressful.

To determine the stress conditions, clinically we use a questionnaire or interview which are subjective method. Physically stress can be understood by examining pupil dilation, blink rate, facial gestures. Physiological biomarkers of stress from automatic nervous system exist in the form of heart rate and hate rate variability, respiration and skin conductance. Stress detection is an ongoing research topic among both psychologist and engineers. Whereas less research is available on reduction of stress method in term of technology but studies are available stress detection methods.

Using wearable sensors and bio signal processing, several technologies are developed for human stress detection. Human bio signals such as electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG), galvanic skin response (GSR), blood volume pulses (BVP), blood pressure (BP), skin temperature (ST) and respiration are used to detect stress. Physiological features are also used to measure the stress level using physiological signals. There is difference between individual physiological features. It changes when faces a stressful condition. In proposed method EEG stress detection technology is applied. EEG is one among the most reliable and trustworthy sources to note down the electrical activity of human brain along the scalp. Voltage fluctuation resulting from ionic current within the neuron of the brain, EEG is used. EEG is also applicable to diagnose the epilepsy, coma, sleep disorder, tumors. Hardware cost is comparatively lower than other techniques or methods thus more preferred compare to other technique, EEG is most superior choice because it is suitable tool for physiological research. When subject perform some behavioural or it is out of laboratory, EEG can be used as wearable sensor. EEG can trace brain changes during different phases of life without disturbing a patient for e.g. EEG sleep analysis.

II.LITERATURE SURVEY

A literature review conducted over analysis of stress using physiological signals and evaluation of stress level.

Related Research

Stress detection is an ongoing research topic among psychologist, scientist and engineers.Previous work on stress measurement has been focused on the collection and analysis of physiological data and the identification of the correlation between perceived stress and multiple physiological features.

Qianli Xu, Tin Lay Nwe, and Cuntai Guan (2015) [1], Proposes a Cluster-Based Analysis method to determine stress using physiological signals, which accounts for inter subject differences. This research work uses the clustering process that assigns the subject into different subgroups, so as to develop the inherent homogeneity of subject's stress response within the cluster. Thus the inter subject differences are repeatedly accommodated, and the overall accuracy of stress evaluation is improvedor increases.

Chee-Keong Alfred Lim and Wai Chong Chia (2015) [2], Focuses on evaluating to what extent a single electrode EEG headset – NeuroSky MindWave is able to classify brainwave in terms of subject's stressor level. In this study they use the MATLAB environment for processing the EEG signals. By reducing the number of electrodes needed, it also means cheaper EEG headset can be used to diagnose various mental disorders.

Tong Chen, Peter Yuen, Mark Richardson, Guangyuan Liu, and Zhishun She (2014) [3], Present a method to detect psychological stress in a non-contact manner using a human physiological response. Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tr¨oster, and Ulrike Ehlert (2010) [4], Developed a personal health system for detecting the stress. They use the discriminative power of electro dermal activity (EDA). This EDA power is use for distinguishing stress from office work load. They do the analysis on EDA data and evaluate the information about stress level of a person.

Awanis Romli, Arnidcha Peri Cha (2009) [5], Suggested a system which deals with stress management. They change the manual system into computerized one. This system does the stress management for particular individual based on their activity interest and provides a solution for stress management. This system using combination of rule based technique and Holland's Self Directed Search Model to determine the best solution for managing the stress. This system use Rule-Based technique. In this technique user is given a test to develop the system's knowledge-based. From this test system can determine the user's interest, behaviour and through their thinking, the system gives the best solution to manage user's stress according to their interest.

Jennifer A. Healey and Rosalind W. Picard (2005) [6], Shows how physiological data is useful during real-world driving task to determine a driver's relative stress level. In this study they show how physiological sensor can be used to obtain electrical signals that can be processed automatically by an on-board computer to give dynamic indications of a driver's natural state under natural driving conditions. The experiment was design to monitor the driver's physiological reaction during real-world driving situation under normal condition.

In recent studies, changes in the EEG absolute power and in connectivity measures such as coherence and mutual information have been shown to vary due to stress [7]. Similarly, asymmetry in EEG alpha power has been shown to be influenced by HRV biofeedback during stress therapy [8].Another study discussed EEG alpha asymmetry and revealed stress-related disorders in a virtual reality environment [9]. EEG eigenvalue decomposition was utilized for stress level classification [10]. Another study proposed an EEG-based brainwave balancing index to assess the stress level of uni-versity students during their studies [11].

III.MATERIAL PARTICIPANTS:

This is a specialized work topic:

A.STUDY PARTICIPANT

Ten healthy subjects, including male and female both (18 to 22 years old) were selected based on having no previous medical record or head injury and not using any medication to increase cardiac activation.

B. EXPERIMENT DESIGE

In this study to create confusion for typical collage students, we have shown them videos of non-familiar topics like quantum mechanics, Stem cell research. Herein basically, we collected EEG signal data from 10 college students while they watched MOOC video clips. We extracted online education videos that are assumed not to be confusing for college students, such as videos of the introduction of basic algebra or geometry. We also prepare videos that are expected to confuse a typical college student if a student is not familiar with the video topics like Quantum Mechanics, and Stem Cell Research. We prepared 20 videos, 10 in each category. Each video was about 2 minutes long. We chopped the two-minute clip in the middle of a topic to make the videos more confusing. The students wore a single-channel wireless MindSet that measured activity over the frontal lobe. The MindSet 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 rated his/her confusion level on a scale of 1-7, where one corresponded to the least confusing and seven corresponded to the most confusing. These labels if further normalized into labels of whether the students are confused or not. This label is offered as self-labelled confusion in addition to our predefined label of confusion.

<u>Content</u>

These data are collected from ten students, each watching ten videos. Therefore, it can be seen as only 100 data points for these 12000+ rows. If you look at this way, then each data point consists of 120+ rows, which is sampled every 0.5 seconds (so each data point is a one minute video). Signals with higher frequency are reported as the mean value during each 0.5 second.

<u>EEG_data.csv</u>: Contains the EEG data recorded from 10 students
<u>demographic.csv</u>: Contains demographic information for each student
<u>video data</u> : Each video lasts roughly two-minute long, we remove the first 30 seconds and last 30 seconds, only collect the EEG data during the middle 1 minute.

C.EEG DATA ACQUISITION

The EEGs were recorded from 128 channels using a Net Amps 300 amplifier (Electrical Geodesic Inc. (EGI), USA). The Ag/AgCl electrodes were mounted into elastic net. All the electrodes were referenced to the Cz position. The impedance of all the electrodes was maintained below 50 K throughout the recording. The signals were digitized at 500 Hz with a notch filter at 50 Hz. The EEG amplifier was placed inside the experiment room. The amplified and digitized EEG signal was transmitted to Net Station 4.43 recording software operating on the computer placed outside the experiment room via fibre optic cables.

D.EEG ARTIFACT REDUCTION

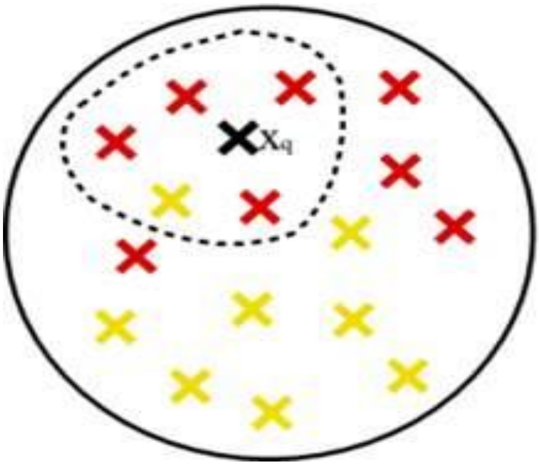
EEG signals were pre-processed offline by employing a 0.1-Hz filter to remove the DC artefacts and a 50-Hz notch filter to remove the line noise. Nineteen EEG channels according to the 10-20 system were selected against the average mastoid reference. Pre-processing of EEG signals was performed in Net Station 4.43. Further processing was performed at a sampling rate of 128 Hz. The eye-blink and muscle artefacts were manually removed by discarding that portion of the recording from the data. Both eye-blink and muscle artefacts were detectable by visual inspection. For example, the eye-blinks created a peak at approximately 10 Hz, while the muscle artefacts appeared at a higher frequency in the power spectral density graph of the EEG signals. The internal consistency and reliability of the cleaned EEG data were measured by computing the split- half reliability, and test-retest reliability measures that were above 90% for every EEG channel. To perform the analysis, sixty seconds of cleaned EEG data were selected.

E. KNN CLASSIFICATION ALGORITHM

KNN is a method of classification, one among the simplest method in ML, great way to understand classification and ML in general. At basic it is essentially classification by finding the most similar data points in training data and making an educated guess based on their classification. It stores all the available cases and classifies the new data or case based on a similarity measure. It uses least distant measure (Euclidean distance and Manhattan distance).

KNN is also called as a lazy learner, it classifies the data points based on the points that are most similar to it. It uses test data to make an educated guess on what an unclassified point should be classified as. It is non parametric that is make of no assumptions. The model is made up entirely from the data given to it.

K identifies neighbor irrespective of labels. For Example , if we have K=5 classes X and X and we need to find class X ?



Than X is classified in X, since its more no. of nearest neighbor belongs to X category. This is how KNN algorithm works.

The KNN Algorithm

1. Loading the data

2. Initialize K to our desired number of neighbours.

3. For each example in the data

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection

4. Sort the ordered collection of distances from sorted collection.

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. For regression, return the mean of the K labels

8. For classification, return the mode of the K labels

Choosing the right value for K

To select the K that's correct for our data, we run the KNN algorithm many times with different values of K and choose the K that reduces the chances of errors we encounter while maintaining the algorithm's ability to precisely make predictions when it's given data it hasn't seen before.

Advantages

1. The algorithm is simple and easy to implement.

2. There's no need to build a model, tune several parameters, or make additional assumptions.
3. The algorithm is versatile. It can be used for classification, regression, and search (as we will see in the next section).

Disadvantages

1. The algorithm gets significantly slower as the number of examples and/or predictors/independent variables increase.

KNN in practice

KNN's main cons of becoming significantly slower as the volume of data increases make it an not practical choice in situations where predictions need to be made rapidly. Moreover, there are faster algorithms that can produce more accurate classification and regression results.

However, provided you have sufficient computing resources to speedily handle the data you are using to make predictions, KNN can still be useful in solving problems that have solutions that depend on identifying similar objects. An example of this is using the KNN algorithm in recommender systems, an application of KNN-search.

IV.RESULT

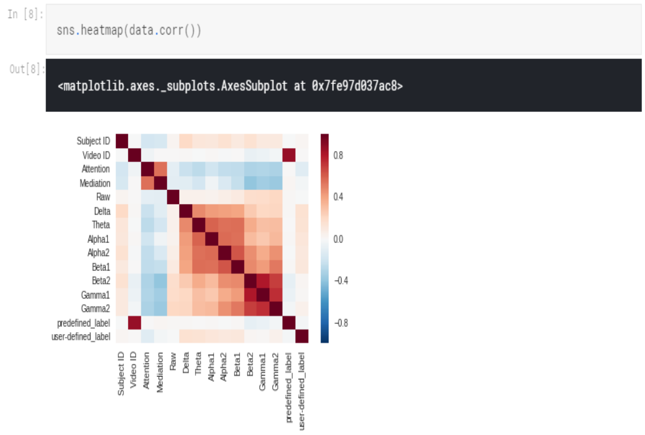
We discuss the prediction results and performance of KNN algorithm employed for this work. In classification reports we have obtained 83% accuracy performance using KNN algorithm for EEG data of students watching videos.

	Prediction	Recall	F1 score	Support
0.0	0.83	0.82	0.83	867
1.0	0.83	0.84	0.84	977
average/total	0.83	0.83	0.83	364

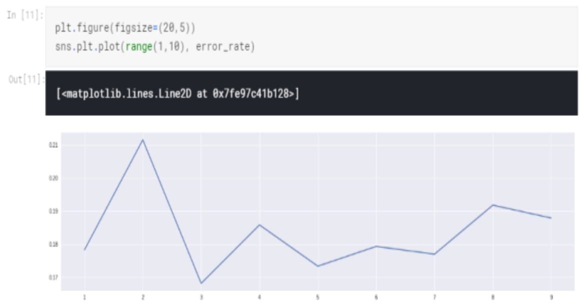
In this work, we have collected EEG signal data from 10 college students while they watched MOOC video clips. We extracted online education videos that are assumed not to be confusing for college students, such as videos of the introduction of basic algebra or geometry. We also prepare videos that are expected to confuse a typical college student if a student is not familiar with the video topics like Quantum Mechanics, and Stem Cell Research. We prepared 20 videos, 10 in each category and applied KNN classification algorithm which gives 83% performance accuracy.

As future work, firstly we can implement this same system by using other stress detection technologies like EEG, EMG, GSR etc. For advanced research we need to go through additional research paper adding up more details and descriptions. Other feature extraction techniques will results into other classification results. Third , for vary large amount of data, deep learning algorithm can be used for evaluating over fitting by implementing Deep Belief Network (DBN) provided by Deep Net package of R language is used for this purpose.

Showing correlation between different fields



Finding value of K with low error rate



V.ACKNOWLEGEMENT:

There have many contributors for this to take shape and authors are thankful to all of them.

VI. CONCLUSION AND FUTURE WORK:

VII. REFERENCES:

[1] Qianli Xu, Member, IEEE, Tin Lay Nwe, Member, IEEE, and Cuntai Guan, Senior Member, IEEE, "Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals", IEEE Journal of Biomedical And Health Informatics, Vol. 19, No. 1, January 2015.

[2] Chee-Keong Alfred Lim and Wai Chong Chia, "Analysis of Single-Electrode EEG Rhythms Using MATLAB to Elicit Correlation with Cognitive Stress", International Journal of Computer Theory and Engineering, Vol. 7, No. 2, April 2015.

[3] Tong Chen, Peter Yuen, Mark Richardson, Guangyuan Liu, and Zhishun She, Senior Member, IEEE, "Detection of Psychological Stress using a Hyper spectral Imaging Technique", IEEE Transactions on Affective Computing, Vol. 5, No. 4, October-December 2014.

[4] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tr¨oster, Member, IEEE, and Ulrike Ehlert, "Discriminating Stress from Cognitive Load Using a Wearable EDA Device", IEEE Transactions On Information Technology In Biomedicine, Vol. 14, No. 2, March 2010

[5] Awanis Romli, Arnidcha Peri Cha, "An Expert System for Stress Management", November 2009

[6] Jennifer A. Healey and Rosalind W. Picard, "Detecting Stress during Real-World Driving Tasks using Physiological Sensors", IEEE Trans. Intell. Transp. Syst., vol. 6, no. 2, pp. 156–166, Jun. 2005.

[7] J. F. Alonso, S. Romero, M. R. Ballester, R. M. Antonijoan, and M.A.Mañanas, ''Stress assessment based on EEG univariate features and functional connectivity measures,'' Physiol. Meas., vol. 36, no. 7, p. 1351, 2015.

[8] I. Dziembowska, P. Izdebski, A. Rasmus, J. Brudny, M. Grzelczak, and P. Cysewski, ''Effects of heart rate variability biofeedback on EEG alpha asymmetry and anxiety symptoms in male athletes: A pilot study,'' Appl.Psychophysiol. Biofeedback, vol. 41, no. 2, pp. 141–150, Jun. 2016.

[9] P. Cipresso et al., ''EEG alpha asymmetry in virtual environments for the assessment of stress-related disorders,'' Studies in Health Technology and Informatics, Newport Beach, CA, USA, 2012, pp. 102–104.

[10] P. M. Pandiyan and S. Yaacob, ''Mental stress level classification using eigenvector features and principal component analysis,'' Commun. Inf. Sci. Manage. Eng., vol. 3, no. 5, p. 254, 2013.

[11] N. Adnan, Z. Hj Murat, R. S. S. A. Kadir, and N. Hj Mohamad Yunos, ''University students stress level and brainwave balancing index: Comparison between early and end of study semester,'' in Proc. IEEE Student Conf.Res. Develop. (SCOReD), Dec. 2012, pp. 42–47.